Electro-Optical and Infrared Systems: Technology and Applications XX

Duncan L. Hickman
Helge Bürsing
Gary W. Kamerman
Ove Steinvall
Editors

3–5 September 2023
Amsterdam, Netherlands

Sponsored by
SPIE

Cooperating Organisations
Cranfield University (United Kingdom)

Published by
SPIE

Volume 12737

The papers in this volume were part of the technical conference cited on the cover and title page. Papers were selected and subject to review by the editors and conference program committee. Some conference presentations may not be available for publication. Additional papers and presentation recordings may be available online in the SPIE Digital Library at SPIEDigitalLibrary.org.

The papers reflect the work and thoughts of the authors and are published herein as submitted. The publisher is not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Please use the following format to cite material from these proceedings:
Author(s), "Title of Paper," in *Electro-Optical and Infrared Systems: Technology and Applications XX*, edited by Duncan L. Hickman, Helge Bürsing, Gary W. Kamerman, Ove Steinvall, Proc. of SPIE 12737, Seven-digit Article CID Number (DD/MM/YYYY); (DOI URL).

Printed in the United States of America by Curran Associates, Inc., under license from SPIE.

Publication of record for individual papers is online in the SPIE Digital Library.

SPIE. DIGITAL LIBRARY

SPIEDigitalLibrary.org

Paper Numbering: A unique citation identifier (CID) number is assigned to each article in the Proceedings of SPIE at the time of publication. Utilization of CIDs allows articles to be fully citable as soon as they are published online, and connects the same identifier to all online and print versions of the publication. SPIE uses a seven-digit CID article numbering system structured as follows:
▪ The first five digits correspond to the SPIE volume number.
▪ The last two digits indicate publication order within the volume using a Base 36 numbering system employing both numerals and letters. These two-number sets start with 00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 0A, 0B … 0Z, followed by 10-1Z, 20-2Z, etc. The CID Number appears on each page of the manuscript.

Contents

iii

Conference Committee

Symposium Chair

Ric Schleijpen, TNO (Netherlands)

Conference Chairs

Duncan L. Hickman, Tektonex Ltd. (United Kingdom)
Helge Bürsing, Fraunhofer-Institut für Optronik, Systemtechnik und
　Bildauswertung (Germany)
Gary W. Kamerman, FastMetrix Industries, LLC
Ove Steinvall, FOI-Swedish Defence Research Agency (Sweden)

Conference Program Committee

Gianni Barani, Leonardo (Italy)
Piet Bijl, TNO (Netherlands)
Rainer Breiter, AIM INFRAROT-MODULE GmbH (Germany)
Judith Dijk, TNO (Netherlands)
Bernd Eberle, Fraunhofer-Institut für Optronik, Systemtechnik und
　Bildauswertung (Germany)
Natan S. Kopeika, Ben-Gurion University of the Negev (Israel)
Robert A. Lamb, SELEX ES (United Kingdom)
Daniel A. Lavigne, Defence Research and Development Canada,
　Valcartier (Canada)
Gino Putrino, The University of Western Australia (Australia)
Stanley R. Rotman, Ben-Gurion University of the Negev (Israel)
Frank Rutz, Fraunhofer-Institut für Angewandte Festkörperphysik IAF
　(Germany)
Philip J. Soan, Defence Science and Technology Laboratory
　(United Kingdom)

High-performance silicon-based nano-thermoelectric bolometers for uncooled infrared sensing

Aapo Varpula*[a], Anton Murros[a], Kuura Sovanto[a], Arto Rantala[a], David Gomes Martins[a],
Kirsi Tappura[a], Jonna Tiira[a], Mika Prunnila[a]

[a]VTT Technical Research Centre of Finland Ltd, Tietotie 3, FI-02150 Espoo, Finland

ABSTRACT

Infrared (IR) sensors and photodetector arrays are employed in various imaging applications (such as night vision), remote temperature measurement, and chemical analysis. These applications are in space and environmental sensing, transport, health and medicine, safety, security, defense, industry, agriculture, etc. Optical chemical analysis employs IR absorption spectroscopy which enables the identification and quantification of gases, liquids, and materials based on their unique absorption spectra which are feature-rich in the IR region.

State-of-the-art (SoA) quantum photodetectors utilize either photoconductivity or the photovoltaic effect. Commercial quantum photodetectors are widely available in the spectral range from UV to short-wave infrared (SWIR), but in mid-wave IR (MWIR) and long-wave IR (LWIR), they require exotic materials and cooling to maintain high sensitivity. Thermal detectors (bolometers) are a competing technology that can reach high sensitivities in IR without the need for cooling and can be manufactured using widely available semiconductor technologies. SoA bolometers include resistive bolometers, diode- or transistor-based bolometers, and thermoelectric bolometers. By utilizing nanomaterials and integrated design, we have minimized the thermal mass and demonstrated fast and sensitive nano-thermoelectric IR bolometers with high thermoelectric efficiency.

We review the application and development of the silicon-based nano-thermoelectric infrared bolometers: modelling, design, fabrication, and electro-optical characteristics. The enabling materials, silicon nanomembranes, are also discussed, and the first devices used to test the potential of these nanomembranes, the electro-thermal devices, are reviewed and new experimental results are presented.

Keywords: Infrared detector, nano-thermoelectrics, mid-wavelength infrared, long-wavelength infrared, thermoelectric bolometer, nanomembrane, silicon

1. INTRODUCTION

Infrared (IR) sensors operating in the short- (SWIR), mid- (MWIR) and long-wave IR (LWIR) regions have various applications ranging from thermal imaging for night vision to remote thermometry and chemical analysis using infrared spectroscopy[1–3]. These applications find use in sectors such as space and environment sensing, transport, health and medicine, safety, security, defense, industry, and agriculture. Targeting the longer wavelengths – MWIR and LWIR – offers possibilities for gathering independent spectral information with passive operation by utilizing the background thermal emission of objects whose IR emission is strongest in this range[1]. IR detection in this range can be utilized for example in optical chemical analysis employing IR absorption spectroscopy to identify and quantify gases, liquids, and materials based on their unique IR absorption spectra[1,2].

The most commonly commercially available IR sensors are quantum detectors based on electron-hole pair generation, and thermal detectors (bolometers) based on radiation-induced heating of an absorber. Quantum detectors typically provide higher performance than bolometers in terms of speed and sensitivity. However, quantum detectors require cooling down from room temperature to cryogenic temperatures to reach their maximum performance, particularly in the LWIR range where photon energies are small[4]. Quantum detectors in the LWIR range also require expensive and toxic materials, such as HgCdTe and InAsSb[4]. On the other hand, state-of-the-art bolometers can reach relatively high sensitivities without cooling and can be manufactured using widely available CMOS-compatible technologies.

*aapo.varpula@vtt.fi; phone 358 40 357-1370; vttresearch.com

Furthermore, by leveraging nano-thermoelectrics – wherein the material properties of the thermoelectric elements are improved by dimensional scaling to the sub-100nm regime – bolometer performance can be improved, and the signal transduction noise can be significantly reduced, leading to a low noise transduction method for thermal sensors[5,6] and bolometers[4,7]. These benefits in conjunction with smart IR detector design allow thermoelectric bolometers to be a competitive technology to SoA quantum detectors, narrowing the gap between cooled and uncooled IR photodetector technologies[4].

It has been demonstrated that by reducing the thickness of silicon in two-dimensional systems, i.e. thin films, the thermal conductivity can be reduced by almost two orders of magnitude[8,9]. Experiments have consistently shown that a reduction in the silicon membrane thickness results in a proportional reduction in thermal conductivity[8,10–13], with an observed experimental minimum thermal conductivity of 9 W/mK measured in a 9-nm free-standing silicon membrane[8]. The reduction in thermal conductivity can be attributed to the spatial confinement of phonon modes in thin films[14–16] and increased phonon scattering at boundaries resulting in shortened phonon mean free path[17–19]. To maximize the thermoelectric figure of merit ZT, and therefore achieve high thermoelectric detector performance, the reduction in thermal conductivity must be balanced without compromising electrical conductivity. This challenges some nanostructuring methods, such as nanomeshes[20] and nanoholes[21], as a method to achieve large reductions in thermal conductivity, since the presence of pores and grain boundaries can reduce charge transport as well[9]. On the other hand, thickness reduction may enhance Seebeck coefficient and electron mobility due to quantum confinement[22–26].

We have demonstrated nano-thermoelectric IR bolometers[4,7] based on ultra-thin silicon membranes. They utilize silicon nanomembranes as phonon-engineered thermoelectric elements with enhanced ZT and metal nanomembranes and low thermal mass absorbers. The principle of operation of the IR bolometer is illustrated in Figure 1. Incident IR radiation is converted into a temperature change by the absorber element, which is thermally isolated from the ambient and substrate by suspension with support beams. Typically, the detector is sealed in a low thermal conductivity atmosphere or vacuum. The support beams are thermoelectric n- and p-type doped Si nanomembranes that act as thermocouple elements to convert the induced temperature gradient between the absorber and ambient (or substrate) into an electrical signal via the Seebeck effect. This mode of operation enables low power operation, as the bolometer does not require any external power for signal generation.
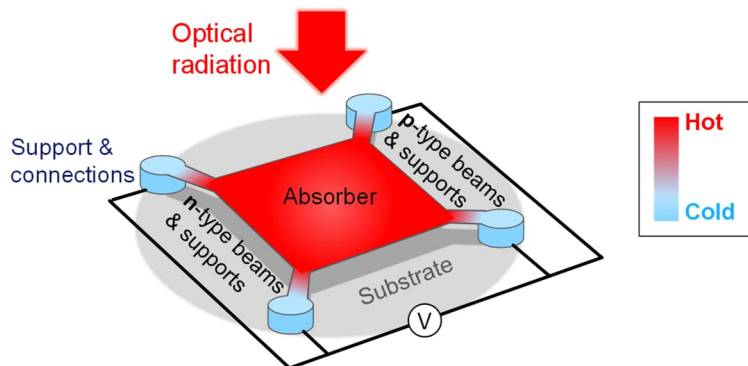
Figure 1. Schematic picture of a thermoelectric bolometer supported by 4 beams[7]. 2-beam bolometers with the same principle are also common. During the operation, the absorber absorbs the incident optical radiation and heats up. The formed temperature gradient is converted into electric voltage using p- and n-type thermoelectric materials (i.e., legs) placed in/on the supporting beams, or they can act simultaneously as supports themselves[4].

Thermoelectric transduction via the Seebeck effect has in principle low noise, with the main noise sources being Johnson-Nyquist and thermal fluctuation noise[5,27,28]. By scaling down the size of the thermocouple beams the phonon thermal conductance can be reduced, potentially allowing the devices to reach thermal fluctuation noise limited performance[6]. The thermal mass of the detectors is minimized by the absence of separate micromechanical supporting layers and by the integrated metal nanomembrane absorber. Further thermal mass reduction and thermal conductivity reduction is achieved by reducing the cross-sectional dimensions of the device. The reduced thermal mass enables high-

speed operation, which can also be exploited as increased sensitivity by altering the design of the bolometer in terms of beam dimensions, detector area, absorber fill factor, and film thicknesses.

In this article we review the development of the silicon-based nano-thermoelectric infrared bolometers. In section 2, the modelling of thermoelectric bolometers[4,7,29,30] is discussed. In section 3, we discuss the key enabling materials, silicon nanomembranes, and their fabrication processes for realizing the dimensionality reduction[39]. The availability of nanomembranes leads to the realization of nano-thermoelectric electro-thermal devices[5,6], which are discussed with new experimental results in section 4. Finally, the experimental IR bolometers[4,7] are discussed in section 5 and their estimated performance potential in section 6. We are also working on the readout integrated circuits, which are discussed in Ref. [7].

2. MODELING OF THERMOELECTRIC BOLOMETERS

2.1 Device model of thermoelectric bolometers

Thermoelectric bolometers are described by a combination of thermal resistor-capacitor (RC) and thermoelectric models[4,29,30]. The speed of a bolometer is described by thermal time constant

$$\tau = \frac{C_{\text{th}}}{G_{\text{th}}}, \tag{1}$$

where C_{th} is the thermal capacitance, i.e. the heat capacity, of the bolometer, and G_{th} is the thermal conductance of the bolometer (i.e. the thermal conductance between the absorber and the substrate and surroundings). The corresponding thermal cutoff angular frequency is $\omega_{\text{c}} = 1/\tau$. At frequencies well below ω_{c}, the dependence of the output voltage of a thermoelectric bolometer on the incident optical power P is characterized by the (voltage) responsivity[4]

$$R_V = \frac{\text{d}V}{\text{d}P}\bigg|_{\omega \ll \omega_{\text{c}}} = \frac{S\eta}{G_{\text{th}}}, \tag{2}$$

where the total Seebeck coefficient of the thermoelectric transducer of the bolometer is given by $S = S_{\text{p}} - S_{\text{n}}$, where S_{p} and S_{n} are the Seebeck coefficients of the p- and n-type thermoelectric elements, and η is the optical efficiency of the absorber.

The sensitivity of a detector is determined by the combination of the responsivity and noise. In thermoelectric bolometers there are the two dominant noise sources; the Johnson-Nyquist noise and the thermal fluctuation noise. The optical noise-equivalent power (NEP) corresponding to the Johnson-Nyquist noise is given by

$$\text{NEP}_{\text{JN}} = \frac{\sqrt{4k_{\text{B}}TR}}{R_V}, \tag{3}$$

where k_{B} is Boltzmann's constant, T the absolute temperature, and R the (total) resistance of the bolometer. The optical NEP corresponding to the thermal fluctuation noise is given by $\text{NEP}_{\text{th}} = \sqrt{4k_B T^2 G_{\text{th}}}/\eta$. These noise sources can be combined into total NEP as[4,29,5]

$$\text{NEP} = \sqrt{\text{NEP}_{\text{th}}^2 + \text{NEP}_{\text{JN}}^2} = \text{NEP}_{\text{th}}\sqrt{1 + \frac{1}{\widetilde{ZT}}}, \tag{4}$$

where the effective thermoelectric figure of merit of the bolometer is given by

$$\widetilde{ZT} = \frac{S^2 T}{G_{\text{th}}R}. \tag{5}$$

The specific detectivity D^* allows for comparison of different detector technologies and detectors with different active areas. It is given by

$$D^* = \frac{\sqrt{A_{\text{abs}}}}{\text{NEP}}, \tag{6}$$

where A_{abs} is the area of the absorber.

2.2 Model of quarter-wave resistive absorber

A quarter-wave resistive absorber illustrated in Fig. 2 is a radiofrequency and optical absorber design which allows effective absorption in bolometers[4,7,29]. It is also known as the Salisbury screen[29,31] or an antiresonant interference structure[32]. A quarter-wave resistive absorber consists of a conducting film and a reflector. The conducting film is placed near the reflector so that an optical cavity is formed. The optical absorption of the structure is maximized when the real or effective sheet resistance of the absorber is matched to the vacuum impedance.
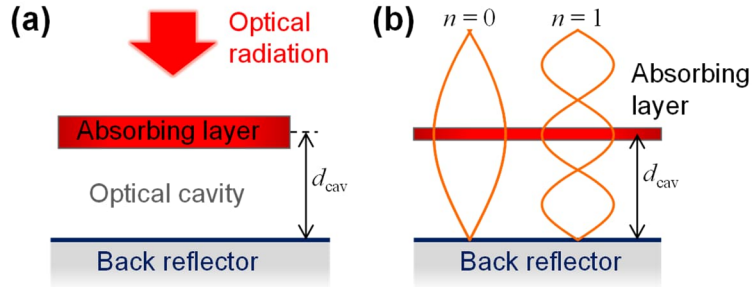
Figure 2. (a) Schematic cross-section of the quarter-wave absorber[29]. The absorbing layer placed above a back reflector. Optical cavity with the depth of d_{cav} is formed between the back reflector and the absorbing layer. (b) Illustration of optical waves in the absorber in the cases corresponding to two absorption peaks with the longest matching wavelengths, i.e. orders $n = 0$ and $n = 1$[29]. The optical absorption is strongest when an antinode of the optical wave is on the absorbing layer.

The optical properties of the quarter-wave resistive absorber structure can be modelled with full-wave electromagnetic simulations[4], transfer-matrix model[29], or analytical optical model[29] derived from a radio-frequency model of the Salisbury screen[33]. The first model is the most general one suitable for complex absorber geometries, but it is computationally heavy as it requires use of the finite-element method (FEM). The transfer matrix model[34,35] describes uniform layers of optical media, and this reduces the computational load considerably. This approach is often suitable for bolometers as their absorbers can be modelled using effective materials. We have shown that the fully analytical optical model matches with the transfer matrix model in idealized cases and agrees reasonably well in realistic cases[29]. The analytical model enables very rapid calculations.

In the analytical model, the absorption spectrum of an ideal quarter-wave resistive absorber consists of periodic absorption peaks at the wavelengths[29]

$$\lambda_{peak} = \frac{4 d_{cav}}{1 + 2n},\tag{7}$$

where d_{cav} is the depth of the cavity and the order $n = 0, 1, 2, 3, \dots$. The relative bandwidth of the peak with the lowest matching wavelength ($n = 0$) is rather broad: 280 % relative full-width-at-half-maximum (FWHM) bandwidth[29]. The cavity depth d_{cav} defines the spectral range of the structure. At the wavelengths of eq. (7), the absorptance of an ideal quarter-wave absorber is given by[7,29]

$$A = \frac{4 \frac{Z_{vac}}{R_{sh}}}{\frac{Z_{vac}}{R_{sh}} \left(\frac{Z_{vac}}{R_{sh}} + 2 \right) + 1},\tag{8}$$

where the vacuum impedance, or the impedance of free space, is $Z_{vac} \approx 376.73\ \Omega$ and R_{sh} is the (effective) sheet resistance of the absorber. Eq. (8) is plotted in Fig. 3, showing that rather high absorption can be achieved with a wide range of sheet resistance R_{sh} values (for example for A > 50%, R_{sh} can range between 65 and 2196 Ω[29]).
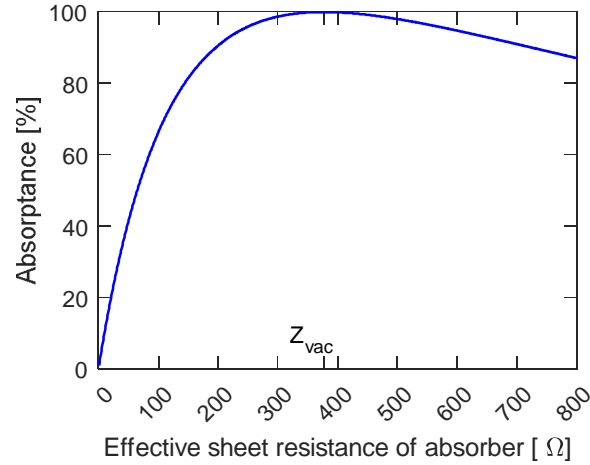
Figure 3. Dependence of absorptance A of the ideal quarter-wave resistive absorber on the effective sheet resistance of the absorber R_{sh} and vacuum impedance $Z_{vac} \approx 376.73\ \Omega$[7], when the wavelength of the optical signal is at the maximum of an absorber peak. Calculated using eq. (9).

The full dependence of the absorptance A on wavelength λ is given by

$$A = \frac{4\frac{Z_{vac}}{R_{sh}}\left[\sin\left(\frac{2\pi d_{cav}}{\lambda}\right)\right]^2}{\frac{Z_{vac}}{R_{sh}}\left(\frac{Z_{vac}}{R_{sh}} + 2\right)\left[\sin\left(\frac{2\pi d_{cav}}{\lambda}\right)\right]^2 + 1}. \tag{9}$$

Figure 4 shows a comparison between the transfer-matrix and analytical model in the cases of pure conductor and metallic VO$_2$ as a quarter-wave resistive LWIR absorber. The transfer matrix and the analytical models agree very well in the case of the pure conductor. In the VO$_2$ case, the peaks are slightly distorted and decay towards the smaller wavelengths. The decay is caused by the Drude-behavior of the complex permittivity and the imaginary part of the electric conductivity, which are present in real materials[29]. Despite these deviations, the analytical model is a powerful tool for rapid bolometer design drafting and description of the absorber behavior.
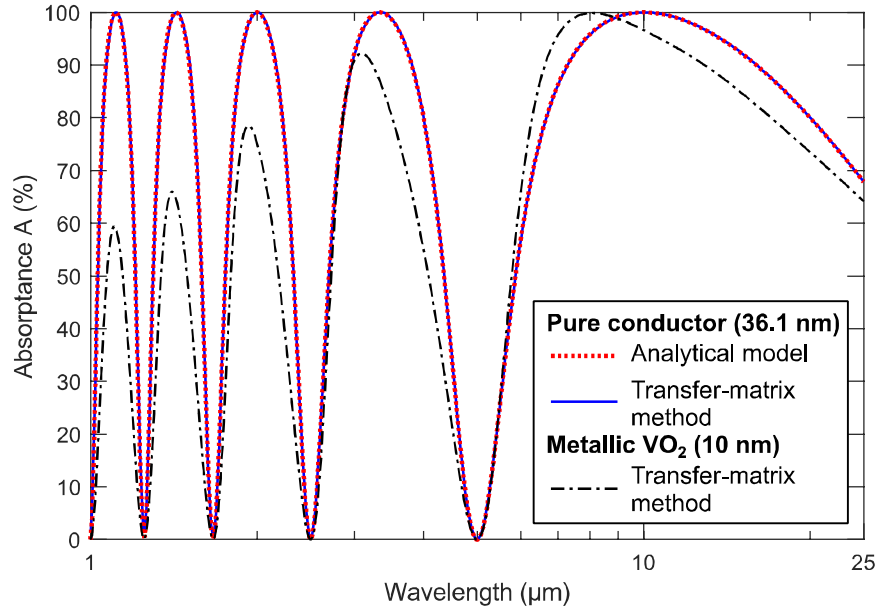
Figure 4. Comparison of absorption spectra of a quarter-wave resistive absorber calculated using the analytical model [eq. (9)] and the transfer-matrix method[29]. The depth of the optical cavity is d_{cav} = 2500 nm. Here the exemplary absorbing layer is metallic VO$_2$, which is modeled either as a pure conductor with the experimental DC electric conductivity or using experimental n and k data[29].

3. NANO-THERMOELECTRIC SILICON MEMBRANES

Silicon is an attractive detector material due to its common use in the semiconductor industry, and its cost-efficient, non-toxic, and generally abundant nature. However, due to its high thermal conductivity of 148 W/mK in the bulk[36], silicon has a low thermoelectric figure of merit, ZT, at room temperature ZT$_{300K}$ $\approx$ 0.01[8,9]. By shaping silicon into nano-scale membranes[8,37] or wires[38], the thermal conductance of silicon is significantly reduced – whilst preserving Seebeck coefficient and electrical conductivity virtually unaltered – resulting in a significantly improved ZT[5,4]. By leveraging nano-thermoelectrics in this manner, Si-based nanostructures form an extremely promising transducer materials platform for thermoelectric detectors.

Previously, we reported on the fabrication and characterization of large-area ultra-thin flat suspended single-crystalline (SC) Si membranes with controlled strain with membrane thicknesses down to 6 nm[39]. For more details, refer to Ref. [39]. The produced devices were thermally studied by Chávez-Ángel and co-workers[8] to produce the results shown in Figure 5. The fabrication process[39] involves thinning down a SOI layer by thermal oxidation and subsequent oxide removal to produce a Si film of the desired thickness. A Si$_3$N$_4$ film under tensile stress is deposited by low-pressure chemical vapor deposition (LPCVD) and patterned to form a stress-compensating frame that allows for compensation of compressive stress and strain tuning of the Si membrane. This stress-compensating frame is also utilized in our IR bolometers, discussed further in sections 4 and 5. Finally, the Si membranes were released by backside deep etching through the wafer.
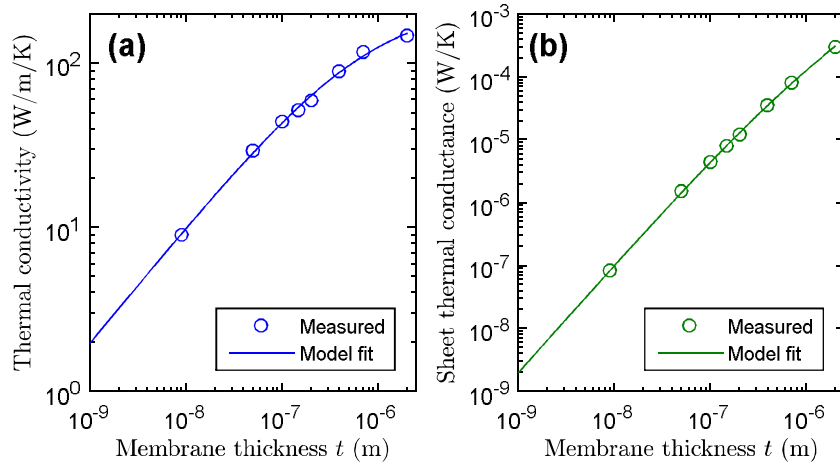
Figure 5. Experimental data[8] and fitted phenomenological model[30], $\kappa = \alpha/(1+\beta t^\gamma)$, of (a) thermal conductivity κ and (b) sheet thermal conductance $G_{\square th} = \kappa t$ of ultra-thin single-crystalline silicon membranes as functions of membrane thickness t. The values of the model fitting parameters are $\alpha = 228.4$ W/m/K, $\beta = 116.0$ /nm, and $\gamma = 0.7149$.

The thermal conductivity and conductance of ultra-thin single-crystalline silicon membranes as a function of membrane thickness are shown in Figure 5. For modeling purposes, a phenomenological model has been fitted to the measured data. The reduction in silicon membrane thickness results in almost two-orders of magnitude decrease in thermal conductivity, with an observed experimental minimum thermal conductivity of 9 W/(mK) in a 9-nm free-standing silicon membrane[8]. The reduction of thermal conductivity is caused by the spatial confinement of phonon modes in thin films[14–16] and increased phonon scattering at boundaries resulting in shortened phonon mean free path[17–19]. Reducing the thickness of the thermoelectric membrane is an effective method to improve ZT and therefore achieve high thermoelectric detector performance without significantly sacrificing electrical conductivity. Thickness scaling is thus an attractive solution to enable room temperature high-sensitivity and cost-effective IR detection.

Nano-thermoelectric Si can be realized by doping the nanomembranes into n- and p-type Si[5–4]. In addition to SC-Si, poly-crystalline Si nanomembranes can also be realized[6]. While higher charge-carrier mobilities allow lower resistivities to be obtained with SC-Si than with poly-Si, poly-Si affords simpler and more cost-efficient device fabrication. The characteristics of the nano-thermoelectric SC- and poly-Si are shown in Table 1. These were obtained from electro-thermal devices and IR bolometers which are described in sections 4 and 5. The Seebeck coefficients in Table 1 were measured with electro-thermal devices (see section 4) by employing the heater resistors as thermometers and thermocouples as Peltier coolers and heaters[5,6].

Table 1. Microstructure and thermoelectric properties of nano-thermoelectric silicon membranes. The total Seebeck coefficients were measured using electro-thermal devices[5,6].

Crystallinity	Membrane thickness	N-type resistivity	P-type resistivity	Total Seebeck coefficient S	Ref.
Single	40 nm	1.9 mΩcm	1.8 mΩcm	0.39 mV/K	[5]
Poly	100 nm	11 mΩcm	9.7 mΩcm	0.59 mV/K	[6]
Poly	80 nm	5.3 mΩcm	3.4 mΩcm	-	[4]
Poly	70 nm	4.7 mΩcm	3.0 mΩcm	-	[4]

4. ELECTRO-THERMAL DEVICES

We have investigated the potential of nano-thermoelectric silicon membranes in sensor and detector devices by fabricating electro-thermal devices out of them. These electro-thermal devices have a resistor for heating and a thermocouple for measuring thermal output voltage[5,6]. They are excellent vehicles for investigating the performance prospects of these materials. These devices can also be used as direct thermal detectors, for example in scientific experiments with scanning thermal microscopy (SThM)[40,41].

The scanning-electron microscopy (SEM) and optical microscopy images of the fabricated electro-thermal devices are shown in Figure 6. These devices utilize a silicon nitride (Si_xN_y) stress-compensation frame to control the strain of the nanomembrane[5,6,39]. In the devices the silicon nanomembranes are patterned into beams supporting a membrane section in the middle. In the middle membrane, there are meandering heater resistors made by doping this part of the membrane with boron (p+). The beams are also either n+ or p+ doped to provide electrical contact to the heater and to form the thermocouple together with aluminum contact metal.
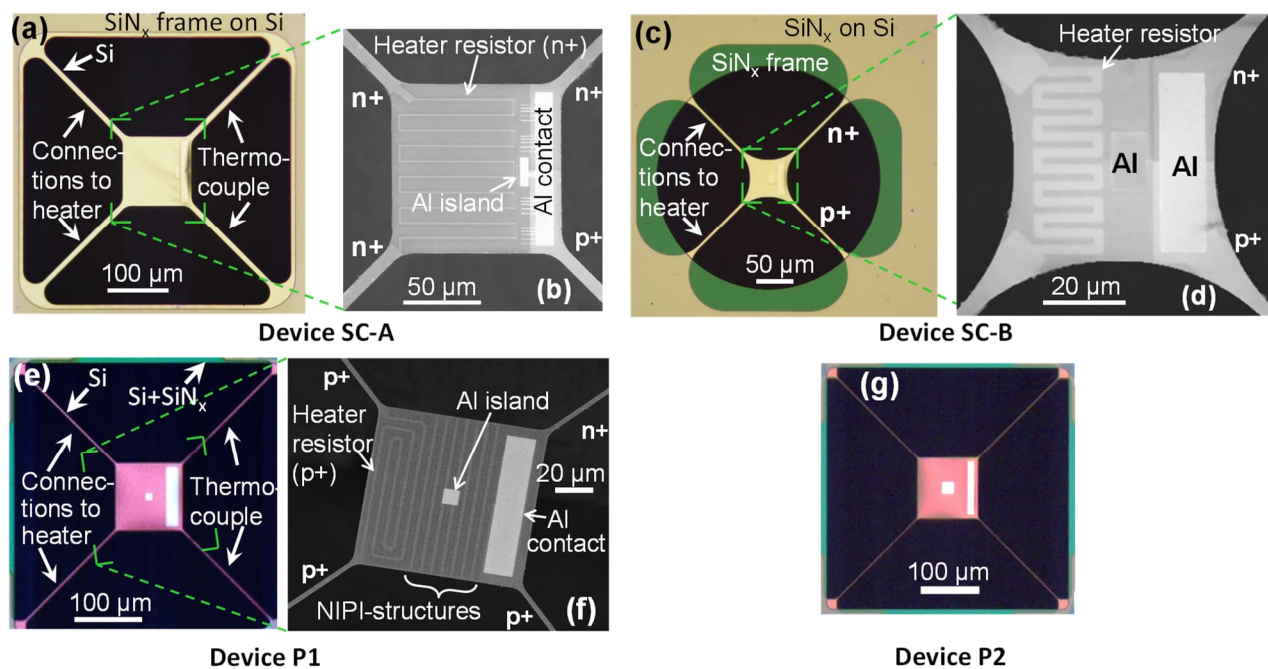
Figure 6. Optical (a, c, e, g) and scanning-electron (b, d, f) micrographs of devices SC-A[5] (a, b), SC-B[5] (c, d), P1[6] (e, f), and P2 (g). These electro-thermal devices have electric heater resistors for characterization and experimentation, and a thermocouple for electric temperature readout. The pair of the silicon beams on the left side of the devices provides the electrical connection for the meandered doped silicon heater resistor patterned in the silicon membrane. The 30-nm Al film forms an electrical contact between the ends of the n+ and p+ silicon beams acting as thermocouple. The devices have Al islands for scanning probing experiments. Device P1 has additional alternating n+ doped, undoped, and p+ doped regions (NIPI) between the heater meanders and the heater and the thermocouple to ensure that the heater is well electrically insulated from the rest of the device[6].

The electro-thermal devices were fabricated on 150-mm silicon wafers using the fabrication processes described in Refs. [5, 6]. In the case of single-crystalline (SC) Si devices, silicon-on-insulator (SOI) wafers were used. The targeted single-crystalline nanomembrane thickness was reached by thinning down the SOI layer by thermal oxidation and oxide stripping. In case of polycrystalline Si devices, standard silicon wafers were first oxidized to form a buried silicon oxide, and then a layer of 100 nm polysilicon was added by LPCVD on them. Next, both Si layers were doped selectively with boron and phosphorus by ion implantation and patterned by plasma etching. Then, a 280-nm-thick stress-compensation Si_xN_y layer was deposited by LPCVD. Next, a 30-nm-thick layer of Al (contact metal) was sputtered and patterned by wet

etching. Finally, the devices were released by deep-reactive etching through the silicon wafer from the back side and HF vapor etching of the buried oxide layer.

The measured characteristics of the electro-thermal devices are summarized in Table 2. The thermal time constants fall into the range 2.5–30 ms and are governed by the thickness of the Si nanomembranes, size of the middle membrane, and the dimensions of the supporting beams. Due to this, the thinner single-crystalline devices are faster. Noise measurements[5] have confirmed the theoretical prediction that the noise in these devices is fully described by the Johnson-Nyquist noise originating from the electrical resistance of the thermocouple. The NEPs in Table 2 were calculated using eq. (3) with electric responsivity, which is eq. (2) in the case $\eta = 1$. The poly-Si devices have larger responsivities mainly due to the larger Seebeck coefficient of the poly-Si (see Table 1), but the higher resistivity of the poly-Si (see Table 1) increases the thermocouple resistances leading to higher noise, which compromises NEP. The device beam dimensions have also an effect, which can be clearly seen between devices SC-A and SC-B, and P1 and P2. Overall, these electro-thermal devices are suitable for rapid detection of small thermal signals.

Table 2. Characteristics of electro-thermal devices[5,6].

Device	Material	Middle membrane size [μm^2]	Thermal time constant τ [ms]	Electrical responsivity [V/W]	Thermocouple resistance [kΩ]	Electrical NEP [pW/Hz$^{1/2}$]
SC-A[5]	40 nm SC-Si	110 x 110	9.4 ms	1180	27	18
SC-B[5]	40 nm SC-Si	50 x 50	2.5 ms	1960	39	13
P1[6]	100 nm poly-Si	100 x 100	13 ms	3090	215	19
P2	100 nm poly-Si	100 x 100	30 ms	6750	797	17

5. EXPERIMENTAL NANO-THERMOELECTRIC INFRARED BOLOMETERS

We have demonstrated nano-thermoelectric bolometers operating in the LWIR regime, targeted to maximize detector output for room-temperature thermal radiation with a wavelength maximum of approximately 10 μm[4,7]. Our bolometers have been fabricated with nano-thermoelectric silicon membranes that operate as both a signal transducer – converting the thermal signal into electrical output – and as mechanical supports for the integrated metal absorber. By utilizing the properties of nano-thermoelectric membranes and smart device design, the thermal mass and conductivity of the detectors can be tuned and optimized to target different application ranges, preferring high sensitivity or fast operation speed[4].

Scanning electron micrographs of the fabricated bolometers with cross-sectional device schematics are shown in Figure 7. The devices consist of an optical absorber supported by integrated n- and p-type ultra-thin silicon membranes, which act as thermocouples transducing incoming IR radiation into electrical signal. The metal optical absorber electrically connects the thermocouple pair and also acts as an infrared absorber, converting incoming IR radiation into a temperature change that is utilized by the thermocouples. Additionally, the devices utilize either a Si_xN_y or Al_2O_3 based stress-compensating frame discussed in sections 3 and 4 that is used to tune the stress of the detector membrane. The details of each layer are included in Figure 7. The metal-poly-Si absorber stack is patterned into a grid with plasma dry etching to control the optical impedance of the grid and to allow for easier removal of the underlying sacrificial oxide during device fabrication. For all device types the detector is electrically isolated from the substrate by suspension over an optical cavity, which acts as an optical resonator. The optical cavity and absorber form a quarter-wave resistive absorber (see section 2.2). The detector's performance can be optimized for specific wavelengths by varying the depth of the cavity. Here the detector output was maximized for the target LWIR wavelength of 10μm by setting the cavity depth to 2.5μm.
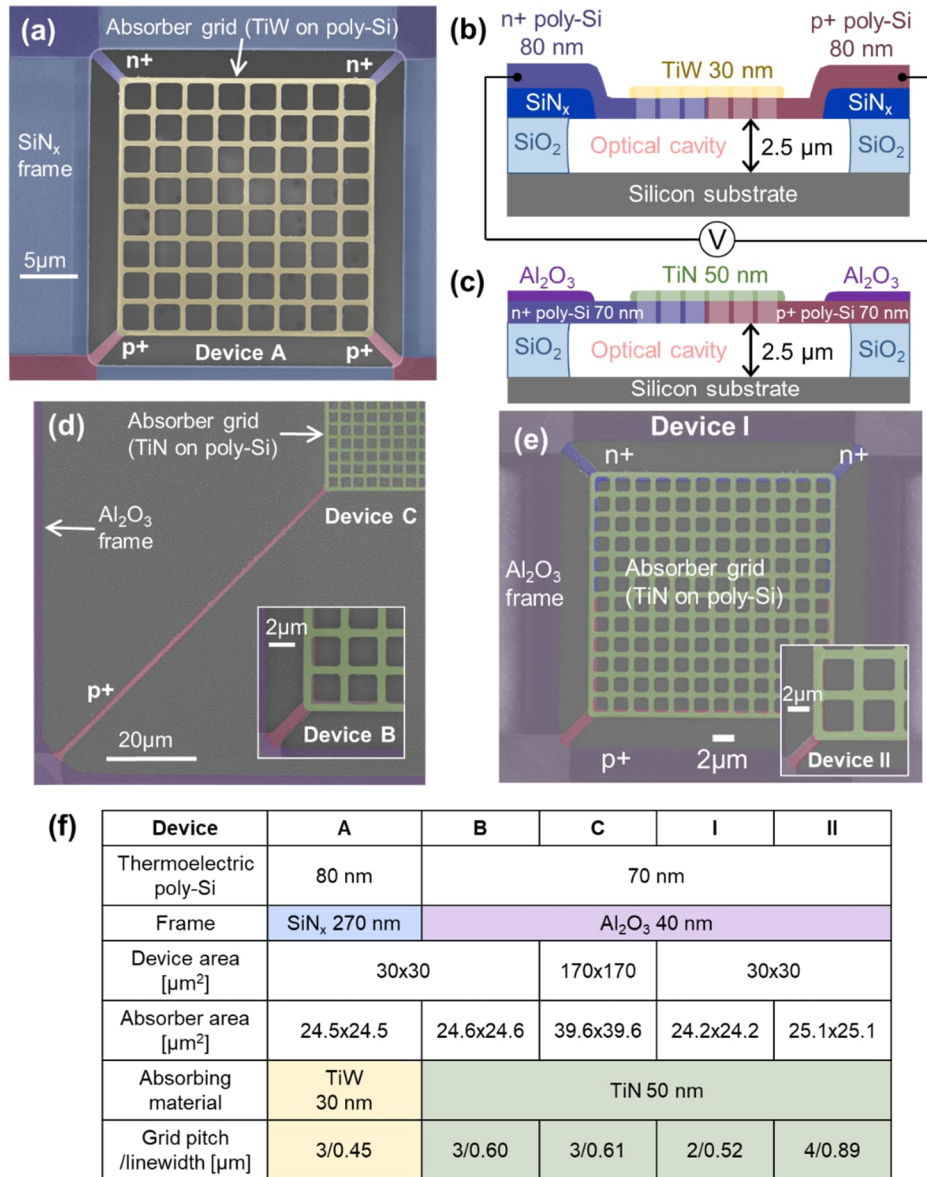
Figure 7. Pseudo-colored scanning-electron micrographs (a, d, e) and cross-sectional schematics (b, c) of nano-thermoelectric infrared bolometers A–C[4] and I–II[7]. The dimensions and material characteristics of the device layers are included in Table (f).

The nano-thermoelectric bolometers have been fabricated on 150 mm single-side polished p-type (1 – 50 Ωm) silicon wafers[4,7]. All wafers started with the deposition of 2.5μm sacrificial SiO_2 using LPCVD with tetraethyl orthosilicate (TEOS). In the case of the wafers with SiN_x stress-compensating frame, the SiN_x was deposited using LPCVD and patterned using plasma dry etching. After SiN_x deposition, the poly-Si layer was deposited by LPCVD, followed by patterned implantation and poly-Si plasma dry etching. In the case of the Al_2O_3 devices, the poly-Si was deposited and processed first, followed by atomic layer deposition of Al_2O_3, which was patterned using wet etching. For the SiN_x and Al_2O_3 detectors the absorber metals were sputtered: 30 nm TiW and 50 nm TiN, respectively. The absorber metal was patterned using plasma dry etching to form the detector grid and to enable access to selectively remove the underlying sacrificial oxide to suspend the detector membrane. Finally, the sacrificial SiO_2 is etched using HF-vapor to release the devices and to form the optical cavity.

The opto-electro-thermal characteristics of the fabricated nano-thermoelectric infrared bolometers are summarized in Table 3. The differences in the device speeds are determined by the absorber sizes (A_{abs}) and beam dimensions as they determine the thermal masses and thermal conductances. The small devices (A, B, I–II) are remarkably fast (for thermal detectors) with thermal time constants τ of 66 µs (Dev. A) and around 200 µs (larger devices) and the corresponding thermal cutoff frequencies around 2400 Hz and 800 Hz, respectively. These detectors allow infrared signals up to several kHz to be measured, depending on the required signal-to-noise ratios.

Table 3. Opto-electro-thermal characteristics of the nano-thermoelectric infrared bolometers[4,7].

Device	A	B	C	I	II
Thickness of thermoelectric poly-Si	80 nm	70 nm			
Absorbing material	TiW 30 nm	TiN 50 nm			
Absorber area [µm^2]	24.5×24.5	24.6×24.6	39.6×39.6	24.2×24.2	25.1×25.1
Measured thermal time constant τ [µs]	66	190	3596	197	208
Total bolometer resistance R_{tot} [kΩ]	50.0	11.9	86.4	6.5	9.1
Resistance of thermoelectric beams R_{beams} [kΩ]	2.5	2.7	58.1	2.7	2.7
Measured responsivity R_V to 200 °C BB radiation [V/W]	179	494	2930	334	398
Optical noise equivalent power[b] (NEP) [pW/Hz$^{1/2}$]	160	28	13	31	31
Optical NEP determined by R_{beams} only[b,c] [pW/Hz$^{1/2}$]	36	13	11	20	17
Specific detectivity[b,c] D^* [10^7 cmHz$^{1/2}$/W]	1.5	8.7	30.9	7.9	8.2
D^* determined by R_{beams} only[b,d] [10^7 cmHz$^{1/2}$/W]	6.9	18.4	37.7	12.3	15.1

[b]200 °C blackbody (BB) radiation.
[c]Sensitivity determined by total resistance, which is limited by additional series resistance such as contact resistance. Calculated using eq. (3) with $R = R_{tot}$.
[d]Limiting sensitivity determined by the resistance of the thermoelectric beams. Calculated using eq. (3) with $R = R_{beams}$.

The differences in the responsivities are caused by the slight differences in the total Seebeck coefficient S of the nanomembrane and larger differences in the absorber areas A_{abs}, thermal conductivities G_{th}, and absorber efficiencies η, which are, in turn, caused by differences in the non-optimal matching to vacuum impedance[4,7] (see section 2.2). The absorber efficiency can be improved by optimizing the absorber grid dimensions for better impedance matching and by improving the reflectance of the back reflector[4]. The fact that relatively high responsivities were obtained demonstrates how well the quarter-wave absorber can work even with non-optimal parameters.

The voltage noise of these bolometers is due to the Johnson-Nyquist noise[4], which is determined by the total resistance of the bolometer R_{tot} [see eq. (3)]. As Table 3 shows, the optical NEPs of the devices are determined by their responsivities R_V and the total resistances R_{tot}. Notably, the optical NEP of Device C is the same as the electrical NEP of electro-thermal device SC-B (Table 2).

The total resistance R_{tot} comprises of the resistance of the thermoelectric beams R_{beams}, the resistance of the absorber grid, and the potential contact resistance between the absorber metal and the poly-Si. The resistance of the absorber grid has ideally only little contribution to the total resistance of the devices as the effective sheet resistance of the absorber grid is close to the vacuum impedance (377 Ω)[4]. The poly-Si–metal contact resistance can be high, as in the case of Device A

(see Table 3). Therefore, the optimization of contact resistance by material and process development is crucial. The performance can be further improved by geometrical optimization of the thermoelectric beams for improved $\widetilde{ZT}$[4]. Optimization of the bolometer resistance and optical absorption will improve the optical NEPs of these bolometers to the same and lower level as the electrical NEPs of the electro-thermal devices.

To compare the nano-thermoelectric infrared bolometers with the state-of-the-art thermal detector technologies, the time constant and specific detectivities D^* of commercial and non-commercial LWIR bolometers are plotted in Figure 8. Our bolometers are fast, and the fastest detector in the graph is Device A. Our devices are also relatively sensitive, and by optimization of the device resistance and optical absorption their specific detectivities can be further improved[4]. In addition, we estimate that by utilizing 9-nm-thick poly-Si nanomembranes, we could, depending on the device design, realize either faster or more sensitive detectors than the state of the art.
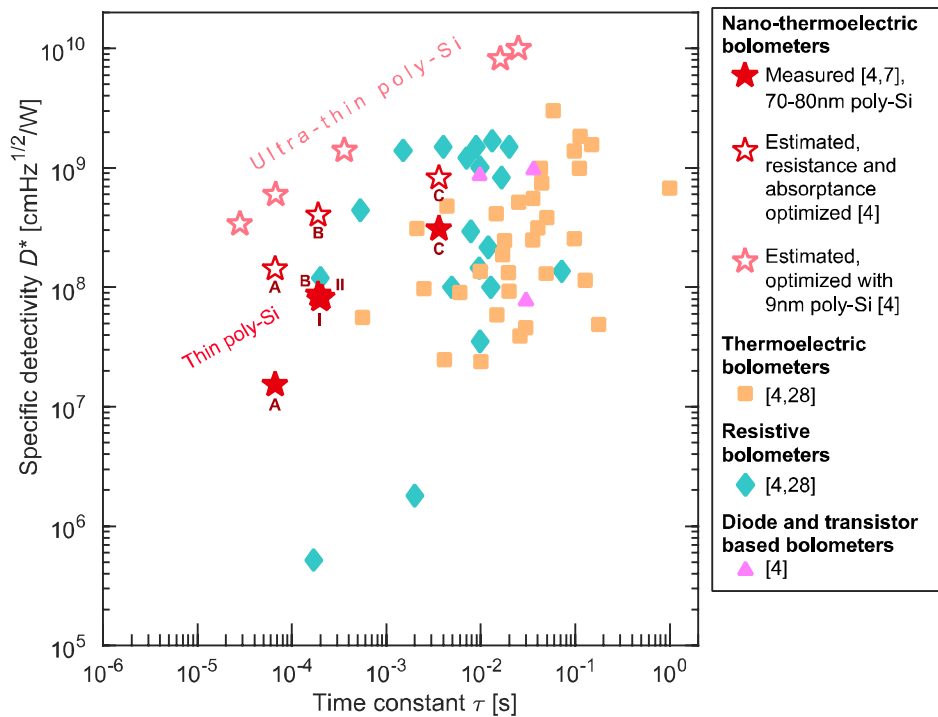
Figure 8. Time constants τ and specific detectivities D^* of the nano-thermoelectric bolometers[4,7] and of commercial and non-commercial LWIR bolometers[4,28].

6. ESTIMATION OF PERFORMANCE OF NANO-THERMOELECTRIC BOLOMETERS

We have used the device models of section 2.1 to show that silicon and metal nanomembranes can be used as high-performance active materials of thermoelectric bolometers[29,30]. The effect of the membrane thickness on the device performance is shown in Fig. 9, where D^* and τ of detectors with Ti absorbers and 3 different absorber sizes and leg geometries are plotted. The absorber area A_{abs} and the leg geometries defined by the number of squares in the legs N, determine the speed of the detectors via the thermal mass (C_{th}) and the thermal link (G_{th}). A_{abs} also determines D^* as it governs the amount of radiation absorbed. The membrane thickness t_{TE} has a profound effect on the device performance as it determines the thermal mass and the thermal conductivity of the nanomembrane (see Fig. 5). When t_{TE} decreases, G_{th} decreases both due to the geometric effect ($G_{th} \propto t_{TE}$) and the drastic decrease of thermal conductivity of the nanomembrane, which leads to increase in $\widetilde{ZT}$. These leads to increase in D^* with decreasing t_{TE}. The dependence of $\tau = C_{th}/G_{th}$ [eq. (1)] is more complex as both G_{th} and C_{th} decrease with decreasing t_{TE}. Dependencies of further device parameters are discussed in Refs. [29, 30]. In overall, these calculations show that in infrared detectors, nano-

thermoelectric silicon has the potential to realize the detector performance close to the fundamental limit determined by the thermal fluctuation noise only[42].
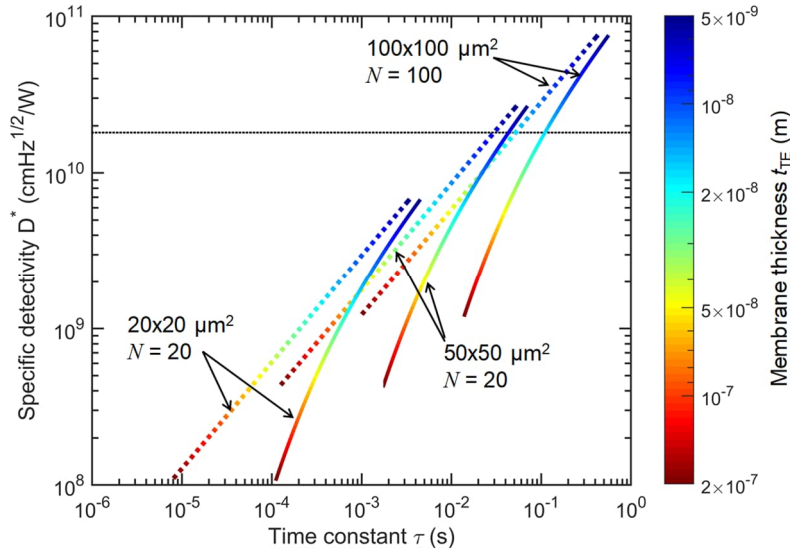
Figure 9. Calculated specific detectivities $D*$ and thermal time constants τ of Si-nanomembrane-based thermoelectric bolometers with 10 nm Ti absorber and different absorber areas and leg geometries as functions of the membrane thickness t_{TE}[29]. The absorber area A_{abs} and the number of squares in the n- and p-type thermoelectric legs N are indicated in the graph. The dotted lines are the case where the absorber comprises of Ti only, and the solid lines are the case where the absorber is made of both Ti and Si nanomembranes. Resistivity of 1 mΩcm and the thermal conductivity model of Fig. 5 was used for the Si-nanomembrane in the calculations. The black dashed line is the background radiation limit for thermal detectors in the full spectrum and half space case at 300 K[42]. Beyond this limit the estimated detector $D*$ is limited only by the external background noise. Further parameter values and details are given in Ref. [29].

7. CONCLUSIONS

Phonon-engineered silicon nanomembranes enable the realization of cost-effective and high-performance electro-thermal devices for thermal experiments and infrared bolometers for a wide range of optical imaging and sensing applications such as night vision and optical chemical sensing. The nano-thermoelectric IR bolometer design is based on the quarter-wave resistive absorber, where the wavelength range of interest can be selected by the cavity depth and absorber design. In principle, designs could be made for the entire infrared range. We have shown that the nano-thermoelectric infrared bolometer technology can exceed the state-of-the-art thermal detectors.

ACKNOWLEDGEMENTS

REFERENCES

[1] Budzier, H. and Gerlach, G., [Thermal infrared sensors: theory, optimisation and practice], Wiley (2011).

[2] Baldini, F., Chester, A. N., Homola, J. and Martelluci S., [Optical Chemical Sensors], Springer (2006).

[3] Popa, D. and Udrea, F., "Towards Integrated Mid-Infrared Gas Sensors," Sensors 19, 2076 (2019).

[4] Varpula, A., Tappura, K., Tiira, J., Grigoras, K., Kilpi, O.-P., Sovanto, K., Ahopelto, J. and Prunnila, M., "Nano-thermoelectric infrared bolometers," APL Photonics 6, 036111 (2021).

[5] Varpula, A., Timofeev, A. V., Schepetov, A., Grigoras, K., Hassel, J., Ahopelto, J., Ylilammi, M. and Prunnila M., "Thermoelectric thermal detectors based on ultra-thin heavily doped single-crystal silicon membranes," Appl. Phys. Lett. 110, 262101 (2017).

[6] Varpula, A., Grigoras, K., Tappura, K., Timofeev, A.V., Shchepetov, A., Hassel, J., Ahopelto, J. and Prunnila, M., "Silicon Based Nano-Thermoelectric Bolometers for Infrared Detection," Proc. Eurosensors, 2(13), 894 (2018).

[7] Varpula, A., Murros, A., Sovanto, K., Rantala, A., Gomes Martins, D., Tappura, K., Tiira, J., and Prunnila, M., "Uncooled nano-thermoelectric bolometers for infrared imaging and sensing," Proc. SPIE 12417, Optical Components and Materials XX, 124170U, 124170U (2023).

[8] Chávez-Ángel, E., Reparaz, J. S., Gomis-Bresco, J., Wagner, M. R., Cuffe, J., Graczykowski, B., Shchepetov, A., Jiang, H., Prunnila, M., Ahopelto, J., Alzina, F. and Sotomayor-Torres, C. M, "Reduction of the thermal conductivity in free-standing silicon nano-membranes investigated by non-invasive Raman thermometry," APL Mater. 2, 012113 (2014).

[9] Mangold, C., Neogi, S. and Donadio, D., "Optimal thickness of silicon membranes to achieve maximum thermoelectric efficiency: A first principles study," Appl. Phys. Lett. 109 (5), 053902 (2016).

[10] Asheghi, M., Leung, Y. K., Wong, S. S. and Goodson, K. E., "Phonon-boundary scattering in thin silicon layers," Appl. Phys. Lett. 71 (13), 1798-1800 (1997).

[11] Ju, Y. S. and Goodson, K. E., "Phonon scattering in silicon films with thickness of order 100 nm," Appl. Phys. Lett. 74 (20), 3005-3007 (1999).

[12] Liu, W., Asheghi, M., "Thermal conduction in ultrathin pure and doped single-crystal silicon layers at high temperatures," Journal of Applied Physics 98 (12), 123523 (2005).

[13] Liu, X., Wu, X., Ren, T., "*In situ* and noncontact measurement of silicon membrane thermal conductivity," Appl. Phys. Lett. 98 (17), 174104 (2011).

[14] Balandin, A. and Wang, K., "Significant decrease of the lattice thermal conductivity due to phonon confinement in a free-standing semiconductor quantum well," Phys. Rev. B 58, 1544–1549 (1998).

[15] Huang, M.-J., Chang, T.-M., Chong, W.-Y., Liu, C.-K. and Yu, C.-K., "A new lattice thermal conductivity model of a thin-film semiconductor," International Journal of Heat and Mass Transfer 50 (1–2), 67-74 (2007).

[16] Lü, X., "Lattice thermal conductivity of Si nanowires: Effect of modified phonon density of states," Journal of Applied Physics 104 (5), 054314 (2008).

[17] Tang, G. H., Zhao, Y., Zhai, G. X. and Bi, C., "Phonon boundary scattering effect on thermal conductivity of thin films," Journal of Applied Physics 110 (4), 046102 (2011).

[18] Liu, W. and Asheghi, M., "Phonon–boundary scattering in ultrathin single-crystal silicon layers," Appl. Phys. Lett. 84 (19), 3819–3821 (2004).

[19] Martin, P., Aksamija, Z., Pop, E. and Ravaioli, U., "Impact of Phonon-Surface Roughness Scaling on Thermal Conductivity of Thin Si Nanowires," Phys. Rev. Lett. 102, 125503 (2009).

[20] Yu, J.-K., Mitrovic, S., Tham, D., Varghese, J. and Heath, J. R., "Reduction of thermal conductivity in phononic nanomesh structures," Nature Nanotech 5, 718–721 (2010).

[21] Tang, J., Wang, H.-T., Lee, D. H., Fardy, M., Huo, Z., Russell, T. P. and Yang, P., "Holey Silicon as an Efficient Thermoelectric Material," Nano Letters 10 (10), 4279-4283 (2010).

[22] Hicks, L. D. and Dresselhaus, M. S., "Effect of quantum-well structures on the thermoelectric figure of merit," Phys. Rev. B 47, 12727 (1993).

[23] Vo. T. T. M., Williamson, A. J., Lordi, V. and Galli, G., "Atomistic Design of Thermoelectric Properties of Silicon Nanowires," Nano Letters 8 (4), 1111-1114 (2008).

[24] Ikeda, H. and Salleh, F., "Influence of heavy doping on Seebeck coefficient in silicon-on-insulator," Appl. Phys. Lett. 96 (1), 012106 (2010).

[25] Neophytou, N., Zianni, X., Kosina, H., Frabboni, S., Lorenzi, B. and Narducci, D., "Simultaneous increase in electrical conductivity and Seebeck coefficient in highly boron-doped nanocrystalline Si," Nanotechnology 24, 205402 (2013)

[26] Neophytou, N., Karamitaheri, H. and Kosina, H., "Atomistic calculations of the electronic, thermal, and thermoelectric properties of ultra-thin Si layers," J Comput Electron 12, 611–622 (2013).

[27] Richards, P. L., "Bolometers for infrared and millimeter waves," J. Appl. Phys. 76, 1 (1994).

[28] Dillner, U., Kessler, E. and Meyer, H.-G., "Figures of merit of thermoelectric and bolometric thermal radiation sensors," J. Sens. Sens. Syst. 2, 85 (2013).

[29] Varpula, A., Tappura, K., Tiira, J., Grigoras, K., Viherkanto, K., Ahopelto, J. and Prunnila, M., "High-performance infrared thermoelectric bolometers based on nanomembranes," Proc. SPIE 11289, Photonic and Phononic Properties of Engineered Nanostructures X, 112891O (2020).

[30] Varpula, A., Timofeev, A. V., Schepetov, A., Grigoras, K., Ahopelto, J. and Prunnila M., "Thermoelectric bolometers based on silicon membranes," Proc. SPIE 10246, Smart Sensors, Actuators, and MEMS VIII, 102460L (2017).

[31] Salisbury, W. W., "Absorbent body for electromagnetic waves," U. S. Patent 2,599,944 (1952).

[32] Talghader, J. J., Gawarikar, A. S. and Shea, R. P., "Spectral selectivity in infrared thermal detection," Light Sci. Appl. 1, e24 (2012).

[33] Fante, R. L. and McCormack, M. T., "Reflection properties of the Salisbury screen," IEEE Trans. Antennas Propagat. 36(10), 1443–1454 (1988).

[34] Katsidis, C. C. and Siapkas, D. I., "General transfer-matrix method for optical multilayer systems with coherent, partially coherent, and incoherent interference," Appl. Opt., AO 41(19), 3978–3987 (2002).

[35] Born, M. and Wolf, E., [Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light], Elsevier (2013).

[36] Glassbrenner, C. J. and Slack, G. A., "Thermal Conductivity of Silicon and Germanium from 3°K to the Melting Point," Phys. Rev. 134, A1058–A1069 (1964).

[37] Neogi, S., Reparaz, J. S., Pereira, L. F. C., Graczykowski, B., Wagner, M. R., Sledzinska, M., Shchepetov, A., Prunnila, M., Ahopelto, J., Sotomayor-Torres, C. M., and Donadio, D., "Tuning Thermal Transport in Ultrathin Silicon Membranes by Surface Nanoscale Engineering," ACS Nano 9 (4), 3820 (2015).

[38] Boukai, A., Bunimovich, Y., Tahir-Kheli, J., Yu, J.-K., Goddard III, W. A. and Heath, J. R., "Silicon nanowires as efficient thermoelectric materials," Nature 451, 168–171 (2008).

[39] Shchepetov, A., Prunnila, M., Alzina, F., Schneider, L., Cuffe, J., Jiang, H., Kauppinen, E. I., Sotomayor Torres, C. M. and Ahopelto, J., "Ultra-thin free-standing single crystalline silicon membranes with strain control," Appl. Phys. Lett. 102, 192108 (2013).

[40] Varpula, A., Renahy, D., Grigoras, K., Tappura, K., Timofeev, A. V., Shchepetov, A., Hassel, J., Ahopelto, J., Gomés, S., and Prunnila, M., "Silicon nano-thermoelectric detectors for sensing and instrumentation applications," Nanoscale and Microscale Heat Transfer VI (NMHT-VI), Eurotherm seminar No 111, Levi, Lapland, Finland, 2 – 7 December 2018.

[41] Gomès, S., Assy, A., Chapuis, P.-O., "Scanning thermal microscopy: A review," Phys. Status Solidi 212 (3), 477 (2015).

Uncooled Thermal MWIR Imagers for High-Temperature Imaging Applications

M. Michel[a], S. Blaeser[a], A. Litke[a], E. Zakizade[a], S. Weyers[a], and D. Weiler[a]

[a]Fraunhofer IMS, 47057 Duisburg, Germany

ABSTRACT

Microbolometers are well-established sensing elements for uncooled thermal imaging applications. Benefits in both costs and power consumption allow microbolometers to be a competitive alternative as compared to cooled infrared detectors in most common infrared imaging scenarios. Until now, microbolometers are designed and optimized for the long wavelength infrared (LWIR) regime ranging from 8 μm to 14 μm. However, the mid wavelength infrared (MWIR) regime ranging from 3 μm to 5 μm is also of great interest for a wide range of applications that can benefit from the advantages of a technological concept relying on microbolometers. For this reason, Fraunhofer IMS developed an uncooled thermal imager based on microbolometers targeting the wavelength spectrum of the MWIR for high temperature imaging applications. A novel imager technology based on Fraunhofer IMS's microbolometer process for lateral leg bolometers providing QVGA resolution (320 x 240) in case of a pixel pitch of 17 μm but transferred to the MWIR regime will be presented here. In order to increase the sensitivity in the MWIR, the transmission characteristics of the vacuum package have been adopted to meet the requirements of this wavelength region. The resulting spectral sensitivity of our MWIR imagers was verified by means of an electro-optical test setup making use of a high temperature black body radiator. In addition, the actual design of the microbolometer membrane has been optimized to reduce the overall thermal capacitance, resulting in thermal time constants up to 30 % lower than those of our standard LWIR imager.

Keywords: high-temperature imaging, uncooled thermal imaging, microbolometer, LWIR, MWIR, spectral sensitivity, thermal time constant

1. INTRODUCTION

According to Planck's law, objects with a heat signature in the mid wavelength infrared (MWIR) ranging from 3 μm to 5 μm also emit thermal radiation in the long wavelength infrared (LWIR) regime ranging from 8 μm to 14 μm. Therefore, in many imaging scenarios, objects that emit thermal radiation primarily in the MWIR can also be visualized using LWIR cameras. However, there is a wide range of situations where an explicit distinction between MWIR and LWIR is beneficial. Different weather conditions (humidity, rain, clouds, ..), background structures, sun reflections as well as the object's temperature and distance may have an impact on the detectability of certain structures or targets.[1,2] Also flame detection, the visualization and analysis of gases or combustions at high temperatures are relevant imaging scenarios in the MWIR.[3,4]

However, uncooled thermal MWIR imaging may not only be of great interest for civil use cases. Also security and military applications might drastically benefit from the introduction of this wavelength regime in addition to the standard LWIR. Uncooled thermal MWIR imaging not only allows for a visualization of reflections resulting in a more natural, i.e. VIS-like, viewing but is also less susceptible to atmospheric conditions stemming from either humidity in case of water vapor or CO_2 in the atmosphere. The latter circumstances result in an increased absorption, scattering, and refraction of LWIR radiation whereas MWIR radiation may penetrate the atmosphere much less affected. In consequence, uncooled thermal MWIR imaging may enable longer detection ranges, which are more than 2 times farther than in case of visualization through the LWIR regime.[5]

Further author information: (Send correspondence to M. Michel)
M. Michel: E-mail: marvin.daniel.michel@ims.fraunhofer.de

There are already established technical solutions for the aforementioned application scenarios using cooled detector technologies. Although some benefits of a higher sensitivity and higher frame rate cannot be easily overcome by an uncooled detector technology, there still is plenty of room for microbolometers in mobile and battery-operated cameras. In this paper we present our first results of our MWIR imager technology based on microbolometers with reduced thermal capacitance in a mid wave optimized vacuum chip-scale package.

2. CONCEPT OF UNCOOLED THERMAL MWIR IMAGERS

Uncooled thermal imaging realized by microbolometers as the actual sensing element relies on two technological key requirements, regardless of whether MWIR or LWIR is the respective wavelength regime of interest: On the one hand, a hermetic sealing of the microbolometers within a compact vacuum package is indispensable to minimize thermal losses to the environment stemming from convection. On the other hand, the absorption of the incoming infrared radiation within the microbolometer itself has to be maximized to ensure a sufficient heating of the sensing membrane correlating to a proportional change in electrical resistance with the latter being evaluated by a readout integrated circuit (ROIC).

However, hermetic sealing of the microbolometers within a compact vacuum package requires an implementation of an IR-transparent lid to allow the incoming infrared radiation to lossless penetrate the latter before reaching the actual microbolometer. Transmission of this IR-transparent lid is typically enhanced by fabricating an anti-reflective coating (ARC) on top. In this paper, we demonstrate an ARC not only optimized for the full MWIR regime but also tunable for the desired peak wavelength in the MWIR.

Absorption of uncooled thermal LWIR imagers is typically optimized by implementation of a so-called $\lambda/4$ resonator enabling standing waves between the microbolometer and a reflective area, typically deposited on top of the wafer substrate. Their resonance results in a maximized amplitude and intensity, respectively, and in consequence, in a maximized absorption in the microbolometer. However, the proven concept of the $\lambda/4$ resonator cannot be directly transfered from the LWIR to the MWIR regime as it would result in a drastically reduced distance between microbolometer and substrate increasing the risk of thermal losses in case of mutual contact. Therefore, alternative design concepts had to be developed to reliably shift the absorption peak from the LWIR to the MWIR regime.

2.1 MWIR Anti-Reflective Coating for Vacuum Packaging

Fraunhofer IMS has developed a simple design concept for anti-reflection coatings in the MWIR by using a multi-layer structure with a total thickness much smaller than the relevant wavelength. Such thin multilayer structures can change the effective refractive index of optical media and allow for high transmission for the wavelength of interest.[6] Compared to the different types of nanostructure anti-reflective coatings,[7] the multilayer coating is a simple and cost-effective method that provides design freedom to tune the transmission of the lid.

In this section, we present the results of simulation and optical investigation of the lid with high transmission in an optimized wavelength range. The simulation of the multilayer structures was performed using OpenFilters, an open-source software for designing and optimizing optical filters. A layer stack of silicon nitride (Si_3N_4) and silicon dioxide (SiO_2) was simulated as a multilayer anti-reflective coating on both sides of a silicon substrate. Fig. 1(a) shows an example of the simulated transmission spectra for different thicknesses of Si_3N_4 and a fixed thickness of 600 nm of SiO_2. Varying the nitride thickness results in a shift in the wavelength range and a change in intensity of the transmission peak. However, a significant reduction in the transmission spectrum is observed at about 9.3 µm due to the high extinction coefficient of SiO_2 at this explicit wavelength. Hence, by using a multilayer anti-reflective coating, control and tuning of the transmission spectrum of the lid for the desired spectral range can be achieved.
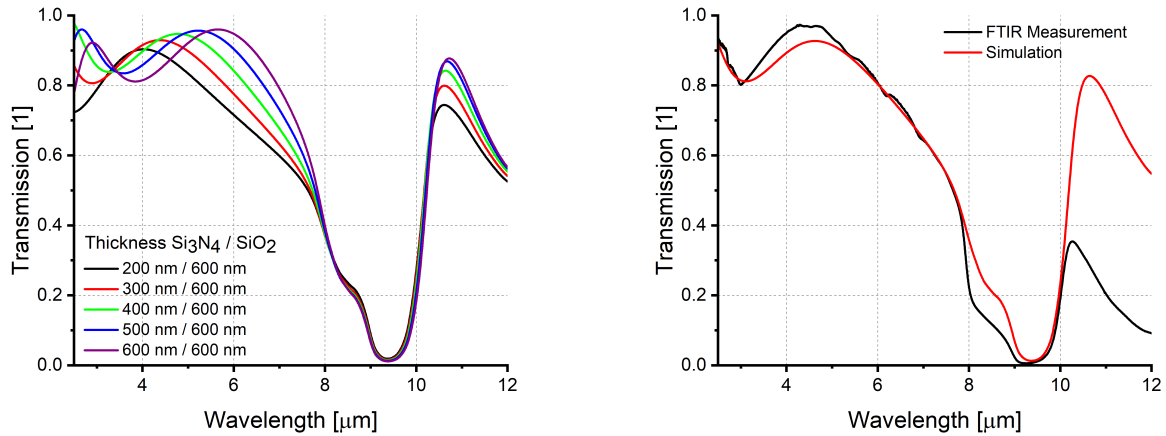
(a) Simulated Transmission spectra of layer stacks composed of Si_3N_4 and SiO_2.

(b) Comparison of simulated and measured transmission spectra for first time manufactured MWIR lid.

Figure 1. Transmission spectrum of MWIR lids.

Based on these simulation results, a MWIR-transparent lid was fabricated by using Si_3N_4 and SiO_2 on both sides of the silicon substrate. The optical characterization of the lid was performed using Fourier transform infrared (FTIR) spectroscopy. Fig. 1(b) shows the simulated and measured transmission spectra for the MWIR lid. A good match between the simulation and experimental results can be observed in the MWIR regime. However, the simulation overestimates the transmission for wavelengths above 8 µm. This overestimated transmission stems from the used material data[8] for Si_3N_4, which matches to the optical properties of our Si_3N_4 in the MWIR, but does not provide any data for the optical extinction coefficient. However, it is known from literature,[9] that Si_3N_4 has a non-negligible optical extinction coefficient in the LWIR. Since this data set did not fit to our material in the MWIR, we focused on a more accurate simulation of the transmission in the MWIR. All in all, an optimized MWIR lid with a high transmission spectrum in the MWIR and a maximum at about 4.5 µm was achieved.

2.2 Absorption Enhancement of Microbolometers in the MWIR

The absorption range of a microbolometer is typically defined by the optical path between a reflective metal on top of the CMOS circuitry and an absorption layer within the microbolometer membrane, which are together arranged as a $\lambda/4$ resonator. This concept has proven itself for many decades of microbolometer designs for uncooled thermal imaging in the LWIR.

The idea of a resonator based microbolometer was adapted here to shift the maximum of the absorbed thermal radiation from the LWIR in the MWIR. For this purpose, the design of the microbolometer was modified in that way that a reduced optical path between reflective area and absorption layer was successfully achieved without increasing the risk of a thermal short circuit between microbolometer and substrate.

3. PERFORMANE OF UNCOOLED THERMAL MWIR IMAGERS

As a measure of the electro-optical performance of our uncooled MWIR imagers, we consider their noise equivalent temperature difference (NETD), spectral sensitivity and thermal time constant. Although MWIR imagers do not have their specified maximum sensitivity at the standard LWIR target wavelength at 10 µm, they are nevertheless analogously evaluated with respect to LWIR imagers in terms of their NETD measured close to standard room temperature at 293 K.[10, 11] However, to be able to evaluate the imaging properties of our MWIR imager to the full extent, we combine these standard NETD measurements with the imagers' spectral sensitivity,

which is measured in a test setup based on a high-temperature black body radiator. These main performance parameters will be supplemented by measurements of the thermal time constant.

The performance parameters mentioned above are evaluated with Fraunhofer IMS's QVGA ROIC.[12] The MWIR lids (compare section 2.1) are finally integrated in our vacuum chip-scale-package (CSP) to combine the adapted microbolometers and lids in a full-featured uncooled MWIR infrared focal plane array (IRFPA).

To illustrate the conceptional functionality of our MWIR imager, we present and compare three variants of our infrared imagers based on our standard LWIR IRFPA (17 µm @ QVGA):

1. LWIR imager (LWIR lid and LWIR-optimized microbolometers)

2. MWIR-like imager (MWIR lid and LWIR-optimized microbolometers)

3. MWIR imager (MWIR lid and MWIR-optimized microbolometers).

As an example, a MWIR variant with MWIR lid (left side) and a LWIR imager (right side) are shown in Fig. 2.
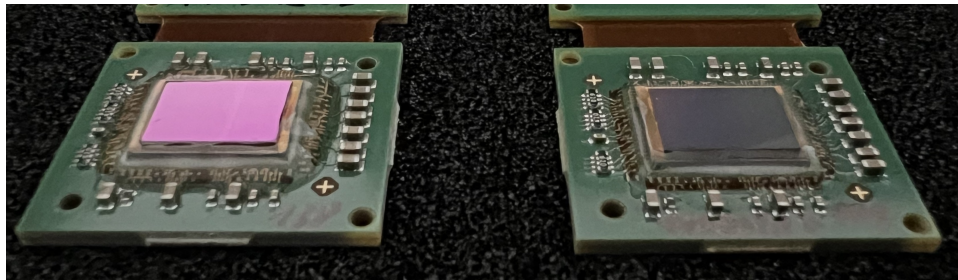
Figure 2. Picture of used uncooled thermal imagers. Left: Imager with MWIR lid. Right: Imager with LWIR lid.

3.1 NETD

The NETD is determined by means of an electro-optical characterization based on a test setup with two black body radiators at different temperatures. Black body temperatures are set to 293.15 K and 308.15 K respectively, thus resulting in a black body temperature difference of $\Delta T_{\mathrm{BB}} = 15$ K. Measurements are performed with $f/1$ optics at 30 Hz. The NETD is a measure of the responsivity $\Re = \Delta D_{\mathrm{out}}/\Delta T_{\mathrm{BB}}$ with ΔD_{out} being the difference of the digital output values at the two different black body temperatures, but additionally considers the total noise of both the microbolometers itself and the ROIC $\langle D_{\mathrm{out}} \rangle$. Thus, NETD is used as the key quantity to evaluate the sensitivity of a thermal imager and is given by Eq. 1:

$$\mathrm{NETD} = \frac{\sqrt{\langle D_{\mathrm{out}} \rangle^2}}{\Re}. \tag{1}$$

Fig. 3 shows the measured NETD values for all three types of investigated imagers composed of different types of microbolometers and lids (compare section 3).
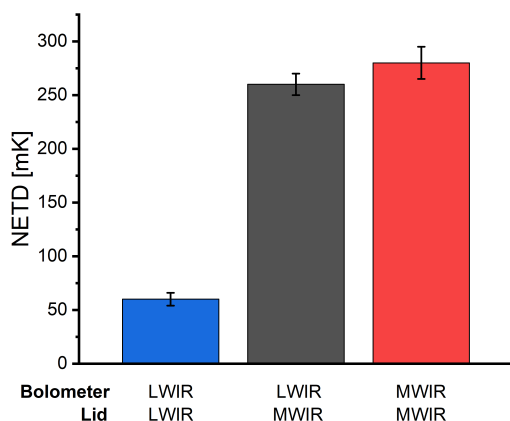
Figure 3. NETD of LWIR and MWIR uncooled thermal imagers measured at 293 K.

While our standard IRFPA for uncooled thermal imaging (composed of LWIR microbolometers and LWIR lid) reveals as expected a low NETD of around 50 mK, the use of our MWIR lid increases the NETD by a factor of ∼ 5. This effect is mainly stemming from the highly reduced transmission of the MWIR lid in the LWIR regime (compare Fig. 1 (b)). On the other hand, the actual microbolometers do not show a significant difference regarding their thermal sensitivity in the LWIR. So even our MWIR-optimized IRFPAs composed of the MWIR microbolometers and the MWIR lid have a certain sensitivity in the LWIR regime, which is mainly limited by the transmission characteristics of the lid.

3.2 Spectral Sensitivity

Spectrally resolved responsivity measurements were carried out to further compare the electro-optical response of the imagers. A high temperature cavity black body set to 800 K was used as radiation source. A series of bandpass filters with center wavelengths (CWL) from 3.5 μm to 10.5 μm and a bandwidth of 0.5 μm completed the setup. The results are shown in Fig. 4(a). The steep drop for responsivities beyond 6 μm is due to the use of a MWIR camera lens, whose transmission is strongly reduced for higher wavelengths and as a reason for this has a limiting effect here. However, the three different types of imagers can be compared qualitatively in terms of their performance in the MWIR range.

According to Plancks law, the spectral exitance $M_e(\lambda)$ of a black body is a function of the corresponding wavelength. Thus, the irradiance received by the device under test differs from filter to filter. In order to account for this effect, the relative irradiance E_r at 800 K was calculated for each filter using empirical data of their respective transmittance $T(\lambda)$:

$$E_r = \frac{\int M_e(\lambda) T(\lambda) d\lambda}{\int M_e(\lambda) d\lambda}. \tag{2}$$

The results weighted by E_r are presented in Fig. 4(b). The maximum sensitivity of the imager adapted for the MWIR is at about 4.5 μm which is in good agreement with the transmission spectra of the lid shown in Fig. 1(b).

(a) Responsivity.

(b) Spectral sensitivity.
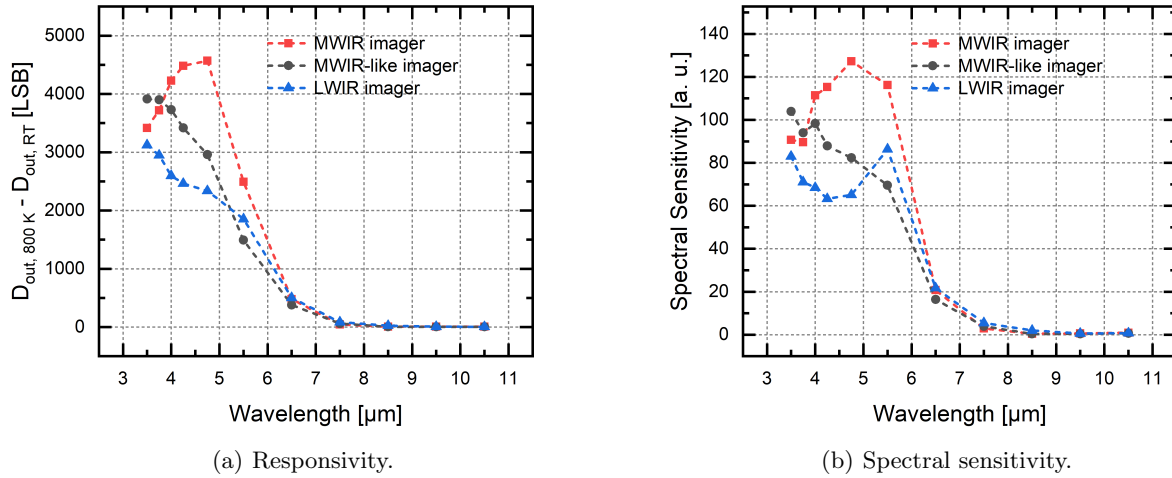
Figure 4. Spectral sensitivity of LWIR and MWIR imagers between 3.5 µm and 10.5 µm.

Fig. 5 compares the spectral sensitivity of both MWIR imager variants to our standard LWIR imager within our region of main interest. Due to the steep decrease of the used MWIR camera lens' transmission, only statements of the relative performance with respect to each other can be made. However, a comprehensive analysis of the full spectral range including both MWIR and LWIR is not possible with the used test setup. Nevertheless, it is demonstrated that the performance of the MWIR imager exceeds the sensitivity of the LWIR imager by a factor of 2 at our MWIR center wavelength of about 4.5 µm.
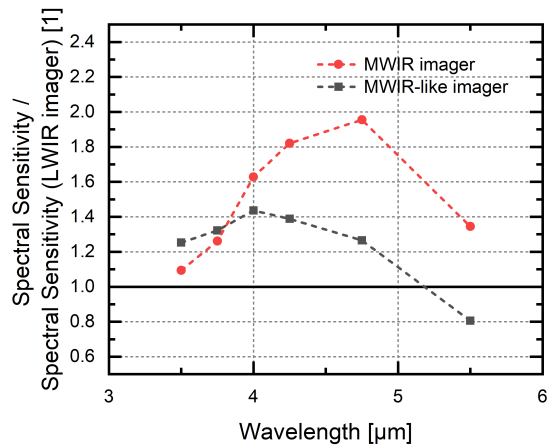
Figure 5. Spectral sensitivity of MWIR imagers relatively compared to our LWIR imager between 3.5 µm and 5.5 µm.

3.3 Thermal Time Constant

The thermal time constant is determined by measuring the required time for a full signal rise and signal drop, respectively, of the digital output values D_{out} towards a stable plateau in a dynamic scene with two defined temperature levels by using a rotating chopper wheel. The frequency of the chopper wheel is set to 4 Hz to ensure saturation of the digital output values D_{out} at both the lower and the upper temperature level, corresponding to the blades of the chopper wheel and the surface of the black body radiator, respectively. Subsequently, the thermal time constant τ is estimated by an exponential fit following $\sim e^{-t/\tau}$ at the rising or falling edge of the time-resolved signal of the digital output values D_{out} for a selected part of the microbolometers in the middle of the QVGA imager.
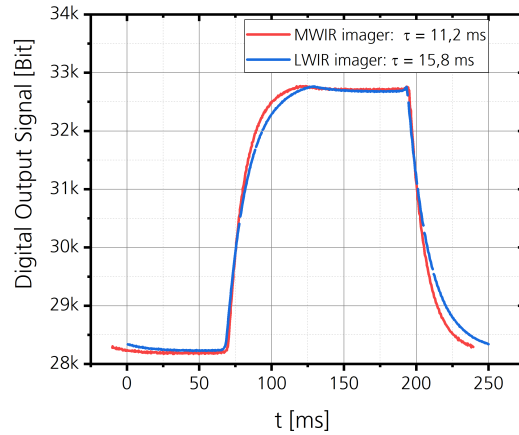
Figure 6. Mesurement of thermal time constant of uncooled LWIR and MWIR imagers.

Fig. 6 depicts both the measurement results for our standard LWIR imager and its modified MWIR version. The imagers discussed here each use the microbolometers optimized for their spectral range and the corresponding lid following the definitions from section 3. Due to our conceptional adjustments to shift the resonator condition towards the MWIR range, a decrease in thermal capacitance resulting in a lower thermal time constant could be achieved. Thus, the thermal time constant of the MWIR imager was reduced by nearly 30 % as compared to the LWIR version.

3.4 MWIR Imaging Sceneries

To demonstrate full functionality and discriminability of both imaging technologies, two test sceneries were captured by our LWIR and MWIR imager respectively. Fig. 7 depicts a lighter as a hot infrared emission source to illustrate the different parts of the emitted spectrum covering both LWIR and MWIR.

(a) Lighter in hand taken by LWIR imager.

(b) Lighter in hand taken by MWIR imager.

(c) Lighter behind silica glass taken by LWIR imager.

(d) Lighter behind silica glass taken by MWIR imager.

Figure 7. Images of two test sceneries with a lighter plus flame, either taken by the LWIR or the MWIR imager.

As expected, the LWIR images in Fig. 7 impress with a sharp visualization of the complete scenery, whereas the MWIR images can only resolve hot areas of the images in high contrast. This observation correlates with the measured NETD values (compare section 3.1). In addition, the setup with the silica glass in front of the lighter is a convenient example to demonstrate silica glass appearing transparent to the MWIR portions of the spectrum emitted by the flame whereas silica glass is opaque for LWIR parts of the infrared spectrum. With the help of this effect, an extended scene contrast can be achieved when combining both images.

4. SUMMARY AND OUTLOOK

In this publication, we presented Fraunhofer IMS's first measurement results of uncooled thermal imagers optimized for the MWIR regime ranging from 3 μm to 5 μm. For this purpose, both the lid of the vacuum packaging as well as the resonating structure of the microbolometers have been accordingly adjusted. Combining these effects, we were able to demonstrate a spectral sensitivity which is twice as high at a wavelength of around 4.5 μm as compared to our LWIR imagers featuring a NETD of $\sim$ 55 mK. Differences in the visualization of the two spectral ranges could be highlighted by means of two specific test sceneries, which at the same time allowed to successfully demonstrate the functionality of the new developed MWIR imagers.

Our developed MWIR lid increases the transmission in the MWIR range for about 20 - 40 % in comparison to our established LWIR version. The use of a multilayer coating composed of Si_3N_4 and SiO_2 allows to effectively tune the absorption maximum to the desired target wavelength. The accompanying reduction of the transmission in the LWIR is mainly caused by the strong absorption of Si-O bonds within the silicon dioxide between 8 μm to 10 μm. This effect explains the increase of the NETD which is measured close to room temperature resp. in the LWIR.

In addition to the optimized lid, we were able to demonstrate a further beneficial impact on the mid wave spectral sensitivity by adjusting the microbolometers' resonating absorption condition. Furthermore, due to these improvements the thermal time constant could be reduced by nearly 30 % as compared to the LWIR imagers to around 11 ms.

The high values of the measured NETD of our uncooled MWIR IRFPA mainly stem from a non-optimized transmission of the lid in the LWIR regime as discussed in detail in the previous sections. However, if solely focusing on high temperature imaging applications for black body temperatures in the range of 800 K or above, this limitation does not come into play enabling already now a wide field of use case scenarios as fire or gas detection. If uncooled thermal MWIR imaging should be expanded to black body temperatures in the range of 300 K, the lid has consequently to be adapted to be reliably transparent over a broad regime in the IR covering both MWIR and LWIR. When optimizing the transmission of the lid in the LWIR, focus of our future research will be put on the analysis of the discrepancy between simulation and FTIR measurement enabling room for an improved transmission up to a factor of 2.5.

REFERENCES

[1] Breiter, R., Cabanski, W., Mauk, K.-H., Rode, W., Ziegler, J., and et al., "Multicolor and dual-band ir camera for missile warning and automatic target recognition," *Targets and Backgrounds VIII: Characterization and Representation* **Proc. SPIE 4718**, 280–288 (2002).

[2] Turgut, B. B., Artan, G. G., and Bek, A., "A comparison of mwir and lwir imaging systems with regard to range performance," *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIX* **Proc. SPIE 10625**, 1062512 (2018).

[3] Linares, R., Vergara, G., Gutirrez, R., Fernndez, C., Villamayor, V., and et al., "Gas and flame detection and identification using uncooled mwir imaging sensors," *Thermosense: Thermal Infrared Applications* **Proc. SPIE 9485**, 94851F (2015).

[4] Olbrycht, R. and Kaluza, M., "Optical gas imaging with uncooled thermal imaging camera - impact of warm filters and elevated background temperature," *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS* **VOL. 67, NO. 11**, 9824–9832 (2020).

[5] Swamidoss, I. N., Bin Amro, A., and Sayadi, S., "Systematic approach for thermal imaging camera calibration for machine vision applications," *Optik* **247**, 168039 (2021).

[6] Jung, S.-M. and et al., "Design and fabrication of multi-layer antireflection coating for iii-v solar cell.," *Current Applied Physics* **11.3**, 538–541 (2011).

[7] Weng, B. and et al., "Responsivity enhancement of mid-infrared pbse detectors using caf2 nano-structured antireflective coatings," *Applied Physics Letters* **104.2**, 021109 (2014).

[8] Luke, K., Okawachi, Y., Lamont, M. R. E., and et al., "Broadband mid-infrared frequency comb generation in a si3n4 microresonator," *Optics Letters* **Vol. 40, No. 21**, 4823–4826 (2015).

[9] Kischkat, J., Peters, S., Gruska, B., and et al., "Mid-infrared optical properties of thin films of aluminum oxide, titanium dioxide, silicon dioxide, aluminum nitride, and silicon nitride," *APPLIED OPTICS* **Vol. 51, No. 28**, 6789–6798 (2012).

[10] Lee, H., Eom, J., Kang, K., Ryu, S., Jang, A., and et al., "640x512 dual-band midwave and longwave infrared focal plane array at i3system," *Infrared Technology and Applications XLIX* **Proc. SPIE 12534**, 125340L (2023).

[11] Demirer, N., Agca, G., Kaldirim, M., Yalinkilic, M., Akca, E., and et al., "Development of mwir t2sl photodetectors at aselsan inc.," *Infrared Technology and Applications XLIX* **Proc. SPIE 12534**, 125340I (2023).

[12] Weiler, D., Hochschulz, F., Wuerfel, D., Lerch, R., Geruschke, T., Wall, S., Hess, J., Wang, Q., and Vogt, H., "Uncooled digital irfpa-family with 17um pixel-pitch based on amorphous silicon with massively parallel sigma-delta-adc readout," *Infrared Technology and Applications XL* **Proc. SPIE 9070**, 90701M (2014).

Mixed oxide provides an efficient solution to near infrared detection

Antoine Cheneau[a,b], Annick F. Dégardin*[a,b], Alain J. Kreisler[a,b]

[a]Université Paris-Saclay, CentraleSupélec, CNRS, Laboratoire de Génie Électrique et Électronique de Paris, 91190 Gif-sur-Yvette, France; [b]Sorbonne Université, CNRS, Laboratoire de Génie Électrique et Électronique de Paris, 75005 Paris, France

ABSTRACT

With the objective of producing high-performance infrared radiation detectors, we have undertaken the study of devices based on the Y-Ba-Cu-O material produced in amorphous film (a-YBCO) by DC sputtering at low temperature (< 150 °C) on p-doped silicon substrates coated with a SiO_x oxide film. Two types of structures have been considered: simple planar structures, where a-YBCO is connected to in-plane metal contacts, and trilayer structures where a-YBCO is sandwiched between the two contacts. The near-infrared response was recorded with a laser source at 850 nm wavelength, amplitude modulated up to 40 MHz. The main characteristics of the responses are: i) a f^{+2} behavior at very low frequency (resulting from two dipolar relaxations); ii) a typically pyroelectric behavior in f^{+1} up to a few tens of kHz; iii) a maximum response followed by a decrease in $f^{-1/2}$, reflecting the heat diffusion through the substrate. All of those results could be interpreted with the help of a theoretical model with adjustable parameters. Small size trilayer devices exhibit a very fast response (time constant: 11 ns). Besides, with noise equivalent power levels as low as 8 pW/Hz$^{1/2}$ and detectivity values of 8×10^8 cm·Hz$^{1/2}$·W^{-1} at 1 MHz, our uncooled devices are standing at the state-of-the-art both in terms of sensitivity and bandwidth.

Keywords: Y-Ba-Cu-O semiconductor, amorphous thin films, pyroelectric detectors, uncooled near-infrared detectors, pyroelectric model, fast pyroelectric response, noise equivalent power, detectivity

1. INTRODUCTION

Pyroelectric detectors are widely exploited in the field of infrared (IR) radiation uncooled sensors. A pyroelectric detector is a thermal detector based on a capacitive structure. By definition, a thermal detector which is wavelength independent exhibits a slow response, typically in the millisecond to few tens of microsecond range. The room temperature operation of such a detector allows a reduction in cost, size, and complexity. In view of miniaturization, dielectric material thin films exhibiting complementary metal oxide semiconductor (CMOS) compatibility for integrating the devices with their readout electronics have been largely investigated: lead zirconate titanate (PZT)[1], lithium tantalate (LiTaO$_3$)[2], lithium niobiate (LiNbO$_3$)[3], zinc oxide (ZnO)[4], aluminium nitride (AlN)[5], hafnium oxide (HfO$_2$)[6], yttrium barium copper oxide (YBCO)[7,8].

The latter compound, YBa$_2$Cu$_3$O$_{6+x}$, is well-known as a high critical temperature superconductor, for oxygen doping x above ~ 0.5. YBCO exhibits also a semiconducting phase when its oxygen content is low (i.e., $x < 0.5$). Semiconducting YBCO can be prepared at low temperature, typically 150 °C, in the form of amorphous films (hereafter called a-YBCO) under conditions that are fully compatible with silicon integration of CMOS readout circuitry[9]. Moreover, a-YBCO exhibits a high pyroelectric coefficient, a high absorption coefficient, and a low noise level, making this material an attractive solution for near-infrared (NIR) uncooled sensing[8].

In this paper, two types of detector geometries, fabricated on silicon substrates, have been investigated: planar devices, where the a-YBCO sensing film is deposited on the top of two metal contacts, and trilayer devices where the a-YBCO sensing film is sandwiched between two metal contacts. This latter geometry allows to decrease the detector access resistance, a possible advantage for preamplifier impedance matching. In section 2 of this paper, we shall detail the manufacturing steps and describe the experimental set-up for the device test at 850 nm wavelength. In section 3, we shall focus on the optical responses of these devices, their modelling in view of understanding the unconventional features of their response in relation with the detector geometry. Noise performances will be also considered before concluding in section 4.

*annick.degardin@centralesupelec.fr; phone 0033 1 6985 1653; https://annfrance100.wixsite.com/website

2. EXPERIMENTAL DETAILS

2.1 Thin film preparation

a-YBCO thin films were deposited by off-axis direct current (DC) sputtering under a 33 Pa atmosphere of oxygen (45%) and argon (55%), using a two-inch cylindrical hollow target made of the superconducting $YBa_2Cu_3O_{6.9}$ phase (Hitec Materials GmbH). The deposition temperature was about 150 °C and no further oxygenation step was performed; hence the so-obtained films were amorphous and exhibited the semiconducting deoxygenated YBCO phase[9]. The substrates were 380 μm thick p-doped silicon wafers coated with a thermally grown 500 nm thick SiO_x layer. The deposition rate was 150 nm/hour, typically. The film surface was of low roughness (i.e., 3 to 5 nm rms) and the a-YBCO material conductivity was $\sigma_{DC} = 1.7$ mS/cm at room temperature[10].

2.2 Device structures

We fabricated planar and trilayer detection structures, sketched in figure 1. The interest of studying both structures is related to the easiness of fabrication (case of planar devices) or to a small DC access resistance R_{DC} (case of trilayer devices).

For both structures, the fabrication process started with a gold / titanium lower metallization level, e-beam evaporated on the SiO_x//p-doped Si substrate. Then the a-YBCO film formed the second technological level. Film thicknesses varied from 270 to 500 nm. The a-YBCO film was patterned using standard optical lithography to define the active area. The planar device (figure 1, left) has the advantage of its fabrication simplicity. However, the distance between contact pads (represented by the ℓ values) ranged in the tens or hundreds of μm: so R_{DC} values were high (a few MΩ), which can present further difficulties for low-noise readout electronics optimization. The trilayer device (figure 1, right) contains a gold / titanium upper metallization level over the a-YBCO sensing film. This type of device offers the advantage to exhibit an access resistance of a few hundred ohms only, mainly because the ℓ value is of a few hundred nm, which corresponds to the film thickness in this geometry. Some physical and geometrical parameters of the a-YBCO devices studied hereafter are gathered in table 1.
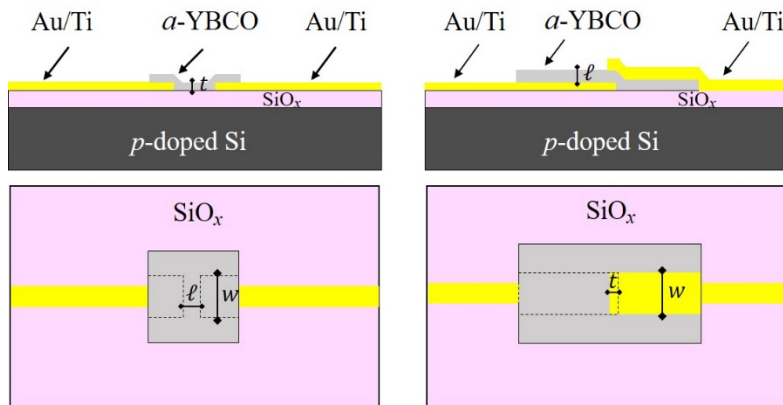
Figure 1. Schematic views of the two a-YBCO-based detector geometries: (left) planar and (right) trilayer. Top: cross sections; bottom: top views. The DC access resistance is given by the well-known formula: $R_{DC} = \rho\,\ell/wt$, where ρ is the a-YBCO material resistivity, and ℓ, w and t are the geometrical dimensions of the a-YBCO sensing element. More specifically, ℓ is the distance between contact pads.

2.3 NIR test bench

A schematic of the NIR experimental set-up is shown in figure 2. The output beam intensity of a VCSEL-type laser source emitting at wavelength $\lambda = 850$ nm was electronically modulated at a frequency f. This output beam was collimated and focused onto the device inserted in a cryostat, which insured efficient shielding, while keeping an operation temperature of ≈ 290 K. The pyroelectric current i_p provided by the unbiased device as a response to the modulation frequency was readout with a low noise transconductance preamplifier, followed by a synchronous detector (lock-in amplifier). The modulation frequency varied from a few Hz to 40 MHz (limit due to the cutoff of our high frequency transconductance preamplifier).

Table 1. Types of detector devices.

Sample #	Type	Film thickness (nm)	Capacitor electrode area (μm²)	$1/\ell^2$ (μm⁻²)	R_{DC} (Ω)
Pl-1	Planar	500 ± 10	250	6.3×10^{-6}	9.4×10^6
Pl-2	Planar	450 ± 10	45	1.1×10^{-3}	3.9×10^6
Tr-1	Trilayer	390 ± 10	4000	6.6	580
Tr-2	Trilayer	270 ± 10	4000	13.7	397
Tr-3	Trilayer	270 ± 10	2000	13.7	794
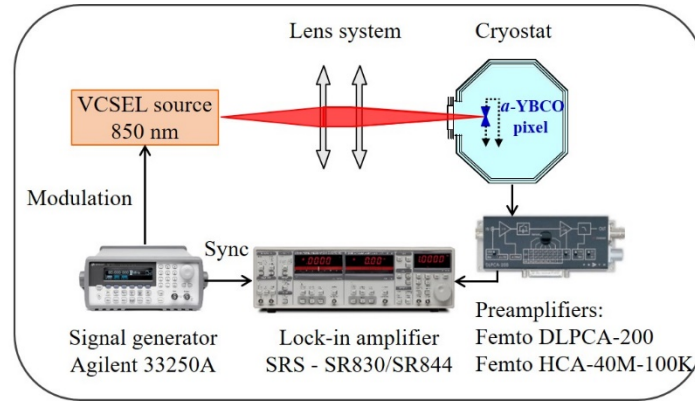
Figure 2. Experimental set-up for NIR tests.

The same lock-in amplifier was also used to measure the noise current density i_n (expressed in A/Hz$^{0.5}$). Due to the noise statistical nature, subsequent adequate statistics treatment was undertaken, as described previously elsewhere[8]. From noise measurements, we extracted the noise equivalent power NEP (in a 1 Hz bandwidth) and specific detectivity D^* values of the devices, according to:

$$NEP = i_n/\mathcal{R}_i, \tag{1}$$

$$D^* = \sqrt{A}/NEP, \tag{2}$$

where $\mathcal{R}_i = i_p/\mathcal{P}_{inc}$ is the device current responsivity (expressed in A/W), with $\mathcal{P}_{inc}$ the effective incident power, and A is the effective pixel area which collects the NIR radiation. $\mathcal{P}_{inc}$ was carefully deduced from a detailed geometrical description of the sample arrangement.

3. NIR PERFORMANCES OF *a*-YBCO THIN FILM DEVICES

3.1 Optical response

Figure 3 exhibits the current output responses of *a*-YBCO planar and trilayer devices, for which similar band-pass type trends can be observed, as for traditional pyroelectric detectors. Indeed, a high-pass behavior is observed at low frequency, which is typical of a pyroelectric response (capacitive effect). However, an unconventional f^2 slope can be identified at very low frequency. A maximum current response is then reached, at a lower frequency ($\approx$ 30 kHz) for the planar devices than for the trilayer ones ($\approx$ 200 kHz): these frequency values are higher than those for devices based on regular pyroelectric materials, which are considered as slow. Finally, an unconventional low-pass behavior close to $f^{-1/2}$ can be observed at higher frequencies, which is typical of heat diffusion through the substrate. These behaviors are discussed in the following subsection.
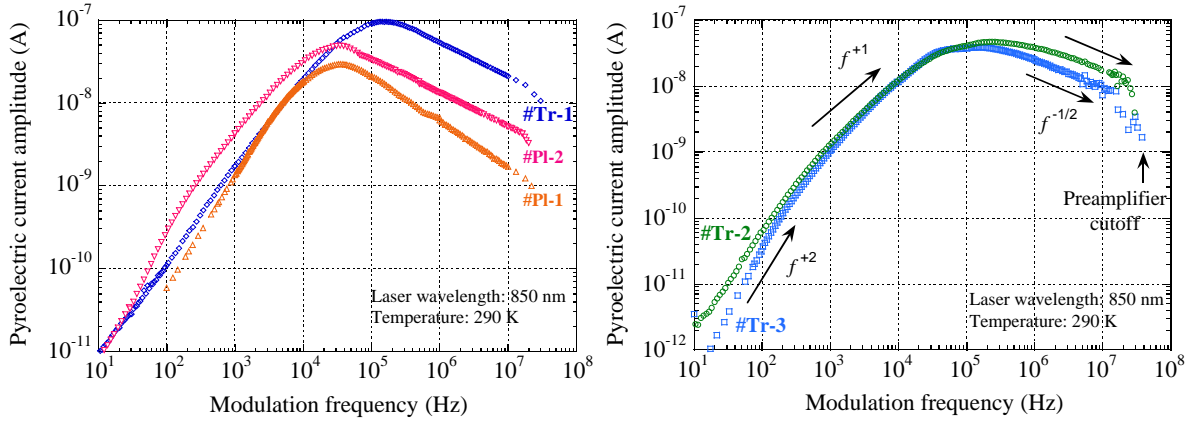
Figure 3. Experimental pyroelectric current as a function of modulation frequency, for a-YBCO based sensing devices: (left) two planar devices (orange and pink curves) and one trilayer device (blue curve); (right) two trilayer devices, with slope values indicated.

3.2 Model and interpretation

3.2.1 Regular pyroelectric model

Before analyzing in detail the results obtained with our a-YBCO based devices, let us consider first the modeling of a regular pyroelectric detector. Figure 4 shows the basic sketches concerning such a regular detector type.
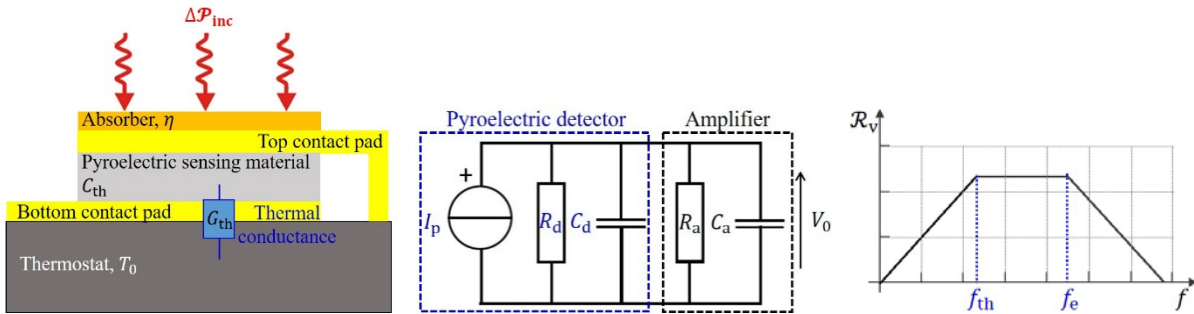
Figure 4. Regular pyroelectric detector. (left) Sketch of a pyroelectric structure based on a thin film capacitor, including an absorbing layer. (center) Diagram of equivalent electrical circuit for the pyroelectric detector combined with its amplifying circuit. (right) Sketch of the voltage responsivity as a function of modulation frequency. See text for details about quantities mentioned in the sketches.

The response in the harmonic regime of a regular pyroelectric detector to a sinusoidally amplitude modulated power is determined in three steps.

i) The incident power variation $\Delta\mathcal{P}_{\text{inc}}$ absorbed by the pyroelectric material implies a material temperature variation ΔT, which can be expressed as follows:

$$\frac{\Delta T}{\Delta\mathcal{P}_{\text{inc}}} = \frac{\eta}{G_{\text{th}}(1+j\omega/\omega_{\text{th}})}, \tag{3}$$

where η is the absorption coefficient, representing the fraction of power actually absorbed by the sensing material, G_{th} is the thermal conductance between the sensing area and the thermostat (at $T_0 = 290$ K), ω is the angular frequency, and $\omega_{\text{th}} = G_{\text{th}}/C_{\text{th}} = 1/\tau_{\text{th}}$ is the thermal cutoff angular frequency, with $\tau_{\text{th}} = C_{\text{th}}/G_{\text{th}}$ the thermal time constant and C_{th} the thermal capacitance of the sensing volume. This corresponds to the behavior of a first order low-pass filter with a thermal cutoff frequency defined as $f_{\text{th}} = \omega_{\text{th}}/2\pi$.

ii) The pyroelectric property of the sensing material implies that a temperature variation ΔT leads to a polarization variation $\Delta P = p\Delta T$, with p the pyroelectric coefficient of the material; ΔP is homogeneous to a variation of polarization charges ΔQ per unit area, hence the occurrence of a pyroelectric current I_p, as follows:

$$I_p(\omega) = j\omega A\Delta P = j\omega Ap\Delta T, \tag{4}$$

where A is the effective area of the pyroelectric detector. Equations (3) and (4) lead to the pyroelectric current response:

$$\frac{I_p(\omega)}{\Delta\mathcal{P}_{inc}} = \frac{Ap\eta}{C_{th}}\frac{j\omega/\omega_{th}}{(1+j\omega/\omega_{th})}, \tag{5}$$

which behaves like a first order high-pass filter with the same thermal cutoff frequency ω_{th}, as defined in step i).

iii) Finally, the pyroelectric current I_p is converted into an output voltage V_0: using the diagram of figure 4 (center), where R_d (respectively R_a) is the resistance and C_d (respectively C_a) is the capacitance of the detector (respectively amplifier), we can deduce the resistance and the capacitance of the equivalent electrical circuit, as: $R_e = R_dR_a/(R_d + R_a)$ and $C_e = C_d + C_a$; hence the output voltage:

$$V_0 = I_pR_e/(1 + j\omega/\omega_e), \tag{6}$$

where $\omega_e = 1/\tau_e = 1/R_eC_e$ is the electrical angular frequency, with $\tau_e = R_eC_e$ the electrical time constant. According to equations (5) and (6), the voltage responsivity $\mathcal{R}_v = |V_0/\Delta\mathcal{P}_{inc}|$ of a pyroelectric detector can be expressed, as follows:

$$\mathcal{R}_v = \frac{Ap\eta R_e}{C_{th}}\frac{\omega\tau_{th}}{\sqrt{1+(\omega\tau_{th})^2}}\frac{1}{\sqrt{1+(\omega\tau_e)^2}}. \tag{7}$$

As shown in figure 4 (right), the frequency response of a regular pyroelectric detector is that of a bandpass filter with a low cutoff frequency f_{th} of thermal origin and a high cutoff frequency $f_e = 1/2\pi\tau_e$ associated with the amplifying electrical circuit. f_{th} values lie in the 0.1 to 10 Hz range and f_e values lie at about tens of kHz, typically.

This regular pyroelectric model could explain the band-pass type behavior of a-YBCO pyroelectric detectors developed by Butler *et al.*[7] but obviously cannot explain the variations in slope of the current amplitude responses for our a-YBCO pyroelectric detectors: e.g., f^{+2} behavior at very low frequencies or $f^{-1/2}$ behavior at high frequencies. Moreover, the low cutoff frequency observed for our pyroelectric detectors cannot be associated with the thermal cutoff frequency usually expected in the model of regular pyroelectric detectors. These inconsistencies led us to make substantial changes to the regular pyroelectric modeling, as explained in the following subsection.

3.2.2   Advanced pyroelectric model

The observed wide bandwidth response, together with the knowledge of the a-YBCO dielectric properties and the specific sensor geometry, led us to reconsider completely the orders of magnitude, as follows[11].

i) The slope variations in the current response at low frequencies (a f^{+2} behavior followed by a f^{+1} behavior) tend to show that the pyroelectric coefficient p assumed to be constant in the regular pyroelectric model could in fact depend on frequency. We have linked this frequency dependence to the dielectric measurements of a-YBCO thin films, which exhibited two Debye's dipolar relaxation frequencies: a low relaxation frequency $f_{dip} \approx 200$ Hz and a high relaxation frequency $f'_{dip} \approx 20$ kHz, observed at 300 K[12]. Thus, the pyroelectric current becomes:

$$I_p(\omega) = j\omega Ap\frac{j\omega/\omega_{dip}}{1 + j\omega/\omega_{dip}} \cdot \frac{j\omega/\omega'_{dip}}{1 + j\omega/\omega'_{dip}}\Delta T, \tag{8}$$

with $\omega_{dip} = 2\pi f_{dip}$ and $\omega'_{dip} = 2\pi f'_{dip}$. We therefore expect to observe a f^{+2} slope at frequencies below f_{dip}, then a f^{+1} slope between f_{dip} and f'_{dip}, followed by an horizontal plateau beyond f'_{dip}.

ii) Likewise, the thermal flux can no longer be described by means of a constant (e.g. interface) conductance G_{th}, but through a frequency dependent heat diffusion process across the substrate thickness. Following the model detailed in Kreisler *et al.*[11], the thermal contribution $1/G_{th}(1 + j\omega/\omega_{th})$ in (3) should be replaced by:

$$1/G_{th0}\left[1 + (\omega/\omega_{dif})^{1/2} + j(\omega/\omega_{dif})^{1/2} + j(\omega/\omega'_{th})\right], \tag{9}$$

where G_{th0} is a constant[11], $\omega_{dif} = 2\pi f_{dif}$ is the diffusion cutoff angular frequency and $\omega'_{th} = G_{th0}/C_{th} = 2\pi f'_{th}$ is the new thermal cutoff angular frequency.

iii) Finally, when combining the previous equations (8) and (9), we obtain the current responsivity $\mathcal{R}_i = |I_p/\Delta\mathcal{P}_{inc}|$ for our a-YBCO pyroelectric detectors as:

$$\mathcal{R}_i = \frac{Ap\eta}{G_{th0}} \frac{\omega}{\sqrt{\left(1+\sqrt{\omega/\omega_{dif}}\right)^2 + \left(\sqrt{\omega/\omega_{dif}} + \omega/\omega'_{th}\right)^2}} \frac{\omega/\omega_{dip}}{\sqrt{1+\left(\omega/\omega_{dip}\right)^2}} \frac{\omega/\omega'_{dip}}{\sqrt{1+\left(\omega/\omega'_{dip}\right)^2}} |R_{el}(\omega)|, \qquad (10)$$

where $R_{el}(\omega)$ is the low-pass transfer function associated with the electronic readout circuitry.

3.2.3 Discussion

Measurements and modeling curves are depicted for two trilayer devices in figure 5. The parameters used to fit the model response curves for all the samples are given in table 2. Several points related to the comparison between experimental and modeling curves are discussed hereafter.

i) Low frequency response – planar devices: two dipolar relaxation cutoff frequencies can be well identified, which values are in line with previous values obtained from a-YBCO film dielectric measurements using the Debye model namely, 200 Hz and 20 kHz, as previously mentioned[12]. The f^{+1} behavior results from a dielectric dipolar relaxation of a-YBCO at $f'_{dip} \approx 20$ kHz, whereas the slope in f^{+2} at lower frequencies results from its combination with another dipolar relaxation at $f_{dip} \approx 200$ Hz.

ii) Low frequency response – trilayer devices: the lower frequency relaxation at $f_{dip} \approx 200$ Hz can be still identified, whereas the high frequency relaxation f'_{dip} is rather located in the 50 to 80 kHz range, as opposed to the 20 kHz value deduced from dielectric experiments.

iii) High frequency response – both device types: as heat from the incident radiation is first absorbed before diffusing through the substrate, it is logical to find a diffusion cutoff frequency f_{dif} lower than the thermal cutoff frequency f'_{th}. Besides, the impact of the device size on the frequency response should be also discussed, as introduced below (see next point). In fact, the trilayer devices exhibit much higher f_{dif} values (than for the planar ones), hence even larger f'_{th} values, which is in favor of small size and very fast devices.

iv) Dimensional considerations: from the device capacitance expression $C = \varepsilon_0\varepsilon_r S/\ell$, it follows that the larger C is, the larger the stored electric dipolar charge is, this latter being related to the whole volume V of the capacitor. As $C/V \propto 1/\ell^2$, the dominant parameter $1/\ell^2$ should be maximized with large facing surfaces of area S; the resulting best choice is therefore the small device #Tr-2 (see table 1), which also offers the advantage of a small access resistance, so enhancing the easiness of impedance matching with readout.

v) Frequency response consideration: besides, for device #Tr-2, our fastest device, the thermal cutoff frequency might be in fact higher than 15 MHz as our measurements, limited by the 40 MHz cutoff frequency of the preamplifier, were somewhat noisy in this frequency range.

Table 2. Cutoff frequency values extracted from response model fitting as well as the computed value of the thermal time constant.

Device #	f_{dip} (Hz)	f'_{dip} (kHz)	f_{dif} (kHz)	f'_{th} (MHz)	τ'_{th} (ns)
Pl-1	300	27	50	0.1	1600
Pl-2	130	20	90	2	80
Tr-1	250	80	650	3	53
Tr-2	150	50	3400	15	11
Tr-3	200	70	650	10	16
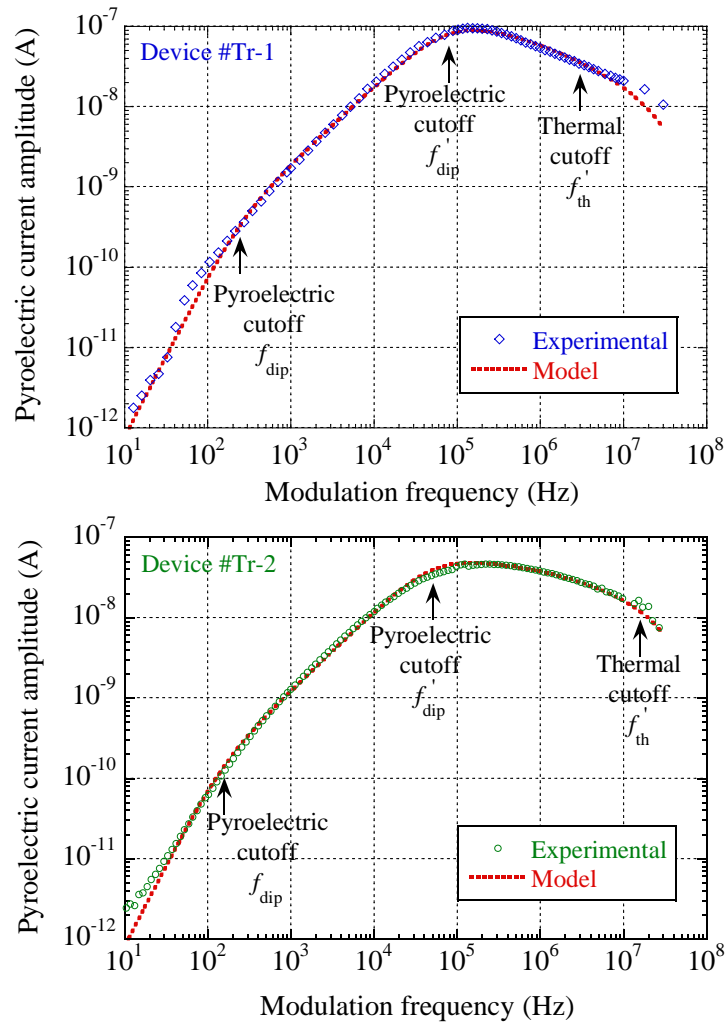
Figure 5. For *a*-YBCO trilayer devices, current amplitude responses as a function of the modulation frequency. The measurements (open symbols) are compared to analytical simulations (red dotted lines). For the model the values were adjusted according to the maximum amplitude. Characteristic frequencies are defined in the text.
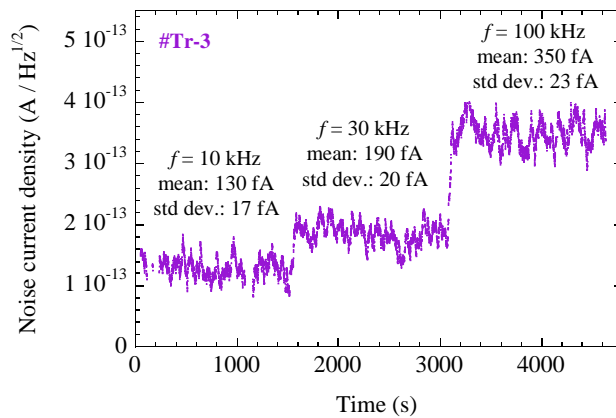
Figure 6. For an *a*-YBCO trilayer device, noise (dark) current spectral density as a function of elapsed time: measurements performed in a 1 Hz bandwidth at three modulation frequencies are here considered, for which the mean and standard deviation values of noise current were computed, as indicated in the graph.

3.3 Detector performance evaluation

Noise values were recorded, as shown in figure 6 for a trilayer device, from which noise equivalent power *NEP* and specific detectivity D^* values were extracted (see table 3).

From table 3, device #Pl-1, which sensing volume is 74 times larger than that of device # Pl-2, exhibits *NEP* values of more than three orders of magnitude higher than those of device #Pl-2; moreover, device #Pl-1 is very slow. These arguments are in favor of realizing small size devices. Device #Pl-2 and device #Tr-1 exhibit similar sensing volume: *NEP* values of device #Pl-2 are between two and eight times higher than those of device #Tr-1. The values of their thermal time constants lie from 80 ns (#Pl-2) to 53 ns (#Tr-1), which is a significant performance improvement. Device #Tr-1 exhibits also a large bandwidth over several hundreds of kHz. These arguments are in favor of realizing trilayer devices, although an optimized small size planar device might also provide correct performances with the advantage of easiness of fabrication.

Considering trilayer devices only, all three devices exhibit a large bandwidth, with a value peaking above 1 MHz for device #Tr-2. The *NEP* lies below 6 pW/Hz$^{\frac{1}{2}}$, and D^* peaks above 10^9 cm·Hz$^{1/2}$·W^{-1} at 100 kHz, which is close to the room temperature theoretical limit for uncooled detectors ($\sim 1.8\times10^{10}$ cm·Hz$^{1/2}$·W^{-1})[13]. The very short time constant values (< 16 ns) for devices #Tr-2 and #Tr-3 are also worth noticing.

It is also worth comparing the performance data of our detectors with those of LiTaO$_3$ or PZT-based sensors commercially available or reported in the literature, as gathered in table 4. It clearly appears that our *a*-YBCO-based devices exhibit better detectivity (as checked up to ~ 1 MHz modulation frequency, even above 1 MHz for device #Tr-2, which cannot be reached by regular pyroelectric detectors) and are much faster by several orders of magnitude.

Table 3. NIR performances of our *a*-YBCO uncooled pyroelectric detectors.

Device #	f (Hz)	*NEP* (W·Hz$^{-1/2}$)	D^* (cm·Hz$^{1/2}$·W^{-1})	Bandwidth (kHz)	τ'_{th} (ns)
Pl-1	10^4 10^5	4.6×10^{-9} 8.8×10^{-8}	9.8×10^6 5.1×10^5	86.2	1600
Pl-2	10^4 10^5	2.0×10^{-12} 1.3×10^{-11}	6.6×10^9 1.0×10^9	87.3	80
Tr-1	10^4 10^5	8.9×10^{-13} 1.6×10^{-12}	5.0×10^9 2.9×10^9	546	53
Tr-2	10^5 10^6 10^7	5.2×10^{-12} 7.9×10^{-12} 2.9×10^{-11}	1.2×10^9 8.0×10^8 2.2×10^8	1620	11
Tr-3	10^5 10^6 10^7	1.7×10^{-12} 3.4×10^{-11} 4.3×10^{-10}	2.7×10^9 1.3×10^8 1.1×10^7	734	16

Table 4. NIR performances of uncooled regular pyroelectric detectors published in the literature.

Reference	Material	D^* (cm·Hz$^{1/2}$·W^{-1})	τ_{th} (ms)
Ref. 14	LiTaO$_3$	2.5×10^8	200
Ref. 15	LiTaO$_3$	3.0×10^8	150
Ref. 2	LiTaO$_3$	1.7×10^8	5.8
Ref. 16	PZT	1.5×10^8	2.0

4. CONCLUSION

We have undertaken the study of both planar and trilayer pyroelectric detectors. The devices were fabricated from a-YBCO films DC sputtered on SiO_x//p-doped Si substrates and tested at 850 nm wavelength. Their current response as a function of the modulation frequency was analyzed in detail with the help of an advanced analytical model.

A small size planar device offered correct performances ($\sim$ 80 ns time constant, $D^* = 1 \times 10^9$ cm·Hz$^{1/2}$·W^{-1} at 100 kHz), which could be improved by optimizing its geometry (e.g. distance between the metal contact pads, film thickness). Largely better performances were obtained with trilayer devices: these exhibited a pyroelectric behavior with very fast response ($\sim$ 11 ns time constant) extending up to the tens of MHz range, a low noise level with highly competitive NEP (8 pW/Hz$^{1/2}$ at 1 MHz) and D^* (8×10^8 cm·Hz$^{1/2}$·W^{-1} at 1 MHz) values.

These a-YBCO detectors offer attractive features for applications requiring fast imaging pixel arrays. Moreover, the low deposition temperature of a-YBCO on Si-based substrates offers integration compatibility with already processed CMOS readout circuitry. Further tests are now in progress to explore the response at other wavelengths.

ACKNOWLEDGMENT

REFERENCES

[1] Ahmed, M. and Butler, D. P., "Modified lead titanate thin films for pyroelectric infrared detectors on gold electrodes," Infrared Phys. Technol. 71, 1–9 (2015).

[2] Zhang, K., Luo, W., Zeng, X., Huang, S., Xie, Q., Wan, L., Shuai, Y., Wu, C. and Zhang, W., "Integration and characterization of LiTaO₃ single crystal film pyroelectric sensor using mid-infrared metamaterial perfect absorber," IEEE Sensors J. 22, 10381-10387 (2022).

[3] Luo, W., Luo, J., Shuaia, Y., Zhang, K., Wang, T., Wu, C. and Zhang, W., "Infrared detector based on crystal ion sliced LiNbO₃ single-crystal film with BCB bonding and thermal insulating layer," Microelectron. Eng. 213, 1-5 (2019).

[4] Doan, A. T., Yokoyama, T., Dao, T. D., Ishii, S., Ohi, A., Nabatame, T., Wada, Y., Maruyama, S. and Nagao, T., "A MEMS-based quad-wavelength hybrid plasmonic-pyroelectric infrared detector," Micromachines 10, 413 (2019).

[5] Calvano, N. P., Chrostoski, P. C., Braithwaite, K., Rana, M. M. and Voshell, A., "Properties of reactively sputtered Al$_x$N$_y$ thin films for pyroelectric detectors," Proc. SPIE 10381, 103810J (2017).

[6] Mart, C., Viegas, A., Esslinger, S., Czernohorsky, M., Weinreich, W., Mutschall, D., Kaiser, A., Neumann, N., Grossmann, T., Hiller, K. and Eng., L. M., "Pyroelectric CMOS compatible sensor element based on hafnium oxide thin films," Proc. Joint Conf. IEEE Intern. Frequency Control Symp. and Intern. Symp. Applications of Ferroelectrics (2020). DOI: 10.1109/IFCS-ISAF41089.2020.9234892

[7] Butler, D. P., Celik-Butler, Z., Jahanzeb, A., Gray, J. E. and Travers, C. M., "Micromachined YBaCuO capacitor structures as uncooled pyroelectric infrared detectors," J. Appl. Phys. 84, 1680-1687 (1998).

[8] Dégardin, A. F., Jagtap, V. S., Galiano, X. and Kreisler, A. J., " Semiconducting amorphous Y-Ba-Cu-O: an attractive material for fast and sensitive thermal sensing in the NIR to THz range," Proc. SPIE 11279, 1127909-1/12 (2020). DOI: 10.1117/12.2546590

[9] Longhin, M., Kreisler, A. J. and Dégardin, A. F., "Semiconducting YBCO thin films for uncooled terahertz imagers," Materials Science Forum 587-588, 273-277 (2008).

[10] Dégardin, A. F., Jagtap, V. S., Galiano, X. and Kreisler, A. J., "Semiconducting Y-Ba-Cu-O thin films sputtered on MgO and SiO$_x$/Si substrates: morphological, electrical and optical properties for infrared sensing applications," Thin Solid Films 601, 93-98 (2016).

[11] Kreisler, A. J., Dégardin, A. F., Galiano, X. and Alamarguy, D., "Low noise and fast response of infrared sensing structures based on amorphous Y–Ba–Cu–O semiconducting thin films sputtered on silicon," Thin Solid Films 617, 71-75 (2016).

[12] Gensbittel, A., Dubrunfaut, O., Jagtap, V. S., Kreisler, A. J. and Dégardin, A. F., "Radiofrequency dielectric properties of amorphous semiconducting Y-Ba-Cu-O oxide thin films for bolometric detection," Thin Solid Films 520, 4749-4753 (2012).

[13] Dillner, U., Kessler, E. and Meyer, H.-G., "Figures of merit for thermoelectric and bolometric thermal radiation detectors," J. Sens. Sens. Syst. 2, 85-94 (2013).

[14] InfraTech. Available online: www.InfraTec.de (accessed on 20 July 2023).

[15] Heinmann Sensor. Available online: www.heimannsensor.com (accessed on 20 July 2023).

[16] Qinqin, W., Xicai, L., Xingyu, Y., Bin, X. and Yuanqing, W., "Pyroelectric infrared device with overlap dual capacitor structure sensor," Sens. Actuator A Phys. 282, 192-196 (2018).

Improved MCT eSWIR modules for various demands of imaging applications

R. Wiegleb*, R. Breiter, H. Lutz, A. Sieck

AIM Infrarot-Module GmbH, Theresienstr. 2, 74072 Heilbronn, Germany

ABSTRACT

The Short Wave Infrared (SWIR) market for defense, aerospace and industry applications is rapidly growing. A wide range of applications is available: From night vision to plastic sorting up to hyperspectral imaging. The classical SWIR detectors, which are based on InGaAs technology, have a cut-off wavelength typically around 1.7 µm. The extended SWIR (eSWIR) technology, with a spectral range up to 2.5 µm however offers significant advantages over traditional SWIR detectors. These are e.g. full exploitation of the nightglow spectrum, enhanced imaging under low light condition or 'out of band' operation with laser illuminators with a wavelength >2 µm. In addition, eSWIR technology includes reflective and thermal imaging, which allows detecting thermal target radiation in complete darkness.

AIM presents its modules with different array formats and pixel pitches in the eSWIR spectral range from 0.8 µm up to 2.5 µm. Additionally we show that the cut-on wavelength can be lowered down to ~0.4 µm to extend the spectrum to the visible range by removing the detector substrate.

A low size, weight and power (SWaP) eSWIR 640x512 10 µm pitch module has been developed and produced by AIM. It is best suited for handheld imaging or hyperspectral imaging applications with e.g. integration in UAVs. Key enabler for low SWaP however is higher operating temperature detector technology. Hence, we present the last improvements on dark current density and electro-optical performance of our eSWIR MCT HOT detector technology.

Keywords: MCT, extended SWIR, eSWIR, SWIR, hyperspectral imaging, SWaP, high operating temperature, HOT

1. INTRODUCTION

SWIR image sensors are commonly associated with InGaAs detectors, typically sensitive to wavelengths between 900 nm and 1700 nm. However, MCT (Mercury Cadmium Telluride) technology offers the ability to adjust the bandgap and thereby the cut-off wavelength, extending it up to 14 µm, while maintaining a low number of material defects. For eSWIR imaging sensors developed by AIM, a cut-off wavelength of approximately 2.5 µm has been chosen, primarily due to limitations imposed by the atmospheric window.

In comparison to traditional SWIR sensors, eSWIR imaging sensors exhibit several advantages. These advantages are particularly noteworthy for reconnaissance applications, where they excel in detecting thermal radiation from targets and functioning with laser illuminators operating beyond 2µm. Additionally, eSWIR sensors have the capability to observe thermal target radiation through optical materials like building windows or vehicles glass, which is not easily achievable with conventional thermal imaging sensors [1]. During daylight conditions, the eSWIR imaging sensor captures reflected sunlight, making it comparable to conventional visible imaging sensors and thus easier to interpret the image for an observer. At moonless nights, the eSWIR sensor captures the entire nightglow spectrum, yielding supplementary information. Another noteworthy benefit of SWIR imagers is their enhanced ability to penetrate fog, dust, or haze compared to traditional visible sensors, thanks to reduced Rayleigh scattering [2].

Standard eSWIR detectors based on MCT have a cut-on wavelength of about 800 nm defined by the CdZnTe substrate. AIM has successfully demonstrated the removal of the CdZnTe substrate, thereby extending the spectral range to the visible spectrum. AIM has therefore presented a 1024x256 Vis-SWIR detector with a 24x32 µm rectangular pixel pitch and a 384x288 Vis-SWIR detector with a 24 µm pitch for extended hyperspectral applications [3].

AIM has introduced various eSWIR imaging sensors with different array formats and pixel sizes, primarily for applications involving hyperspectral imaging devices [4]. Nevertheless, there exists a demand for Low SWaP (Size, Weight, and Power) detectors suitable for handheld use or integration into UAVs. These devices hold potential applications in remote sensing for areas such as night-vision, vegetation analysis, agriculture, mining and geology. Notably, AIM has already presented an eSWIR Clip-on targeting sight for weapons, powered by a battery pack [1]. In this paper, an eSWIR module with a 640x512 array, 10 µm pitch and F/1.6 cold shield will be presented, which is suitable for Low SWaP applications.

One main drawback of MCT eSWIR imagers is that the detector needs to be cooled down to a typical temperature of 175 K. Recent improvements however at AIM detector technology have enabled up to 20 K higher detector temperatures while maintaining the same dark current.

2. ADVANTAGES OF ESWIR IMAGING

eSWIR sensors are often used in hyperspectral imaging applications since the extended SWIR spectral sensitivity up to 2.5 µm provides relevant fingerprints for many materials [5][6]. For imaging in low light conditions a benefit of eSWIR imagers, in contrast to traditional SWIR imagers, lies in their capacity to detect thermal targets with temperatures of around 300 K. AIM has demonstrated this advantage for reconnaissance purposes [1]. The spectral radiance across the wavelength range of 1.5 µm to 2.5 µm for an ideal blackbody source at T=300 K is illustrated in Figure 1. Notably, at 2.5 µm, the spectral radiance is approximately 300 times greater than at 1.8 µm.
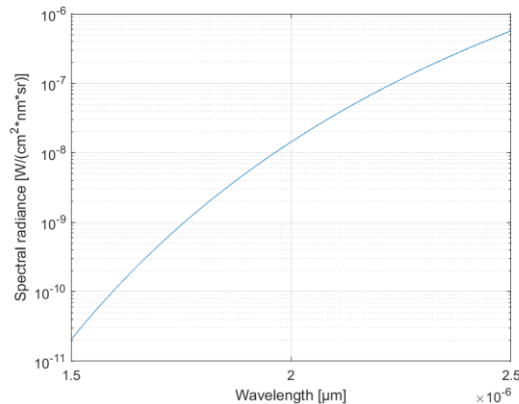
Figure 1: Spectral radiance in W/(cm²·nm·sr) of an ideal blackbody source at T=300 K blackbody temperature between 1.5 µm and 2.5 µm emission wavelength.

When comparing the spectral radiance with the night glow, as depicted in [7], the spectral radiance at approximately 1.8 µm demonstrates a similar order of magnitude. Consequently, the thermal target radiation is more pronounced in the eSWIR range than the night glow effect. This indirect observation is discernible in Figure 2, where two images were captured—one using a visible image sensor (on the left), and the other utilizing an eSWIR 640x512 10 µm F/1.6 image sensor (on the right). These images were taken under conditions of low photon flux (<1lux) during the nighttime. Both images exhibit the reflection of the street light.

Figure 2: Detection of thermal and reflected radiation with an eSWIR 640x512 10μm pitch image sensor (right) compared to a conventional visible camera (left) in a low light condition.

Nevertheless, within the eSWIR image, the thermal signatures of two individuals remain discernible, even under the influence of diffusely reflected light cast by the street lamp. Moreover, the heat signature emitted by the nearby car's exhaust is also observable.

For reconnaissance purposes it can be helpful to see objects behind glass like in buildings or in vehicles. As shown in Figure 3, eSWIR combines the unique feature to see reflective and thermal signatures behind an optical window (fused silica). Conventional fused silica has its cut-off around 3 μm and transmits all wavelength in the eSWIR. In Figure 3 four images are depicted of a person with an optical glass plate in front of the face under different illumination condition. In (1) a quartz tungsten light bulb was used, in (2) and (3) a 1.57 μm and 2.2 μm laser diode and in (4) no light source. In any of the cases, the light was transmitted through the glass.
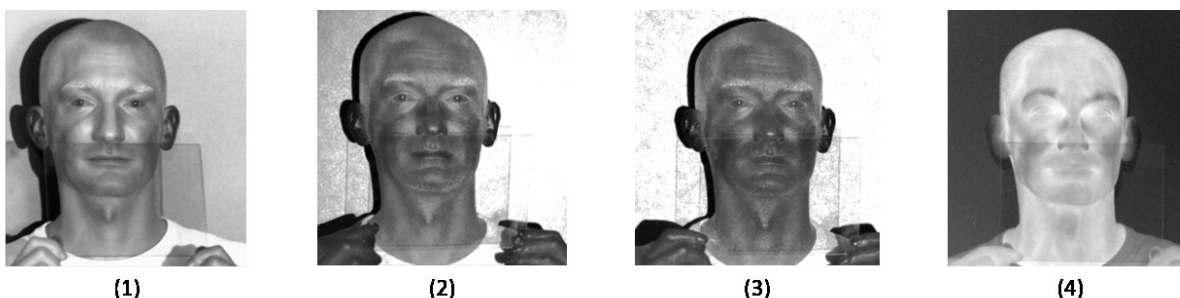
Figure 3: Images taken with an eSWIR MCT 640x512 10 μm pitch image sensor from a person with an optical window under illumination with different light source. (1) With a quartz tungsten light bulb, (2) With a 1.57 μm laser diode, (3) With a 2.2 μm laser diode and (4) No light source (thermal).

Another advantage of SWIR in general is the ability to better see through fog, haze or dust. As shown in Figure 4, the eSWIR image sensor (right) shows much more details at larger distance than a conventional visible imager (left) at foggy weather condition. This is due to limitation of Rayleigh scattering, which magnitude is inversely proportional to the fourth power of the wavelength. This makes eSWIR with its extended cut-off wavelength even more suitable for such application fields [2].

Figure 4: Comparison of images with a visible imager (left) and eSWIR 640x512 10µm pitch F/1.6 imager (right) at foggy weather condition.

AIM has already shown the capability of active imaging with a laser designator with 1.5 µm [8]. However, illumination with higher wavelengths than 1.8 µm can be beneficial for tactical purposes. Figure 5 demonstrates the use of a laser illuminator >2 µm with an eSWIR 640x512 10 µm pitch F/1.6 imager.
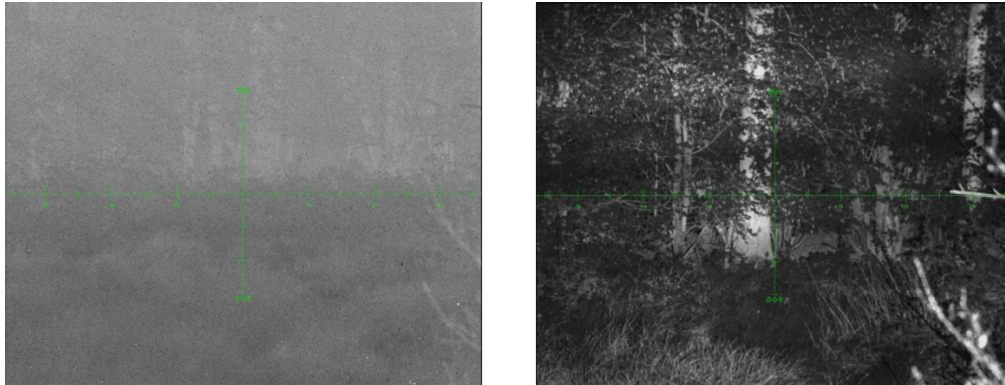
Figure 5: Demonstration of the use of laser illuminator >2 µm (right) with an eSWIR 640x512 10 µm pitch F/1.6 imager.

2.1 VIS enhanced eSWIR

All eSWIR modules produced by AIM can be optionally equipped with a detector that is sensitive to the visible spectrum. This capability has already been demonstrated by AIM in references [3] and [4]. In [3], a cross-section of the FPA (Focal Plane Array) is shown. The PV (Photovoltaic) array comprises a photosensitive epilayer, grown on a CdZnTe substrate. The PV array is interconnected with Indium bumps to the silicon ROIC (Readout Integrated Circuit). PV and ROIC are attached to a ceramics carrier. The cutoff value for spectral sensitivity is determined by the chemical composition between Mercury and Cadmium. Conversely, the cut-on value is determined by the spectral sensitivity of the CdZnTe substrate, which absorbs incident light below 800 nm.

With an AIM internal process it is possible to remove the detector substrate and hence lower the sensitivity down to the visible spectrum. The external quantum efficiency of a conventional eSWIR detector and a VIS enhanced eSWIR detector is depicted in [3]. The overall quantum efficiency of the VIS enhanced eSWIR detector shows slightly lower values than the standard eSWIR detector. This is due to the fact that a different AR coating has to be used which comprises the wide spectral range from 400 nm up to 2.5 µm.

3. AIM ESWIR MODULES

3.1 Overview of eSWIR modules

AIM offers eSWIR modules in various formats and pixel pitches to provide to different applications and system needs. The primary application field for eSWIR modules is hyperspectral imaging. Figure 6 illustrates the most commonly used modules for this purpose. The 384x288 array module with a 24 µm pitch (on the left) comes with a default rectangular cold shield design, while the 1024x256 module with a 24x32 µm pitch (on the right) features a circular F/1.8 cold shield. Higher pixel pitches always benefit from a higher signal to noise ratio (SNR) [9] and might be useful for applications with low photon fluxes (e.g. use of a narrow band pass filter). All modules at AIM can be configured with an individual cold shield design. The 384x288 module is cooled using an AIM MCC030 single-piston linear cooler, having a Mean Time To Failure (MTTF) exceeding 25,000 hours, which makes it suitable for continuous 24/7 operation. On the other hand, the 1024x256 module has an even higher MTTF of over 50,000 hours thanks to its AIM SF100 pulse tube cooler. Notably, the 1024x256 module is substantially heavier at 3.5 kg compared to the 384x288 module, which weighs ~1 kg. Both modules support ITR and IWR as operating readout modes, featuring two gain stages. A unique characteristic of both modules is the ability to independently adjust the gain mode for each row. Both modules are equipped with an interface for power supply, external frame triggering and serial communication over RS232. The video data, featuring a 16-bit resolution, is transmitted through a Camera Link compatible interface operating in base mode. The Camera Link interface can also serve as an optional interface for external frame triggering and serial communication. Key performance parameters for both modules are listed in Table 1.

Figure 6: Photo of an eSWIR 384x288 24 µm pitch module (left) and an eSWIR 1024x256 24x32 µm pitch module (right – without electronics).

The 640x512 module with a 15 µm pitch is well-suited for hyperspectral imaging applications. However, its low SWaP make it particularly advantageous for handheld imaging applications. It features a cold shield design with an F/2.0 aperture, and the detector is cooled using an AIM MCC030 cooler.

This module combines video data with a 16-bit resolution, external frame triggering, and serial communication via RS232 into a single interface. Optionally, an adapter board with standardized electrical interfaces can be provided. Figure 7 displays a photo of the module. It's important to note that the module requires an additional cooling electronics unit, not shown in the photo.

Figure 7: Photo of the eSWIR 640x512 15 µm pitch module with an F/2.0 cold shield design (without cooler electronics).

Table 1: Key performance parameter of 384x288 24 µm, 1024x256 24x32 µm and 640x512 15 µm pitch module

	384x288 SWIR 24 µm pitch module	1024x256 SWIR 24x32 µm pitch module	640x512 SWIR 15 µm pitch module
Material	HgCdTe		
Format	384x288	1024x256	640x512
Pixel pitch	24 x 24 µm	24 x 32 µm	15 x 15 µm
Detector spectral response	0.9 – 2.5 µm (optionally 0.4 – 2.5 µm)		
Readout modes	Selectable ITR / IWR		Selectable ITR / IWR and rolling shutter
Charge handling capacity	> 1 Me- (Low Gain) > 0.34 Me- (High Gain)	> 1.2 Me- (Low Gain) > 0.3 Me- (High Gain)	> 1.5 Me- (Low Gain) > 0.09 Me- (High Gain)
Max. full frame rate	450 Hz	245 Hz	125 Hz
Output video	Camera Link LVDS 16 Bit		
Cool down time (at ambient room temperature)	Typically 5 Min.	Typically 7 Min.	Typically 6 Min.
Power consumption steady state (at ambient room temperature)	Typically 9 W	Typically 20 W	Typically 6 W
Operability	Typically 99.7 %	Typically 99.7 %	Typically 99.7 %
Mean SNR (half well condition)	> 650 (Low Gain) > 350 (High Gain)		> 800 (Low Gain) > 180 (High Gain)

3.2 Low Swap eSWIR 640x512 10 µm pitch module

An IR-module with a resolution of 640x512 and a pitch of 10 µm has been successfully developed by AIM. The module is housed within a compact frame, which includes the detector-dewar and a cooler. The system electronics are attached to the frame, alongside the cryogenic cooler MCC020, which is a linear stirling cooler manufactured by AIM [10].

The cooling process for the detector is facilitated by the MCC020 cooler, with its drive electronics seamlessly integrated into the system electronics. These system electronics are responsible for digitizing the analog video data signal from the detectors with a 14 bit resolution. The communication interface employs RS232 protocol, and the video interface follows a camera link compatible standard. Remarkably, all power supply, communication, and video data interfaces are unified within a single physical interface.

Additionally, the module can be optionally upgraded to function as a camera core. For optical attachments, the module features a mechanical interface, as depicted in Figure 8. Furthermore, an image processing board can be attached to the system electronics. This image processing board, which is shown in Figure 8, provide basic functionalities such as 2-point NUC, bad pixel replacement, automatic brightness adjustment, etc. The video signals can be output either through Camera Link or DVI.

Figure 8: Photo of an eSWIR 640x512 10µm pitch module with image processing unit.

The detector has two operating modes: IWR (Integrate While Read) and ITR (Integrate Then Read). It offers two gain mode with different charge handling capacities (CHC). In high gain mode (HG) the CHC amounts to 120000 electrons, while in low gain mode the CHC has a typical amount of 1.2 Me-. The frame rate is 109 Hz in "standard mode" and maximum 164 Hz in "advanced mode".

They key performance parameter are listed in Table 2.

Table 2: Key performance parameter of 640x512 10µm pitch module

	640x512 SWIR 10 µm pitch module
Material	HgCdTe
Format	640x512
Pixel pitch	10 x 10 µm
Detector spectral response	0.9 – 2.5 µm (optionally 0.4 – 2.5 µm)
Readout modes	Selectable ITR / IWR and rolling shutter optionally on request
Charge handling capacity	> 1 Me- (Low Gain) > 0.1 Me- (High Gain)
Max. full frame rate	164 Hz
Output video	Camera Link LVDS 14 bit
Cool down time (at ambient room temperature)	Typically 3:30 Min.
Power consumption steady state (at ambient room temperature)	Typically 5 W
Operability	Typically 99.5 %
Mean SNR (half well condition)	> 600 (Low Gain) > 180 (High Gain)
Weight	< 400 g

4. RECENT DEVELOPMENT PROGRESS IN HOT ESWIR MCT

A crucial performance parameter for eSWIR image sensors is the overall temporal system noise. This includes the readout noise, the dark shot noise from the FPA and the photon shot noise as a Poisson distributed value. MWIR and LWIR imagers are operated in nearly background limited conditions. That means that the noise is associated with the photons from the background radiation. eSWIR imagers however are typically limited by the intrinsic detector noise. AIM has continuously improved its n/p planar MCT LPE grown eSWIR detector technology in recent years in order to decrease the dark current and thus optimize noise performance as a consequence.

A key e/o figure of merit for evaluating the performance of eSWIR detector is SNR at half well condition (50% of the detector signal range). The SNR of a 640x512 10 μm pitch detector was measured in IWR high gain operating mode with 1ms integration time. The light source used is an integrating sphere with a highly spatially uniform output characteristic and radiance wavelength ranging from the visible up to 2.5 μm. Figure 9 shows two graphs of the signal to noise ratio at half well (x-axis) depicted as a histogram distributed by the pixel values (y-axis) at two different detector operating temperatures 175 K and 235 K. The red curve indicates the gaussian fit, which fits well with the distribution. The mean value and the standard deviation of the SNR remains nearly constant at 235 K compared to 175 K.
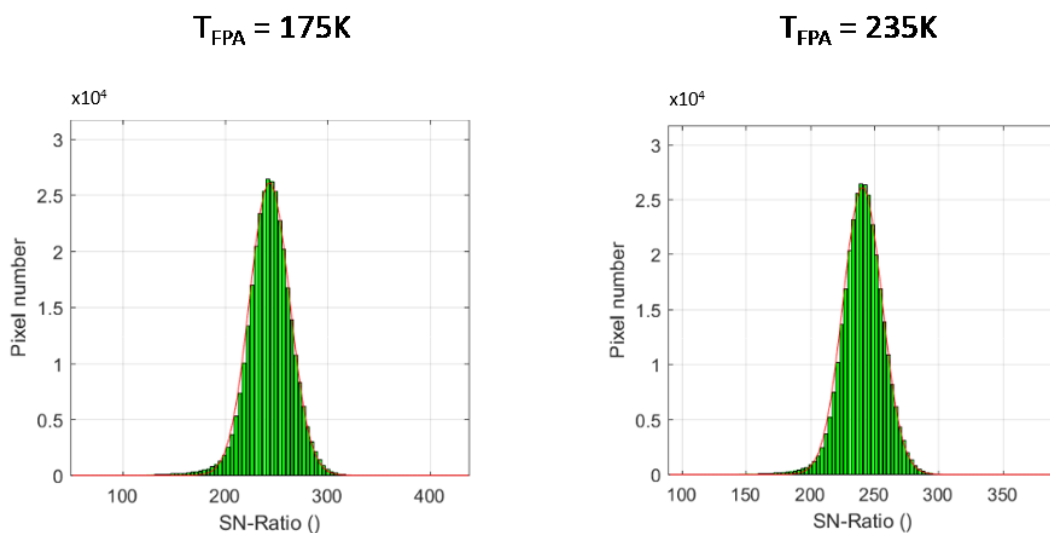
T_{FPA} = 175K T_{FPA} = 235K

Figure 9: Comparison of the SNR pixel distribution at half well condition of a 640x512 10 μm eSWIR detector in IWR high gain with 1ms integration time at two different FPA temperatures. (175K (left) and 235K (right)). The red curve indicates the gaussian fit to the distribution.

Another important parameter is the defect pixel count. In this evaluation a pixel (i=row,j=column) is declared as defective if the following conditions is met:

$$2 \cdot \text{Response}_{mean} < \text{Response}(i,j) < 0.8 \cdot \text{Response}_{mean}$$

$$\text{Noise}(i,j) > 3 \cdot \text{Noise}_{median}$$

The number of defective pixel at typical detector operating temperature of 175 K of an improved FPA is 0.34 %. The defective pixels are mainly dominated by noise defects. At 235 K, the relative number of defective pixels are 0.24 %. According to this evaluation, the amount of defective pixels is slightly reduced due to different noise distribution at these two different operating temperatures. The defective pixels as a 640x512 array map are shown in Figure 10. The black dots are the marking the defective pixel.
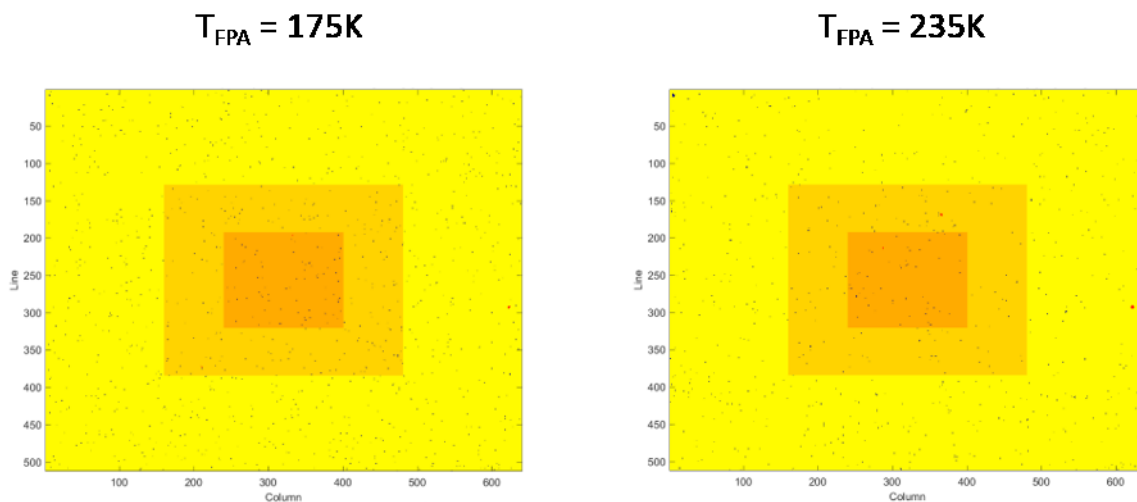

Figure 10: Comparison of the defective pixel maps at 175 K (left) and at an increased operating temperature of 235 K (right).

The noise performance at higher integration time values is mostly dominated by photon noise and the dark current. Since the photon noise is virtually independent from the detector, it is of great interest to reduce the dark current and thus to increase the detector operating temperature towards room temperature. This shall be demonstrated at the two dark current density curves in Figure 11. Here, dark current density values in A/cm² have been acquired with a gold mirror in front of the entrance window in a complete dark room. The diagram is semi-logarithmic with the reciprocal of the cut-off value and the operating temperature on the x-axis. The blue dashed line shows the dark current density of the first AIM MCT n/p 2.5 μm technology, where the orange dashed line shows the improved technology. One can clearly see the lowering of the dark current density by nearly one order of magnitude in the temperature range between 175 K and 210 K. As a reference the Rule07 level is plotted as a yellow line in the diagram. The dark current density values at around 235K of the improved technology falls below Rule07, where the dark current density is mostly dominated by diffusion mechanism. Interestingly, the dark current density values exceeds the Rule07 at lower temperatures down to 150 K. This could be explained by the background radiation limited by the experimental setup.

As illustrated in Figure 11, the dark current density level for the initial n/p technology (blue dashed line) at the typical operating temperature of 175 K was approximately 2-3·10⁻⁹ A/cm². With the improved technology (orange dashed line) it is now achievable to increase the operating temperature up to 195 K while maintaining the same level of dark current density.
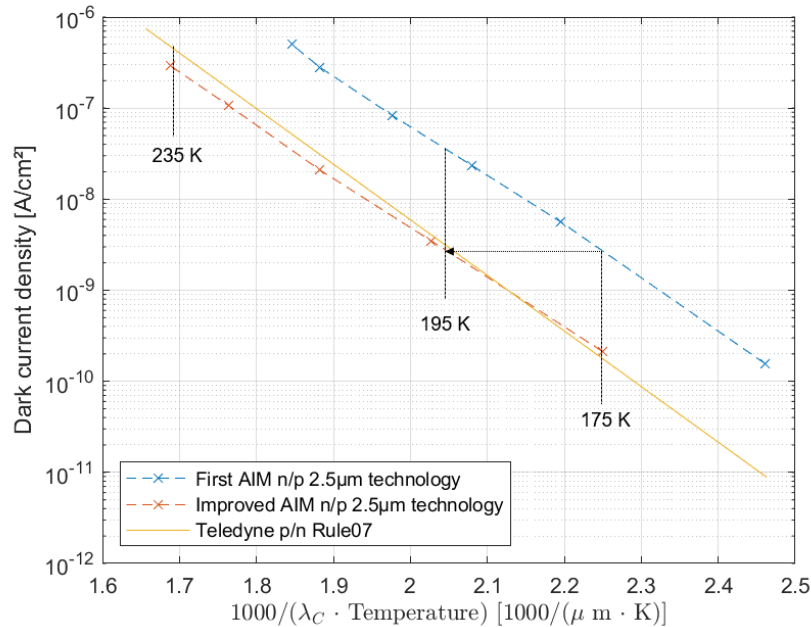
Figure 11: Comparison of the dark current density of a 640x512 10µm pitch eSWIR detector with the first AIM n/p 2.5 µm technology (blue dashed line) and the improved (orange das line) technology design, compared with Teledyne p/n Rule07 (yellow line). The dark current has been acquired with a gold mirror in front of the entrance window.

To analyze benefits of higher operating temperatures on system level for low power consumption application it is important to know what power saving can be achieved at higher operating temperatures. Therefore, the steady state power consumption of an eSWIR 640x512 10 µm pitch module, as presented in chapter 3.2, has been measured at 175 K, 195 K and 235 K operating temperature. At 175 K, the total power consumption of the IR-module including ~1.7 W for the system electronics was at 4.6 W, at 195 K 4.1 W and at 235 K drops to two thirds.

The cool down time (time to image) was 03:10 Min. for 175 K operating temperature, 02:40 Min. for 195 K and 01:40 Min. for 235 K operating temperature.

5. CONCLUSION

The advantages of eSWIR have been demonstrated, such as thermal target detection or out-of-band laser illumination. In addition, AIM has the ability to extend the spectral sensitivity down to the visible range by removing the CdZnTe substrate.

AIM offers eSWIR modules in different configurations and pitch sizes to respond to the wide range of applications, namely a 1024x256 24x32 µm pitch module, a 384x288 24 µm pitch module, a 640x512 15 µm and 10 µm pitch module. In particular, the 640x512 10 µm pitch module is designed for applications where low size, weight and power characteristics are essential, such as UAVs or portable devices.

To further benefit from lower power consumption, which could extend the battery life of a portable device, AIM has presented promising electro-optical results at a higher operating temperature of up to 235 K. Recent development progress of AIM's eSWIR n/p MCT technology has been improved the dark current density by nearly an order of magnitude, which makes it possible to increase the operating temperature by up to 20K and hence lowers the time to image and power consumption of the cooler. Ultimately, the presented dark current density data shows lower values than Teledyne's well established Rule07 at operating temperatures around 235K.

*Robert.wiegleb@aim-ir.com; phone +49 7131 6212 485; aim-ir.com

REFERENCES

[1] Breiter, R., Benecke, M., Eich, D., Figgemeier, H., Ihle, T., Sieck, A., ... & Wendler, J. (2018, May). Extended SWIR imaging for targeting and reconnaissance. In *Infrared Technology and Applications XLIV* (Vol. 10624, pp. 11-21). SPIE.

[2] Wiley, L., Cavanaugh, R., Follansbee, J., Burrell, D., Grimming, R., Pimpinella, R., ... &   Driggers, R. (2023, June). Comparison of reflective band (Vis, NIR, SWIR, eSWIR) performance in daytime reduced illumination conditions. In *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXXIV* (Vol. 12533, pp. 66-80). SPIE.

[3] Höhnemann, H., Hanna, S., Sieck, A., Thöt, R., Benecke, M., Wendler, J., ... & Gross, W. (2019, July). VIS/SWIR IR detectors for space applications at AIM: models and qualification status. In *International Conference on Space Optics—ICSO 2018* (Vol. 11180, pp. 1297-1309). SPIE.

[4] Weber, A., Benecke, M., Wendler, J., Sieck, A., Hübner, D., Figgemeier, H., & Breiter, R. (2016, May). Extended SWIR imaging sensors for hyperspectral imaging applications. In *Image Sensing Technologies: Materials, Devices, Systems, and Applications III* (Vol. 9854, pp. 42-56). SPIE.

[5] Thenkabail, P. S., Gumma, M. K., Teluguntla, P., & Irshad, A. M. (2014). Hyperspectral remote sensing of vegetation and agricultural crops. *Photogrammetric Engineering & Remote Sensing (TSI)*, *80*(8), 695-723.

[6] Wang, S., Zhou, K., Wang, J., & Zhao, J. (2022). Identifying and mapping alteration minerals using HySpex airborne hyperspectral data and random forest algorithm. *Frontiers in Earth Science*, *10*, 871529.

[7] De Gaspari Jan Veldeman Patrick Lamerichs Siegfried Herftijd Patrick Merken and Jan Vermeiren Xenics Nv, D. (2016). The night glows brighter in the Near-IR. *www.photonics.com*. https://www.photonics.com/Article.aspx?AID=50540

[8] Breiter, R., Benecke, M., Eich, D., Figgemeier, H., Lutz, H., Sieck, A., ... & Wiegleb, R. (2019, May). MCT SWIR modules for active imaging. In *Infrared Technology and Applications XLV* (Vol. 11002, pp. 209-218). SPIE.

[9] Jia, J., Wang, Y., Chen, J., Guo, R., Shu, R., & Wang, J. (2020). Status and application of advanced airborne hyperspectral imaging technology: A review. *Infrared Physics & Technology*, *104*, 103115.

[10] Lutz, H., Benecke, M., Breiter, R., Eich, Figgemeier, H., Weber, A., Wendler, J., "Combining compactness and high performance: next generation of cooled MCT IR-modules," OPTRO 2018, 8th International symposium on Optronics in Defence and Security, Paris, (2018).

Investigation of InGaAsSb-based heterojunction photodiodes for extended SWIR imaging

A. Wörl*, R. Müller, V. Daumer, J. Niemasz, F. Rutz and R. Rehm

Fraunhofer Institute for Applied Solid State Physics IAF, Tullastr. 72, 79108 Freiburg, Germany.

ABSTRACT

The detection in short wavelength infrared (SWIR) band, ranging from 1–3 µm, provides a wide range of applications in earth observation, plastics recycling, biology and hyperspectral imaging, gas analysis and defense. In this paper, uncooled InGaAsSb-based detectors for the wavelength range beyond 1.7 µm, the extended SWIR (eSWIR), are investigated for the later use in a thermographic traffic monitoring system that is supposed to localize potentially dangerous overheated hot-spot regions.

Up to the wavelength of 1.7 µm, InGaAs lattice-matched with InP is used for photodetection in the SWIR. To reach a longer cutoff wavelength, "extended InGaAs" can be employed. This requires strained growth that leads to more growth defects and reduced yield, though. InGaAsSb, however, provides a tunable bandgap for detection beyond 1.7 µm and still enables lattice-matched growth on GaSb, which makes it a viable alternative for photodetection in the eSWIR.

We have demonstrated that the bandgap of InGaAsSb can be tuned in the eSWIR by modifying the stoichiometry for lattice-matched growth on GaSb. Furthermore, we have successfully realized InGaAsSb heterojunction photodiodes with an AlGaAsSb hole barrier. At room temperature, the diodes achieve a dark current density of 0.5 mA/cm² and a responsivity better than 1 A/W resulting in an excellent peak detectivity of 9×10^{10} cm Hz$^{1/2}$/W. Thus, the high-performance detector arrays operating at room temperature are within reach in order to meet application demands.

Keywords: InGaAsSb Photodiodes, SWIR, eSWIR, traffic monitoring, infrared imaging, room temperature operation

1. INTRODUCTION

The detection in the short wavelength infrared (SWIR) band ranging from 1–3 µm has attracted high attention over the recent years. Much SWIR research has been focused on the wavelength regime from 1–1.7 µm, mostly due to the availability of the lattice-matched InGaAs on InP material. However, the full SWIR band extends up to 3 µm which is called the extended SWIR (eSWIR). It provides a wide range of applications in, e.g., earth observation, plastics recycling, biology and hyperspectral imaging, gas analysis, and defense[1,2].

To date, eSWIR photodetectors are mostly based on HgCdTe and InGaAs[3,4]. However, the devices based on these materials have intrinsic drawbacks. The development of HgCdTe material is limited by the poor availability of CdZnTe substrate and it does not comply with European rules on the Restriction of Hazardous Substances (RoHS). Extended-wavelength InGaAs photodiodes suffer relatively high dark current because of the increased occurrence of growth defects and dislocations in comparison to lattice-matched InGaAs. Photodiodes utilizing type-II superlattices have also been reported to operate in the eSWIR[5]. Nevertheless, the overall performance of these devices is not yet satisfactory for applications at room temperature. The quaternary material InGaAsSb lattice-matched to GaSb can provide photodetectors operating at room temperature in a wavelength range beyond 1.7 µm[6-10]. For the development of a thermographic traffic-monitoring system that is supposed to increase safety on road and rail by localizing potentially dangerous overheated hot-spot regions, an uncooled InGaAsSb-based detector line array for imaging in the eSWIR is under investigation.

*andreas.woerl@iaf.fraunhofer.de

2. DETECTOR FABRICATION AND RESULTS

In order to investigate the dependency of bandgap energy on stoichiometry composition 500-nm-thick InGaAsSb calibration structures were lattice-matched grown on GaSb substrates by molecular beam epitaxy (MBE). Four different samples with increasing Indium content were grown and analyzed. High-resolution X-ray diffractometry (XRD) yield sharp peaks of the $In_xGa_{(1-x)}AsSb$ layers demonstrating high crystal quality and excellent lattice matching while a smooth surface morphology was achieved. Figure 1 shows the Photoluminescence (PL) characterization at 10 K of the samples with increased concentration of Indium. The sharp peak again shows good material quality of all samples while the measured PL-energy scales with increasing Indium content corresponding to a reduction in bandgap energy of the semiconductor material. Obtained PL-energy at 300 K yields a span of the corresponding 50% cutoff wavelength of the samples from 2.2 to 2.6 microns, which meets application demands. This demonstrates that the bandgap of InGaAsSb can be tuned in the eSWIR-region by modifying the stoichiometry for lattice-matched growth on GaSb.
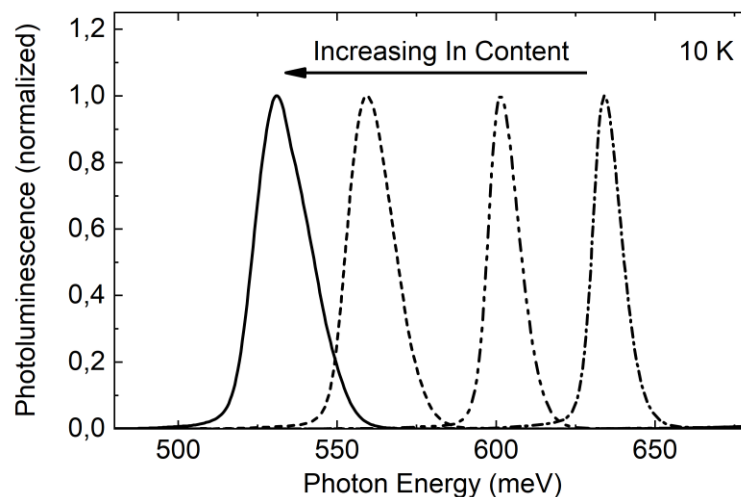

Figure 1. Normalized Photoluminescence of four samples with different Indium content of the lattice-matched grown InGaAsSb structures on GaSb substrate vs photon energy depicting excellent material quality and demonstrating the possibility of bandgap engineering of the detectors for the eSWIR.

After successfully calibrating the quaternary material, detector fabrication was carried out. The devices consist of a 50-nm-thick GaSb n-type (2×10^{18} cm^3) contact layer, a 500-nm-thick AlGaAsSb n-type (5×10^{17} cm^3) hole-barrier layer, a 2500-nm-thick InGaAsSb p-type (6×10^{16} cm^3) absorption layer, and a 1500-nm-thick InGaAsSb p-type (1×10^{18} cm^3) contact layer grown lattice-matched on top of GaSb substrates. The variation in composition and doping concentration of the heterojunction semiconductor was confirmed by secondary-ion mass spectrometry (SIMS). Photodetectors were fabricated with standard photolithographic process. The device mesa was defined by using inductively coupled plasma (ICP) etching down to the InGaAsSb p-type contact layer. Metal layers were deposited on top of the mesas and on the p-contact layer in a lift-off process to form ohmic contacts. Mesas have quadratic and rectangular shapes with different sizes. The latter ones have an opening window at the top-contact metal for front-side illumination purpose. A dielectric passivation layer but no specific antireflection coating was applied.

Figure 2 left shows the current-voltage characteristics of a typical photodiode at room temperature under dark condition. The dark current density at a reverse bias voltage of 100 mV was measured to be 0.5 mA/cm^2. Figure 2 right shows the spectral responsivity at room temperature. The measurements were performed under a reverse bias voltage of 100 mV with the same photodiode as in Figure 2 left illuminated from the frontside at normal incidence. The device exhibits spectral responsivity up to 2.5 μm at room temperature. The 50% cutoff wavelength of the spectral responsivity is 2.4 μm. The peak room-temperature responsivity exceeds 1 A/W.
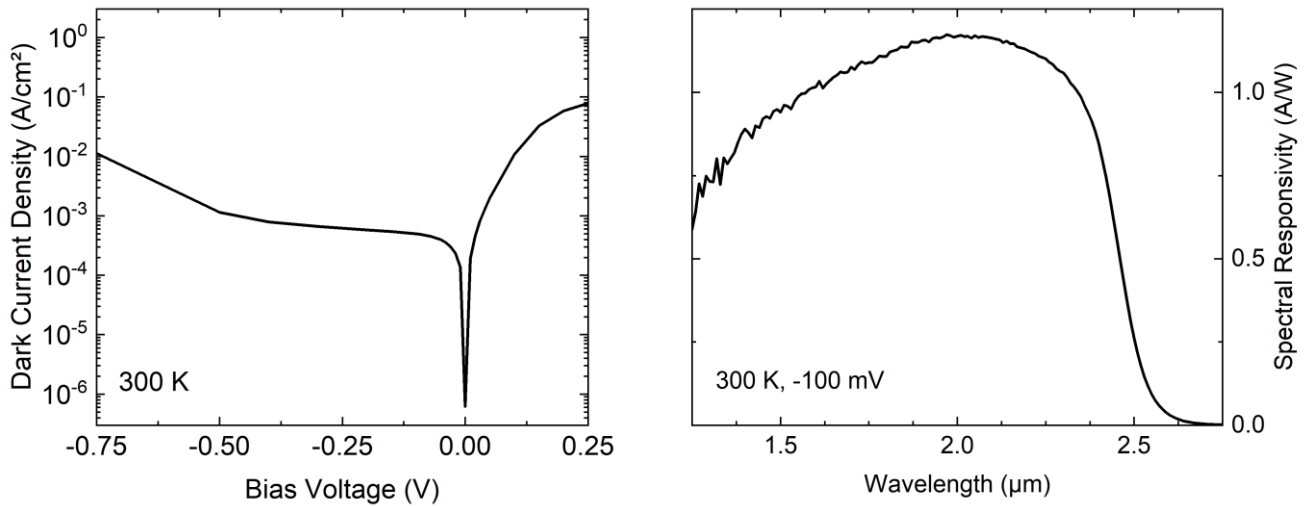
Figure 2. Dark current density of a typical photodiode at room temperature plotted against bias voltage (left) and spectral responsivity of the same photodiode at room temperature and 100 mV reverse bias voltage vs wavelength (right).

Figure 3 shows an Arrhenius plot of the temperature depended dark current density. The fitted activation energy yields a result of 526 meV, which is close to the obtained room-temperature PL-energy and the corresponding responsivity of the device. This correspondence indicates that the dark currents are diffusion limited for temperatures from room temperature down to 240 K. Shockley-Read-Hall (SRH) limited dark currents are known to result in activation energies of around half the bandgap. Therefore, diffusion current was identified as the dominant dark current source at the higher temperature range of interest confirming the suppression of SRH by the barrier design.
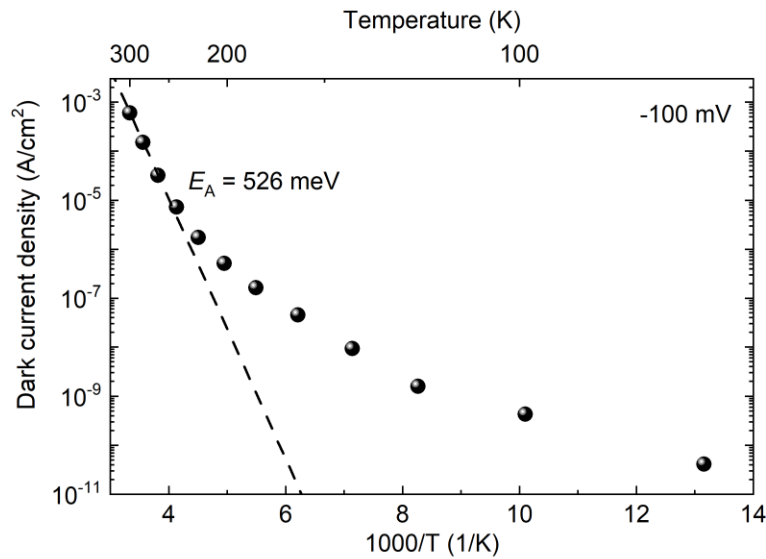
Figure 3. Arrhenius plot of the dark current density of the device at a reverse bias voltage of 100 mV. The dashed line represents the fitted activation energy to be 526 meV.

The shot noise limited detectivity for the InGaAsSb heterojunction photodiode was obtained from the measured responsivity and dark current resulting in a peak detectivity of 9.0×10^{10} cm $Hz^{1/2}$/W at 300 K and under a reverse bias voltage of 100 mV.

3. CONCLUSION

In conclusion, InGaAsSb heterojunction photodiodes lattice-matched grown on GaSb substrates have been demonstrated. High-resolution XRD and low-temperature PL indicated good material quality of the InGaAsSb layer. The heterojunction photodiodes exhibited a low dark current density of 0.5 mA/cm^2 at room temperature and 100 mV reverse bias voltage while the peak spectral responsivity exceeds 1 A/W with a 50% cutoff wavelength at 2.4 µm. A specific detectivity of 9.0 x 10^{10} cm Hz$^{1/2}$/W was achieved at 300 K and 100 mV reverse bias voltage. The dark current of InGaAsSb heterojunction photodiodes is dominated by the diffusion mechanism at temperatures from 300 K down to 240 K. Data reported here demonstrate that the lattice-matched grown InGaAsSb heterojunction material is a very promising candidate for various detection applications in the eSWIR.

ACKNOWLEDGEMENTS

REFERENCES

[1] R. W. M. Hoogeveen and A. P. H. Goede, Infrared Phys. Technol. 42(1), 1 (2001).

[2] M. P. Hansen, V. P. Vavilov, D. S. Malchow, and D. D. Burleigh, Proc. SPIE 6939, 69390I (2008).

[3] M. Vuillermet, D. Billon-Lanfrey, Y. Reibel, A. Manissadjian, L. Mollard, N. Baier, O. Gravrand, and G. Destefanis, Proc. SPIE 8353, 83532K (2012).

[4] C. Besikci, Proc. SPIE 10540, 105400P (2018).

[5] A. M. Hoang, G. Chen, A. Haddadi, S. A. Pour, and M. Razeghi, Appl. Phys. Lett. 100, 211101 (2012).

[6] I. Shafir, N. Snapi, D. Cohen-Elias, A. Glozman, O. Klin, E. Weiss, O. Westreich, N. Sicron, M. Katz, Appl. Phys. Lett. 118, 063503 (2021).

[7] H. Shao, A. Torfi, W. Li, D. Moscicka, W.I. Wang, Journal of Crystal Growth, 311, 7 (2009)

[8] A. P. Craig, M. Jain, G. Wicks, T. Golding, K. Hossian, K. McEwan, C. Howle, B. Percy, and A. R. J. Marshall, Appl. Phys. Lett. 106, 201103 (2015).

[9] G. R. Savich, D. E. Sidor, X. Du, G. W. Wicks, M. C. Debnarh, T. D. Mishima, M. B. Santos, T. D. Golding, M. Jain, A. P. Craig, and A. R. J. Marshall, J. Vac. Sci. Technol. B 35, 02B105 (2017).

[10] D. Z. Ting, A. Soibel, A. Khoshakhlagh, S. A. Keo, S. B. Rafol, A. M. Fisher, B. J. Pepper, E. M. Luong, C. J. Hill, and S. D. Gunapala, Proc. SPIE 10624, 1062410 (2018).

Saddle-shape Whispering Gallery Mode resonators: µm- to cm- scale

Eyal Yacoby[*] and Yosef London

Applied Physics Division, Soreq NRC,Rd 4111, Yavne, Israel

ABSTRACT

Optical Whispering Gallery Mode (WGM) of various sizes and axisymmetry-shapes have been studied and used for a variety of optical sensors. Recently, we suggested a new type of WGM resonators with a saddle shape. These structures consist of two resonators with a bridging region resembling a valley. The unique shape of the saddle-shape resonators may allow the coupling of light into the resonator using a tapered fiber, by placing the tapered fiber at the structural minima point of the valley region. This coupling configuration provides high mechanical stability while maintaining the quality (Q) factor of the joint structured resonator. Here we present saddle shape resonators of various shapes and sizes, suitable for a variety of sensing applications.

Keywords: Whispering Gallery Modes, Microsphere, Mechanical tuning, Saddle-shape, Silica rod resonators.

1. INTRODUCTION

Optical Whispering Gallery Mode (WGM) resonators of various sizes and axisymmetry-shapes have been studied and used for different kinds of optical sensors [1-5]. Among them are sphere [1-4], toroid [5], rod [6-7], bubble [8-9], and bottle [10-13] resonators. Recently, we have suggested a new type of WGM resonators with a saddle shape [14]. The saddle-shape resonators comprise two resonators (spheres or rods in this work) with a bridging region resembling a valley. An illustration of a saddle-shape microresonator built of two microspheres is shown in Fig. 1a.

The unique shape of the saddle-shape resonator allows the coupling of light into the resonator using a tapered fiber by placing the tapered fiber at the minima point of the valley region [14]. This method provides high mechanical stability while maintaining the quality (Q) factor of the resonator [14]. Excitation of WGMs via the valley region is enabled due to the confinement of WGMs of the saddle-shape resonator to the joint structure, including the valley region [14]. Confinement is maintained so long as the cut-off condition of saddle-shape resonators is satisfied, requiring that the diameter of the valley D_{valley} is larger than that of the stems D_{stem} , as shown in fig 1b [14].

Here we present two types of saddle shape resonators. Microresonators, which consist of two welded microspheres of 60 - 200 µm diameters, and cm-scale resonators imprinted on a spinning silica rod using a CO_2 laser. As will be discussed here, the shape and size of the resonators make them suitable for different sensing applications.
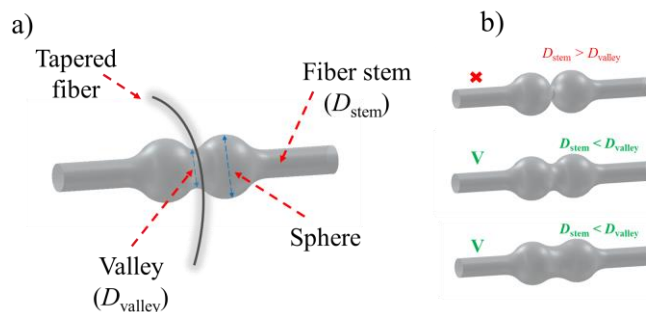
Figure 1. (a) Saddle-shape resonators comprised of two resonators (spheres or rods in this work) with a bridging region resembling a valley. Here, two microspheres are spliced together, forming the saddle shape resonator. The unique shape of the saddle-shape resonator allows the coupling of light into the resonator using a tapered fiber by placing the tapered fiber at the minima point of the valley region. (b) A cut-off condition requires that the diameter of the valley is larger than that of the stems in order to provide confinement of the optical mode.

2. SADDLE-SHAPE MICRORESONATORS

Saddle shape microresonators can be manufactured from two microspheres on a tip of an optical fiber. The microspheres, in turn, are fabricated by welding the tip of the fiber using a fusion splicer [2-3]. The diameters of the microspheres – defining the size and shape of the saddle-shape resonator – can be controlled via the diameter of the melted optical fiber and\or via the intensity of the fusion arcs [2-3]. Examples of saddle-shape microresonators of different diameters and Q – factors are shown in Fig 2. The microresonator shown in Fig. 2a was fabricated by tapering an optical fiber down to a diameter of 26 µm. The tapered fiber was placed inside the arc fusion splicer (Fujikura, FSM-100P) and cut at its waist by pulling one of the fiber ends using the fusion splicer. Two microspheres were manufactured at both ends of the tapered fiber, which were then spliced together. The diameter of the valley (40 µm) is larger than that of the tapered fiber stems (26 µm) as required for exceeding the cut-off condition of saddle-shape resonators [14]. A Q factor of $Q = 3 \cdot 10^7$ is evaluated from the transmission spectrum of the tapered fiber used to couple light into the microresonator (at 1550 nm). The microresonator shown in Fig. 2b was fabricated by the same procedure, using an SMF-28 fiber (not tapered).
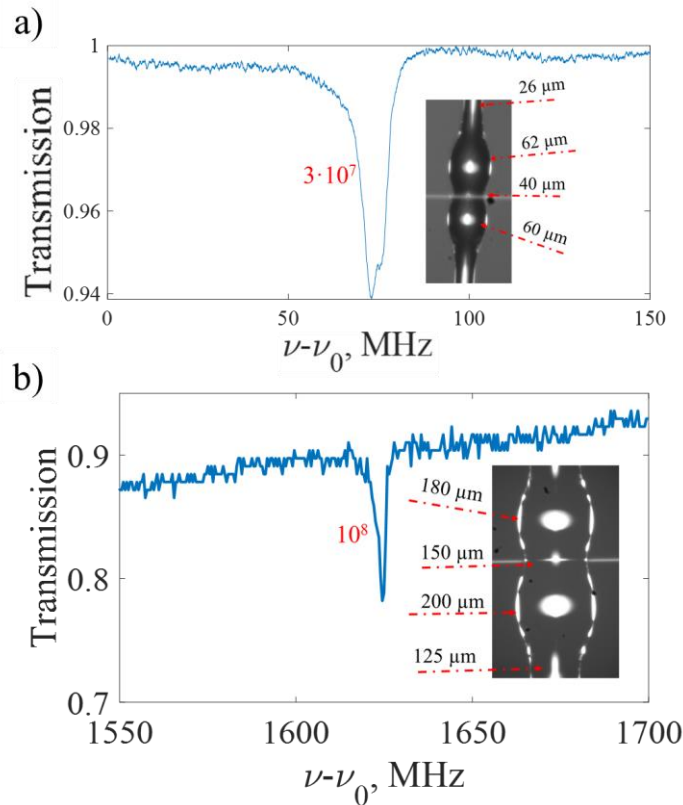
Figure 2. Examples of saddle-shape microresonators of different diameters and Q factors. The saddle-shape microresonators are fabricated by splicing together two silica microspheres, using an arc fusion splicer.

The fiber stems of the saddle-shape microresonators can be used for inducing mechanical strain to the microresonator, which enables the tuning of its WGMs mechanically. This feature may be used as a strain sensor [2], or as a platform for magnetic sensing [3] or any other attribute that may be leveraged to induced strain. Here mechanical strain is induced to a saddle-shape microresonator ($D_{valley} = 165$ µm, $D_{spheres} = 240$ µm and 250 µm) by pulling one stem of the microresonator – using a single axis piezo stage – while the other stem remains fixed. An example of a mechanical responsivity measurement is shown in Fig. 3. The mechanically induced spectral shift is measured while varying the voltage values supplied to the piezo stage, corresponding to different mechanical displacements. The measured mechanical responsivity is 0.045 ± 0.005 pm/$\mu\varepsilon$.
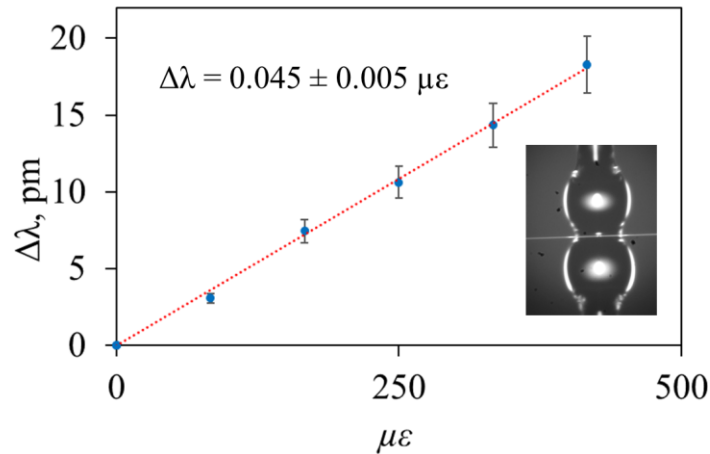
Figure 3. Mechanical responsivity measurement. Mechanical strain is induced to a saddle-shape microresonator (D_{valley} = 165 µm, $D_{spheres}$ = 240 µm and 250 µm) by pulling one stem of the microresonator while maintaining the other stem fixed. The mechanically induced spectral shift is measured while varying the voltage values supplied to the piezo stage, corresponding to different mechanical displacements.

As mentioned above, the WGMs of the saddle-shape resonators are confined to the joint structure, including the valley region as well as the bridged resonators. A picture of an excited resonator is shown in Fig. 4. Fig. 4 a and b show the excited microresonator while placing the tapered fiber on top of the microspheres forming the saddle-shape resonator. It can be seen that most of the light is scattered from the microsphere to which the tapered fiber is coupled, suggesting that WGMs of the microsphere – and not of the joint structure – are excited. On the other hand, in case where the tapered fiber is placed in the valley region, the entire structure is illuminated, as shown in Fig. 3c. It can be explained by the fact that the tapered fiber, when placed in the valley, excites "saddle" WGMs which are guided by the joint structure. This demonstration (purely for illustration purposes) was carried out in the visible regime using a wavelength of λ = 637 nm (TOPTICA).

A numerically simulation supporting this observation is shown in Fig 4d. The spatial distribution of the absolute electric field of a specific WGM – calculated using the commercial COMSOL software – demonstrates the fact that the WGM of the saddle-shape resonator is supported by the joint structure, with a significant portion confined to the valley region.

Figure 4. WGM excitation using a tunable laser (λ = 637 nm, TOPTICA). (a-b) The excited microresonator while placing the tapered fiber on top of the microspheres forming the saddle-shape resonator. Most of the light is scattered from the microsphere to which the tapered fiber is coupled, suggesting that WGMs of the microsphere – and not of the joint structure – are excited. (c) Placing the tapered fiber in the valley results in the excitation of a WGM guided by the joint structure. (d) A numerically simulation supporting this observation. The spatial distribution of the absolute electric field of a specific WGM – calculated using the commercial COMSOL software – demonstrates the fact that the WGM of the saddle-shape resonator is supported by the joint structure, with a significant portion confined to the valley region.
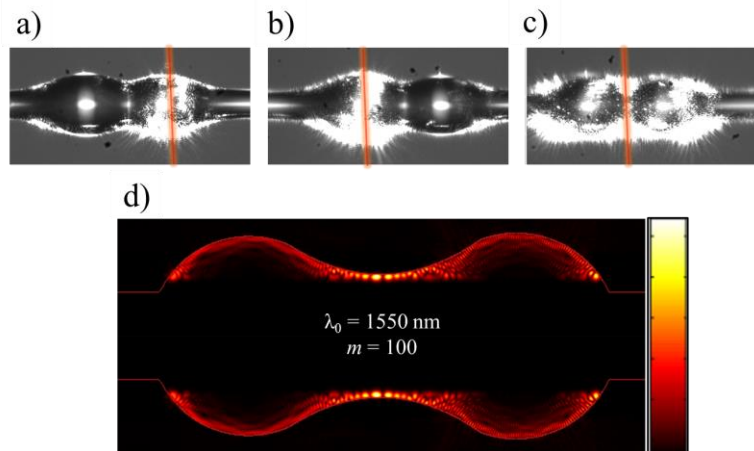
3. SADDLE-SHAPE ROD RESONATORS

The micrometric diameters of WGM microresonators are undesirable for WGM-based gyroscopic sensing, in which the resonator serves as a Sagnac loop [15]. The sensitivity a WGM-based gyroscope increases with the resonator diameter, and therefore cm-scale resonators are required [15].

In [7] we presented the manufacturing of cm-scale ultra-high Q rod (disc) resonators. These rod resonators were fabricated by imprinting slits on a silica rod (while spinning in a motorized spindle) using a high power CO_2 laser. The region between two adjacent slits served as the body of the rod resonator. An illustration of the rod resonator manufacturing is shown in Fig. 5. Here we repeated this procedure using a silica rod of 14 mm diameter, but with an additional (and a smaller) slit in the middle of the rod resonator as illustrated in fig. 5c. Consequently, a saddle-shape rod resonator is obtained. We believe that high-Q, cm-scale, saddle-shape resonator will open the route for the manufacturing of fiber-coupled navigation-grade photonic gyroscope of high mechanical stability.

Following the slit fabrication (ablation), the surface roughness of the saddle-shape ring resonator is high, resulting in a very low Q factor. Efficient excitation of low-Q WGMs of a cm-scale resonators using a tapered fiber may be challenging. The reason is that in order to obtain a significant dip depth which may be used for sensing applications, the coupling loss (=coupling strength/efficiency), induced by the tapered fiber, should be comparable to the internal loss (absorption, radiation and scattering losses) accumulated over a single round-trip. For an uncured resonator of 14 mm diameter, the internal losses accumulated over ~ 44 mm long round-trip may be very high. Therefore, the slitting step should be followed by a curing step in which the resonator surface is heated to high temperature (1700℃ < T < 2700℃) allowing the silica to reflow [7]. This is done by moving the CO_2 laser beam back-and-forth along the resonator (see Fig. 5d) [7].

Fig. 6 shows two different saddle-shape rod resonator imprinted on the same silica rod. One (on the left) is uncured, and therefore of very poor optical quality, whereas the second (on the right) is cured, with a Q factor of $Q = 3 \cdot 10^6$. As shown in Fig. 6, the curing process flattens the contour of the resonator, and decreases the depth of the valley region. The valley depth is crucial for harnessing the saddle shape mechanical advantage and anchoring the coupled tapered-fiber with high mechanical stability. Efforts for improving the fabrication process of the saddle shape rod resonators in terms of deepening the valley and increasing the Q factor are undergoing in our lab.
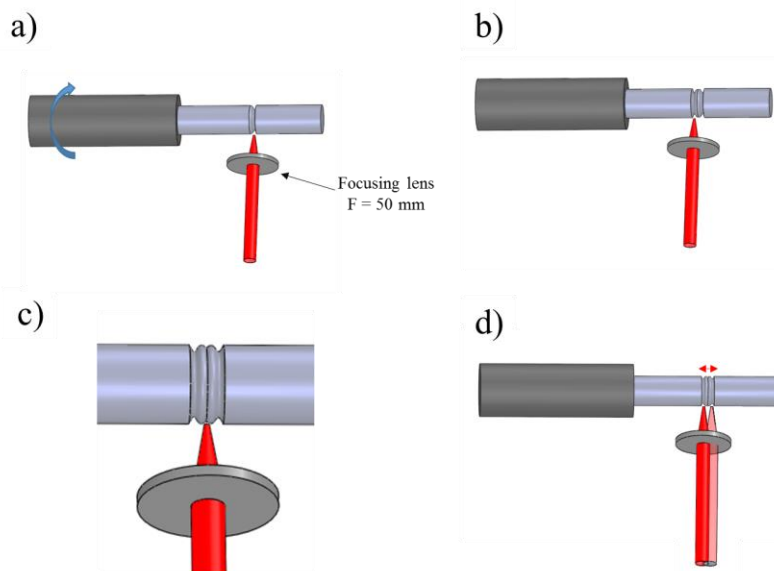
Figure 5. (a-b) Fabrication of WGM rod resonators by imprinting slits on a silica rod using a high power CO_2 laser. The region between two adjacent slits served as the body of the rod resonator. (c) An additional (and a smaller) slit is imprinted in the middle of the rod resonator to form a saddle-shape rod resonator. (d) The slitting step is followed by a curing step in which the resonator surface is heated to high temperature, allowing the silica to reflow. This is done by moving the CO_2 laser beam back-and-forth along the resonator.
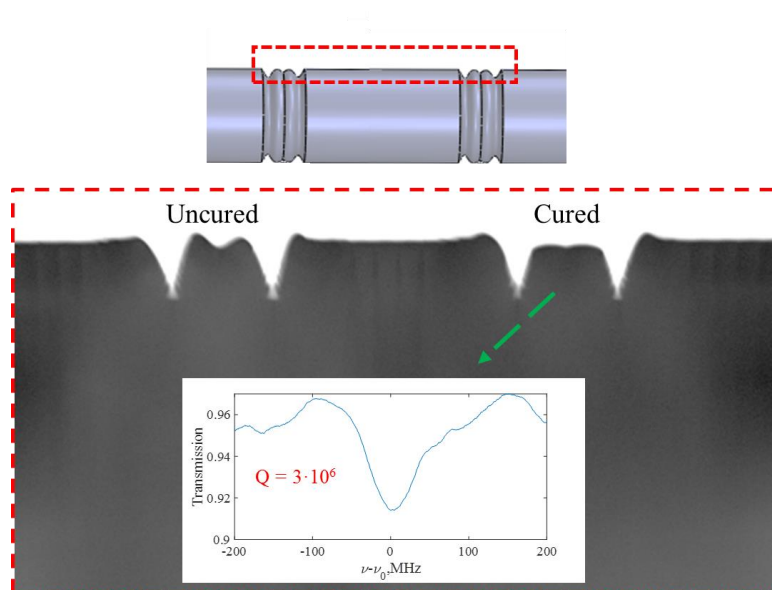
Figure 6. Two different saddle-shape rod resonator imprinted on the same silica rod. On the left – uncured resonator of very poor optical quality. On the right – cured resonator with a Q factor of $Q = 3 \cdot 10^6$. The curing process flattens the contour of the resonator, and decreases the depth of the valley region, which is crucial for harnessing the saddle shapes for anchoring the coupled tapered-fiber.

4. SUMMARY

In this paper we present saddle shape resonators of various sizes, ranging from 60 µm in diameter to 14 mm diameter. We discuss some of the optical properties of said resonators, and overview possible fabrication methods. The micrometric-scale saddle-shape resonator can be tuned mechanically, and therefore may serve as miniature strain and/or magnetic sensors. The cm-scale saddle-shape resonator, on the other hand, may open the route for the manufacturing of miniature, fiber-coupled navigation-grade photonic gyroscope with high mechanical stability.

REFERENCES

[1] Yu, J., Lewis, E., Brambilla, G. and Wang, P., "Temperature sensing performance of microsphere resonators," Sensors (Basel, Switzerland) 18(8), 1-15 (2018).

[2] Yacoby, E. and London, Y., " Mechanical tuning of WGM silica microspheres and WGM-based strain gauge - geometrical considerations," J. Light Technol., to be published, DOI:10.1109/JLT.2021.3138134.

[3] Yacoby, E., Meshorer, Y. and London, Y., "Magnetic sensor based on a Whispering Gallery Mode double-tailed silica microsphere," Opt. Laser. Technol., 151(108019), (2022).

[4] Vollmer, F., Arnold , S. and Keng ,D., "Single Virus Detection from the Reactive Shift of a Whispering-Gallery Mode," in Proceedings of the National Academy of Sciences of the United States of America, PNAS, 105(52), 20701-20704 (2008).

[5] Li, B. B., Bílek, J., Hoff, U. B., Madsen, L. S., Forstner, S., Prakash, V., Schäfermeier, C., Gehring, T., Bowen, W. P., and Andersen, U. L., "Quantum enhanced optomechanical magnetometry," Optica 5 (7), 850-856 (2018).

[6] Del'Haye, P., Diddams, S. A. and Papp, S. B., "Laser-machined ultra-high-q microrod resonators for nonlinear optics," Appl. Phys. Lett. 102(221119),1–4 (2013).

[7] Yacoby, E., London, Y. Harari, G., "Fabrication and packaging of whispering gallery mode rod resonators with adjustable coupling." Optics and Lasers in Engineering 166(107583), (2023).

[8] Sumetsky, M., Dulashko, Y., and Windeler, R. S., "Optical microbubble resonator," Opt. Lett., 35, (7), 898-900 (2010).

[9] Sumetsky, M., Dulashko, Y., and Windeler, R. S., "Super free spectral range tunable optical microbubble resonator," Opt. Lett., 35(11),1866-1868 (2010).

[10] Sumetsky, M., "Optical bottle microresonators," Prog. Quantum Electron., 64, 1-30 (2019).

[11] Qin, H., Yiheng, Y., and Ming, D., "Strain-induced tunable dual-bottle-shaped optical microresonator." Opt. Lett., 44(24), 6017-6020 (2019).

[12] Pöllinger, M., and Rauschenbeutel, A., "All-optical signal processing at ultra-low powers in bottle microresonators using the Kerr effect," Opt. Express, 18(17), 17764-17775 (2010).

[13] Sumetsky, M., "Optical bottle microresonators with axially uniform eigenmode field distribution," Opt. Lett., 45(15), 4116-4119 (2020).

[14] Yacoby, E., and London, Y., "Saddle-Shape Whispering Gallery Mode Microresonators," J. Lightwave Technol. 41, 3139-3144 (2023).

[15] Liang,W., Ilchenko, V., Eliyahu, D., Dale, E., Savchenkov, A., Matsko, A. and Maleki, L., "Whispering gallery mode optical gyroscope," in Proceedings IEEE International Symposium on Inertial Sensors and Systems (2016).

Compact ultra-high-definition mixed reality holographic projections with optical fiber lasers

Jana Skirnewskaja, [a] Yunuen Montelongo, [b] Timothy D. Wilkinson [a]

[a] Electrical Engineering Division, Department of Engineering, University of Cambridge, 9 JJ Thomson Avenue, Cambridge, CB3 0FA, UK; [b] Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK

ABSTRACT

3D mixed reality displays increase accessibility in personal transportation. Personalized Head-up Display (HUD) layouts and the provision of visual depth cues in the replay field can additionally increase safety and security by retaining the driver's focus on the road. Current options for commercial displays fail to provide visual depth cues based on their optical system component arrangements. This work presents a compact mixed-reality volumetric display using virtual lenses, a 4K Spatial Light Modulator (SLM), a combiner, and fiber lasers. The generated mixed reality holographic head-up display aims to contextually enhance the user's perception of the real world with additional information without presenting distractions such as wearable devices or small projections on the windshield. Hence, the optical assembly was designed to be compact such that optical lenses were mostly replaced with virtual Fresnel lenses and the algorithms were accelerated for real-time utilization. Additionally, the accuracy and precision of the replay field results were enhanced with the introduction of pigtailed fiber lasers to reduce speckles. This work has demonstrated 3D ultra-high-definition compact mixed reality holographic replay field results for various applications due to accuracy and precision.

Keywords: Mixed Reality, Liquid Crystal on Silicon, 3D Computer-Generated Holography, RGB Fiber Lasers, Head-Up Displays, Inclusive Design.

1. INTRODUCTION

Safer transportation requires evolving standards. Taking everyone into account when developing technology is crucial to inclusivity and transparency. Holographic Head-Up Displays (HUDs) provide a basis for the transportation sector to build on inclusive design strategies [1]. However, current proposals of holographic HUDs lack a pull-parallax effect, the number of pixels needed to recreate a duplicate holographic mixed reality (MR) object of the original object matching in size and distance, and the speed to reach real-time projections [2]. Holography provides an opportunity for safer and more secure transportation by generating depth cues in the replay field results perceivable for the driver [3]. In consequence, mixed reality head-up displays (HUDs) implement realistic 3D objects and add them to the driver's gaze in the far field on the road without any distractions. This natural visualization technique of physical 3D objects is an attractive tool in the automotive sector. Hence, this study focuses on accelerating the generation of ultra-high-definition (UHD) holographic road signs and obstacles for real-time use during the driving experience. Various existing approaches tackle the challenge of speed for generating computer-generated holograms from real-object data such as the polygon-based method, the depth plane technique, and the point cloud approach [4-6]. To generate real-time high-resolution 3D floating replay field results, a layered holographic method based on phase retrieval and virtual lenses was utilized [7, 8]. A modified and accelerated Gerchberg-Saxton algorithm was generated to project holograms through a Liquid Crystal on Silicon (LCoS) display panel of a UHD Spatial Light Modulator (SLM). Ghost images in mixed reality mode appeared as replay field results at variable focal distances of up to 500 mm. This was achieved through computational techniques by reducing optical lenses in the setup and introducing virtual Fresnel lenses[9]. The customizable 3D holographic projection algorithm allows for the integration of people with various conditions into the transportation sector.

2. METHODS

2.1 Materials

A UHD SLM (3840×2160 px, EXULUS-4K1 from Thorlabs), three single-mode fiber pigtailed laser diodes with FC/PC connections (488 nm at 20 mW), (520 nm at 15 mW) and (642 nm at 20 mW), three plano-convex lenses (f=150 mm, Ø1", N-BK7, ARC: 350-700 nm), and a non-polarizing beamsplitter (50:50 split, 30 mm) were utilized in the study. The UHD SLM was manufactured by Jasper Display Corporation and was found to have an operating wavelength of 400-850 nm after calibrations studies with full factor of >90%, a panel active area of 15.6 mm × 9.2 mm, the pixel pitch of 3.74 μm, phase/retardance range of 2π at 488 nm and at 642 nm, and a frame rate of 30 Hz. The replay field images were taken with a digital camera and equipment from Sony (α7S III E-mount, full-frame sensor (35.6 mm × 23.8 mm), 24 MP) and a camera lens (FE 16-35 mm, F2.8 GM).

2.2 Computer-Generated Hologram generation and calibration.

All CGHs were created by utilizing SolidWorks 2022 (SP3.0, Dassault Systèmes) CAD modeling, and importing 3D models into MATLAB (R2022a, MathWorks). The CGH data was transferred to the UHD SLM via HDMI for replay field projection data. The processing time to generate the CGH via the MATLAB code on a Lenovo ThinkPad laptop (i9-9880H, 2.30 GHz, 64 GB RAM) with a NVIDIA GeForce GTX 1650 Max-Q 4GB GDDR5 graphics card required an average of 1.03 s. Fig. 1 presents the calibration of the developed optical system to determine the recorded resolution of the UHD replay field results. The three different calibration charts in Fig. 1 were specifically chosen to examine the system regarding luminance, accuracy and precision. The analysis in Fig. 1b shows that each chart generated different Gray values and the smoothness of the peaks of the graphs was observed. The spiral chart achieved greatest accuracy and precision due to its sharp peaks, but the highest luminance was achieved by the Siemens star chart. The pattern follows a nearly constant luminance desirable for the replay field results as the optimal luminance for the driver was determined to be an evenly distributed one. Hence, different test charts proof the suitability of the developed optical setup for the application in automotive HUDs.
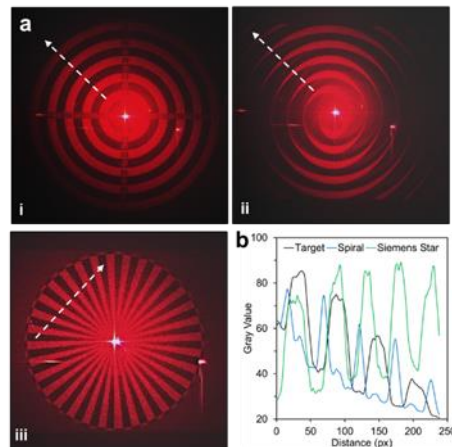
Fig. 1. Calibration of the UHD holographic replay field results. a) Resolution and brightness test charts. b) Recorded resolution and brightness results of the UHD holographic optical system.

3. RESULTS

The use if RGB fiber laser diodes increased the application area of the HUD towards more accessibility options with increased accuracy in the replay field results compared with other light sources such as He-Ne lasers and the use of of several independently controlled light sources. Additionally, multi-phase modulation with the UHD SLM offers absence of conjugate orders and the residual zero-order surface reflection was removed with the improvement of the algorithm.

Another focus of this work was the improvement of perspective and defocus blur. The field of view was determined to be equal to the size of the beamsplitter. However, a panoramic view of several objects aligned with real-life objects was achieved in this study. The fast-processing time for the application of real-time compact and personalizable head-up displays was paramount in this work. Additionally, the 4k resolution and accuracy of the replay field results as 3D objects was achieved as presented in Fig. 2.

Fig. 2. Optical setup and calibration results. (a) Optical system showing a 4K SLM, RGB fiber lasers (488 nm at 20 mW), (520 nm at 15 mW), (642 nm at 20 mW), a combiner, and a beam splitter (BS). (b) Resolution targets: street sign, and construction sign to illustrate the independently controlled fiber laser sources and their alignment in full-color.

The results achieved in Fig. 3. illustrate the achievement of mixed reality holographic replay field results with the RGB fiber lasers. After the calibration of the optical system, the complete assembly was tested with a single wavelength. The test was carried out to determine the accuracy of the replay field results and the alignment level with a real-life object when placed at the same distance. The virtual object was placed at various distances ranging from 50 mm to 500 mm and at the same time the real-life object, a physical 3D object was placed at the same distance with the virtual object to assess the system readiness level. As the virtual holographic objects aligned with the real-life object the size of the replay field results decreases with increasing distance. The holographic projections at 500 mm distance appear more accurate than the closer 3D projections in Fig. 3a. In the far field it would be possible to introduce multiple holographic layers to project several objects at different distances.
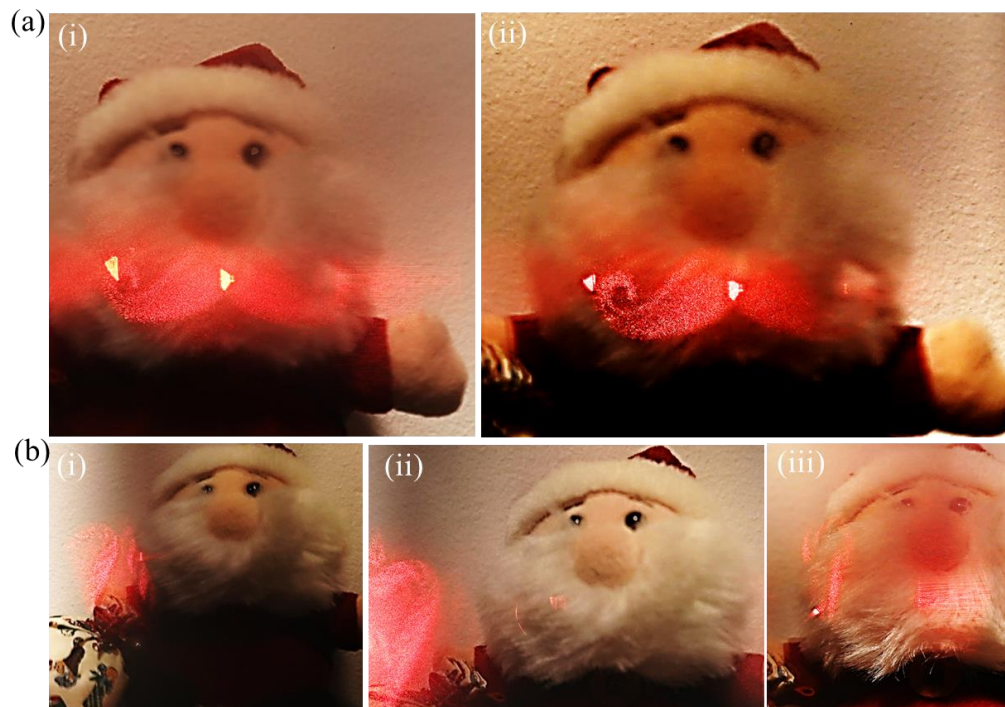
Fig. 3. Mixed Reality holographic replay field results. a) Real-life object at a distance of 50 mm away from the beamsplitter with a holographic 3D projection matching the close distance. b) Real-life object at 500 mm away from the beamsplitter with a holographic 3D projection matching the far distance to mimic the driver's gaze on the road (far distance) and on the windshield (close distance).

4. CONCLUSION

This work has demonstrated real-time holographic mixed reality projections for the application in inclusive head-up displays. As the holographic 3D projections were aligned with real-life objects in the far field, the hypothesis that the driver could maintain focus on the road was upheld. This process could be optimized computationally by analyzing additional accelerating methods such the polygon method to increase the surface area processing. The accuracy of the replay field images could be further enhanced by utilizing a despeckler. However, the accuracy of the replay field results appears more precise with fiber lasers compared with other light sources such as He-Ne lasers as was demonstrated in Fig. 2. The compact size of the optical setup was further reduced with software enhancements such as virtual Fresnel lenses, a LabVIEW program controlling the RGB fiber laser intensity and flicker rate and the accelerated CGH algorithm. As Fig. 2 illustrates, the RGN fiber laser results could be utilized for full-color mixed reality holographic projections or in their single wavelength RGB mode catering to the personal preferences and conditions of the driver. The LabVIEW control of the fiber lasers offers additional precise monitoring of the luminance and the flicker rate of the light sources which can be varied in a range between 1ms to 60s. Future research could focus on the improvement of mixed reality holographic projections to align several holographic objects with real-life objects at varying distances.

Author Information

Corresponding Author: * js2459@cam.ac.uk

REFERENCES

[1]   S. Keates, P. J. Clarkson, L.-A. Harrison *et al.*, "Towards a practical inclusive design approach." 45-52.

[2]   O. Kunieda, and K. Matsushima, "High-quality full-parallax full-color three-dimensional image reconstructed by stacking large-scale computer-generated volume holograms," Applied Optics, 58(34), G104-G111 (2019).

[3]   K. Matsushima, Y. Arima, and S. Nakahara, "Digitized holography: modern holography for 3D imaging of virtual and real objects," Applied Optics, 50(34), H278-H284 (2011).

[4]   K. Matsushima, and S. Nakahara, "Extremely high-definition full-parallax computer-generated hologram created by the polygon-based method," Appl Opt, 48(34), H54-63 (2009).

[5]   D. Arai, T. Shimobaba, T. Nishitsuji *et al.*, "An accelerated hologram calculation using the wavefront recording plane method and wavelet transform," Optics Communications, 393, 107-112 (2017).

[6]   J. Skirnewskaja, and T. D. Wilkinson, "Automotive Holographic Head-Up Displays," Advanced materials, 34(19), 2110463 (2022).

[7]   J. Skirnewskaja, Y. Montelongo, and T. D. Wilkinson, "3D computer-generated holograms for augmented reality applications in medical education." 11842, 431-441.

[8]   Z. Wang, G. Lv, Q. Feng *et al.*, "Enhanced resolution of holographic stereograms by moving or diffusing a virtual pinhole array," Optics Express, 28(15), 22755-22766 (2020).

[9]   J. Skirnewskaja, Y. Montelongo, and T. D. Wilkinson, "Panoramic ultra-high-definition augmented reality 360° color holograms as inclusive tools in transportation." 12445, 84-87.

Design of an imaging system for turbulence mitigation using phase diversity

Nicolas Boehrer, Robert P.J. Nieuwenhuizen, Judith Dijk*, Bart van de Laar, Benjamin Brenny, Niek Doelman

TNO, PO Box 96864, 2509JG The Hague, The Netherlands

Abstract. Long range observation through atmospheric turbulence is hampered by spatiotemporally randomly varying shifting and blurring of scene points in recorded imagery. The image quality degradation induced by turbulence will limit the performance of the system. To mitigate the effect, various hardware strategies combining wavefront sensors together with adaptive optics have been proposed. Both components are associated in a control loop designed to compensate in real time for the aberrations induced by turbulence and reach a system performance close to the diffraction limit. Those techniques are designed for observations of punctual sources within a narrow field of view under which the effect of turbulence can be considered as spatially constant. Unfortunately, the majority of long range horizontal path imaging applications deal with extended sources that are wider than the area under which the turbulence effect is assumed constant. For long range horizontal observation, we devise here a method implementing wavefront sensing using a high speed camera. The system relies on two images of the same scene and same atmosphere realization, having one of the images distorted by a controlled aberration. We describe the simulation and the design choices made for the implementation of such a system. We show that this method allows to measure the relevant shape of the wavefront and provides a way to correct for the effect of atmospheric turbulence.

Keywords: phase diversity, shifting focus, turbulence mitigation, image reconstruction

*Judith Dijk, E-mail: judith.dijk@tno.nl

1 INTRODUCTION

Light traveling through the atmosphere encounters turbulent regions which modify the optical path length [1]. As the light propagates, the effects of turbulent regions accumulate leading to a random phase distortion of the wave front which causes time-varying blurs and shifts in the image recorded by a camera. Atmospheric turbulence therefore limits the effective resolution of optical imaging in many long range observation applications like surveillance or astronomy.

The effect of turbulence can be mitigated by using hardware capable of measuring and correcting for the wavefront distortion while recording (adaptive optics) and/or by using image processing techniques [2]. For astronomy, adaptive optics often performs well to correct the wavefront distortion for the observation of point sources like stars. As an alternative to a real time correction of the wavefront distortion with dedicated hardware, other sensing method were proposed to measure the shape of the distorted wavefront using images recorded at the focal plane of the system. The measured wavefront distortion is then used to enhance images in a post processing step.

The first phase retrieval methods rely on alternatively transforming between the pupil domain and the PSF domain, applying constraints at each iteration (pupil image, PSF, aperture shape) [3]. [4] proposed an extension of this method, by using two PSF images differing by a known defocus. [5] extends the method and shows that by using a pair of images of a same scene with one of the images distorted by a known aberration, phase diversity phase retrieval can be used to estimate the distorted wavefront.

The use of phase diversity phase retrieval has been studied for many applications ranging from astronomical observation, ophthalmology or microscopy. Long range observation for surveillance application is a special case of imaging through turbulence. The horizontal imaging path, which can reach a few kilometers, has the specificity to be entirely located in the atmosphere layer which is the closest to the ground. This layer, also called the surface layer of the atmosphere, is the part of the atmosphere where the turbulence activity is the highest. Apart from the strong turbulence, in surveillance images objects of interest are usually extended. The turbulence effect on the image is not identical on the entire object and varies across the picture.

In this paper, we devise a the approach proposed by [6] of using of time-varying phase diversity phase retrieval for the correction of surveillance images. According to [6], time-varying phase diversity would produce more accurate results and present the advantage to rely on a single optical path, thus not diving the light intensity and signal into multiple path such as done for beam-splitter based setups. Section 2 describes the principle of phase diversity for the correction of surveillance image distorted by atmospheric turbulence, section 3 outlines the design of our system and the experiments. The results are summarized in Section 4.

2 PHASE DIVERSITY

Phase diversity

It is useful to first consider the case of imaging through turbulence under isoplanatic condition. The image formation model can then be locally described by the convolution of the object with the shift invariant point spread function (PSF) combining the aberrations induced by the system and by the atmosphere:

$$y = h \star o + n$$

With y, the observed image, h the PSF, o the object and n additive noise.

The PSF corresponds to the squared Fourier transformation F of the wavefront in the aperture plane W(x,y).

$$h = |F(W(x,y))|^2$$

The wavefront function can be described by:

$$W(x,y) = A(x,y)e^{i\phi(x,y)}$$

With $A(x,y)$ the amplitude and $\phi(x,y)$ the phase distribution in the aperture plane. The phase $\phi(x,y)$ can be decomposed on the orthonormal basis of the Zernike polynomial [7].

In the case of phase diversity phase retrieval, the process relies on recording two images from the same object under the same turbulence realization. The image formation model for the two images can then be described by

$$y_0 = h(c) \star o + n$$
$$y_1 = h(c + \Delta) \star o + n$$

With, c the coefficients of the decomposition of $\phi(x,y)$ on the Zernike basis and Δ the controlled phase diversity added in order to create the aberrated image. One solves the inverse problem of finding the best estimate for the object, ô and the phase coefficient $= \hat{c}$ by iteratively minimizing the cost function

$$\mathcal{C} = (h(\hat{c}) \star \hat{o} - y_0)^2 + (h(\hat{c} + \Delta) \star \hat{o} - y_1)^2$$

The reconstruction algorithm implements the minimization procedure described in [6] and follows an iterative approach. Each iteration consists in, first, finding the object estimate ô which minimizes $\mathcal{C}$ while keeping the coefficients on the Zernike basis $\hat{c}$ constant. This step is done using a gradient descent search. Then in finding the phase coefficients vector estimate $\hat{c}$ which minimizes $\mathcal{C}$ while keeping the object estimate ô constant. This is done using a subspace trust-region Newton-like method. In surveillance imagery, objects of interests will likely cover an angle much larger than the isoplanatic angle. The image is split in tiles, having a size corresponding to the size the isoplanatic patch. Each tile is processed using the algorithm described here, which returns for each tile the best estimate for the object and for the phase coefficients corresponding to the wavefront distortion induced by the turbulence.

Design considerations

For Kolmogorov optical turbulence and for a spherical wave, the isoplanatic patch size S in pixels units is approximately:

$$S \approx \left[115\lambda^{-2}L^{5/3}\int(1 - z'/L)^{5/3}C_n^2(z')dz'\right]^{-3/5}F/P$$

With $C_n{}^2(z)$, the refractive index structure constant at position z, P, the pixel pitch of the sensor, F, the focal distance of the imaging system, and λ, the wavelength of the detected light.

A useful parameter to quantify the effects of turbulence is the atmospheric coherence diameter r_0 (also known as Fried parameter for plane waves). It is a measure of the diffraction limited aperture corresponding to the turbulence condition. For spherical wave propagation, it is defined as:

$$r_0 = \left[16.7\,\lambda^{-2} \int (z'/L)^{5/3} C_n^2(z')dz'\right]^{-3/5},$$

where the integration runs from z=0 at the scene to z=L at the camera. When the C_n^2 is constant, this expression becomes:

$$r_0 = 0.33[\lambda^{-2} L C_n^2]^{-3/5}$$

For constant turbulence strength, the isoplanatic image patch size (which results from the projection of the isoplanatic angle into the image) is approximately equal to:

$$S \approx 0.314\,r_0 F/(L\,P)$$

From this result, we find that the aperture size does not influence the isoplanatic patch size. It is noteworthy though that the patch size above translates to a physical size in the scene of

$$S' \approx 0.314\,r_0$$

To understand the implications of these formulas, Table 1 shows typical values for r_0 assuming λ=550 nm:

Table 1 Atmospheric coherence diameter r_0 for different turbulence strength and different path length

	L = 2 km	L = 5 km	L = 10 km	L = 20 km
$C_n^2 = 10^{-16}m^{-2/3}$	42.3 cm	24.4 cm	16.1 cm	10.6 cm
$C_n^2 = 10^{-15}m^{-2/3}$	10.6 cm	6.1 cm	4.0 cm	2.7 cm
$C_n^2 = 10^{-14}m^{-2/3}$	2.7 cm	1.5 cm	1.0 cm	0.7 cm
$C_n^2 = 10^{-13}m^{-2/3}$	0.7 cm	4.0 cm	0.3 cm	0.2 cm

From these values we can deduce that the isoplanatic patch size will often be only a few centimeters in the scene. Therefore the application of phase diversity will require a very large magnification to provide sufficiently many pixels within the isoplanatic patch. Considering for example a camera with a pixel size of 5.60 by 5.60 µm, and a focal distance of 1m, the image size (in pixels) of the isoplanatic patch corresponding to the scenarios described above are shown in Table 2.

Table 2 Corresponding isoplanatic patch size for a system with F=1m and a pixel pitch of 5.6µm

	L = 2 km	L = 5 km	L = 10 km	L = 20 km
$C_n^2 = 10^{-16}m^{-2/3}$	11.9 pix	2.7 pix	0.9 pix	0.3 pix
$C_n^2 = 10^{-15}m^{-2/3}$	3.0 pix	0.7 pix	0.2 pix	<0.1 pix
$C_n^2 = 10^{-14}m^{-2/3}$	0.7 pix	0.2 pix	<0.1 pix	<0.1 pix
$C_n^2 = 10^{-13}m^{-2/3}$	0.2 pix	<0.1 pix	<0.1 pix	<<0.1 pix

Another relevant aspect is rate at which the image distortions are fluctuating as a result of the atmospheric turbulence. For this we consider the phase structure function for a spherical wave and the von Karman turbulence spectrum [8]:

$$D_{sp}(\rho, L) \approx 1.09 C_n^2 k^2 L \rho^{5/3}[1 - 1.31\left(\rho/L_0\right)^{1/3} + 0.39\left(\rho/L_0\right)^{4/3}]$$

For temporal analysis we replace the separation variable ρ by $v\tau$, in which a constant (transverse) wind speed v along the path is assumed and τ is the time difference. Then, the structure function gives the variance of the phase differences over an interval of τ seconds. Since the purpose of our diversity measurement is to estimate the actual phase distortion, a requirement needs to be set on the relative phase fluctuation with respect to the actual phase. The variance of the phase for a spherical wave is given by [9]:

$$\sigma_{sp}^2(L) = 0.787 \, C_n^2 k^2 L \kappa_0^{-5/3} [1 - 0.16 Q_0^{5/6}]$$

in which $\kappa_0 = 2\pi/L_0$ and $Q_0 = L\kappa_0^2/k$. The mean, relative phase fluctuation over τ seconds is determined by the fraction $\sqrt{D(\tau)/\sigma^2}$. The table below shows the relative phase fluctuation as a function of wind speed and time interval.

Table 3 Mean relative phase fluctuation for different wind speeds and time intervals, when imaging an object at 2km distance. The outer scale of turbulence is 10m.

	v=1m/s (1 Bft)	v=2.5m/s (2 Bft)	v=4.5m/s (3 Bft)	v=6.5m/s (4 Bft)	v=9.5m/s (5 Bft)
$\tau = 2$ ms	0.43 %	0.91 %	1.5 %	2.0 %	2.7 %
$\tau = 4$ ms	0.76 %	1.6 %	2.6 %	3.5 %	4.7 %
$\tau = 6$ ms	1.1 %	2.2 %	3.6 %	4.8 %	6.4 %
$\tau = 8$ ms	1.3 %	2.8 %	4.5 %	6.0 %	8.0 %
$\tau = 10$ ms	1.6 %	3.4 %	5.3 %	7.1%	9.6 %

If we set 5% as the maximum relative phase fluctuation, the allowed time – wind speed product equals: $(v\tau)_{max} = 0.04 \, m$. Note, that this requirement scales almost linearly with the outer scale. For 10% maximum relative phase fluctuation, the allowed time-wind speed product is $(v\tau)_{max} = 0.1 \, m$.

Finally, in order to create diversity, a controlled defocus aberration is introduced while recording one of the frame. [10] indicates that choosing a defocus distance d such that the peak-to-valley (PV) path difference Δ corresponds to the wavelength λ brings optimal performance.

$$\Delta = \frac{d}{8(F/D)^2}$$

Table 4 shows defocus distances for our system having F=1m, D=0.15m and for $\lambda = 550$nm

Table 4: Defocus distance corresponding the different PV distances

Δ	d
0.25 λ	49 µm
0.50 λ	98 µm
1.00 λ	196 µm
2.00 λ	391 µm

Simulations

With the main design parameters defined, we simulated the shape of the system PSF using the Fast Fourier Transform (FFT) method of Zemax [11]. The FFT PSF results were compared to the more complete but much slower Huygens PSF method. The differences were small enough such that the faster FFT method could be used. Calculations were run with a 1024*1024 pupil grid and a 2048*2048 image grid. Figure 1 shows the result of the numerical simulations.
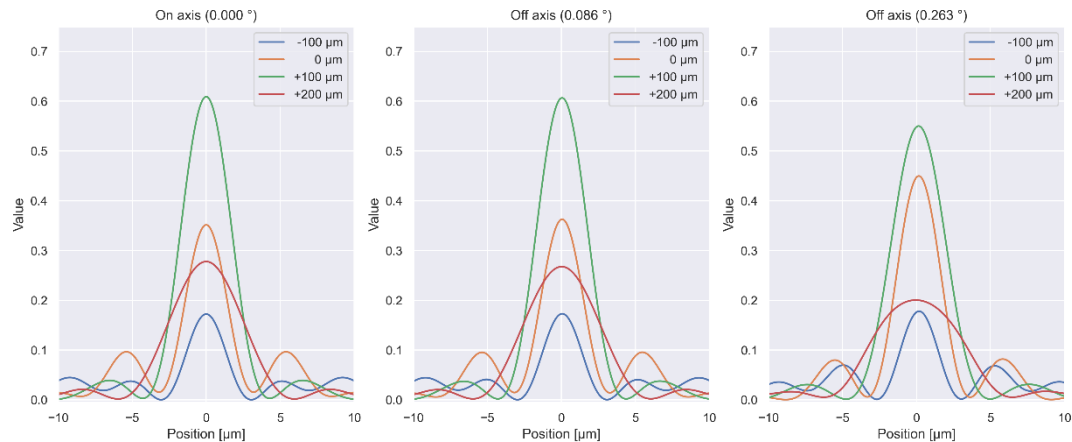

Figure 1: Simulation of the PSF cross sections for different defocus distances and at different locations in the image (left to right)

The figure shows, for different angles between the optical axis and the chief ray and for different defocus distances along the optical axis, the shape of the expected PSF. One can observe that the PSF full width at half maximum (FWHM) is close to the one expected for the diffraction limit (3.4 µm). The shape of the PSF is also compact enough such that the energy of a point like source is concentrated on an area smaller than a pixel (5.6µm). Surprisingly, the simulation shows that the PSF corresponding to a defocus distance of 100µm is more compact than the one corresponding to no defocus distance (0µm). This could be due to the fact that spot size was not the only optimization metric, but others aspects such as telecentricity may also play a role. Next to that, the wavelength and the field angle variation between the recorded images and simulations could also explain that the optimal image sharpness is not found for 0µm defocus shift. These results indicate that the image recorded with a defocus distance of 100µm would be sharper than the one recorded with applying a defocus.

Using the PSF simulation results, we also simulated the output quality of the final reconstruction. The principle of the simulation is shown in Figure 2. A ground truth image (cameraman) is degraded with an arbitrary distortion on the phase coefficient, with c_x a slope running from +5 to +30 and c_y constant +5. The degraded image is then convolved with both PSF's to simulate the two images (called PD-0 and PD-1) that would have been recorded by the system. The correction algorithm (described in section 2 above) generates a corrected image from the two simulated images. The corrected image is compared to the original distortion free ground truth image and the peak signal to noise ratio (PSNR) between the corrected image and the ground truth is computed. The results are summarized in Table 5.
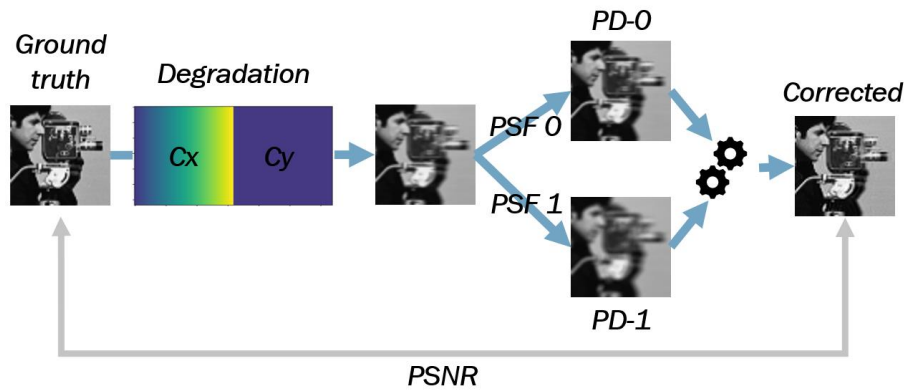
Figure 2 Simulation principle

Table 5: PSNR of the corrected image for different combinations of defocus

	Defocus [µm]		PSNR Field position (0.000° - 0.086° - 0.263°) [dB]
	PD0	PD1	Corrected image
A	0	-100	28.4 – 28.6 – 30.5
B	0	100	29.0 – 29.3 – 31.4
C	0	200	28.9 – 29.2 – 31.1
D	-100	100	24.6 – 24.7 – 25.7
E	-100	200	24.5 – 24.7 – 25.7
F	100	200	**35.4 – 35.5 – 35.0**

These second set simulated data tend to show that the highest PSNR for the corrected image can be achieved when the sharpest image (100µm defocus) is used like in the simulation B and F. Furthermore, when referring to Figure 1, the best results (F) are obtained when using the combination of sharpest and the blurriest images (200µm defocus). In our particular case the optical axis position difference between the two frame that give the best results is 100µm (0.51λ PV distance).

3   EXPERIMENTAL SETUP

Implementation

Figure 3 shows a picture of our setup and Table 6 summarizes the main parameters of the setup. The system consists of a high speed camera Phantom Miro C210 [12] recording the image of the scene focused by an achromat doublet lens (1m focal length and 150mm aperture diameter). To simplify the model, we perform the correction for a single wavelength and we select a 10nm wide region around λ=550nm using a bandpass filter mounted on the camera.

The defocus variation is generated by translating the image plane along the optical axis. Since the time span during which the turbulence effect can be considered constant is relatively short (Table 3), one needs a translation rapid enough to move the imaging plane by the defocus distance (Table 4) during that interval. The fast translation required to record the two images is generated by stacking two translating stages (each of them having a maximum velocity of 20mm/s) carrying two mirrors that fold the rays back to the camera and multiply the displacement by a factor 2. This construction allows to reach

a speed of 80mm/sec, being able to create phase diverse images with 1.0λ PV distance suitable for correction of turbulence effects with a time scale τ_0 up to 2.2ms.
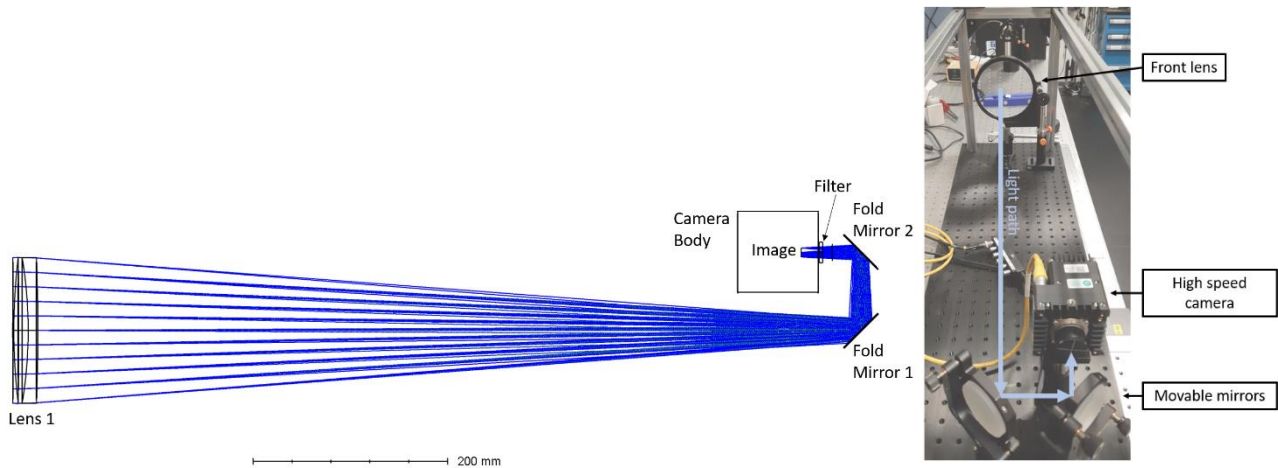

Figure 3 Picture of the setup. Left the Zemax layout, where the mirrors are the movable parts. Right an image of the setup on the optical bench.

Table 6: Main parameters of the phase diversity setup used in the experiment

Optics	
Focal length	1000 mm
Aperture diameter	150 mm

Detector	
Resolution	768 x 768 pix
Pixel pitch	5,6 µm
Spectral bandwidth	550nm /10nm
IFOV	5,6 µrad
Frame rate	2 kHz
Image plane velocity	80 mm/s

As shown in the previous sections, a good system resolution is necessary to resolve sufficient detector pixels within the isoplanatic patch size. Given the fixed detector pixel pitch, the resolution will scale with the focal length, with a better resolution for longer focal lengths. This must be balanced with the aperture size, which will impact the turbulence observed, the diffraction limit and the instrument f/# and will determine the depth of focus. The latter is preferably small, to generate large phase diversity within small scales. For practical reasons, the system design was limited to the choices of components with long focal lengths (70-200cm) and large apertures (70-200mm). An additional constraint was that even with a 10nm bandwidth spectral filter, the chromatic variations are sufficiently present that an achromatic lens was required. The best balance of performance was found for the parameters shown in Table 6, with 1m focal length and 150mm aperture

diameter. For a target distance of L = 2 km, the expected resolution per detector pixel is ~1.1cm, with a FOV of 8.6m in horizontal and vertical directions.

Experimental setup

Figure 4 shows the principle of the experiment. An USAF chart is placed at 50m distance from the phase diversity setup. The chart is illuminated by a halogen light source.
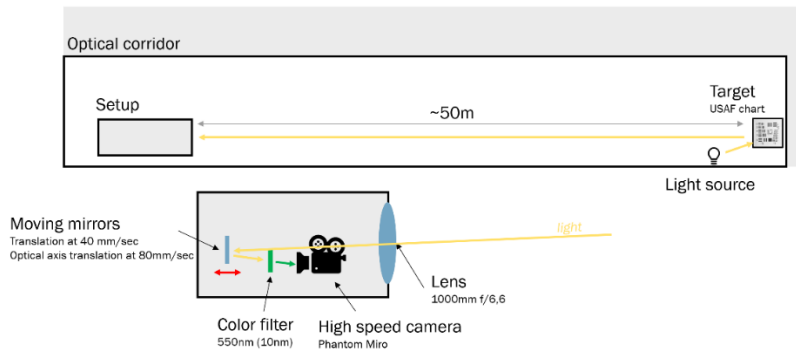
Figure 4 Experimental setup

We recorded static and dynamic sequences as shown in Figure 5:

- 18 static sequences with the stage driven at a fixed position. The defocus distance stays constant during the recording of the sequence. Each sequence consists in 200 frames recorded at 100Hz using an integration time of 2.5 msec.

- 5 dynamic sequences, with the stage moving while recording the sequence. The stage motion induces a displacement along the optical axis at 80mm/sec. We recorded 2 sequences of 2000 frames at 2kHz with 498 µsec integration time, 2 sequences of 2000 frames, recorded at 1kHz with 498 µsec and 998 µsec integration time and 1 sequence of 200 frames, recorded at 100Hz with 2.5msec integration time.
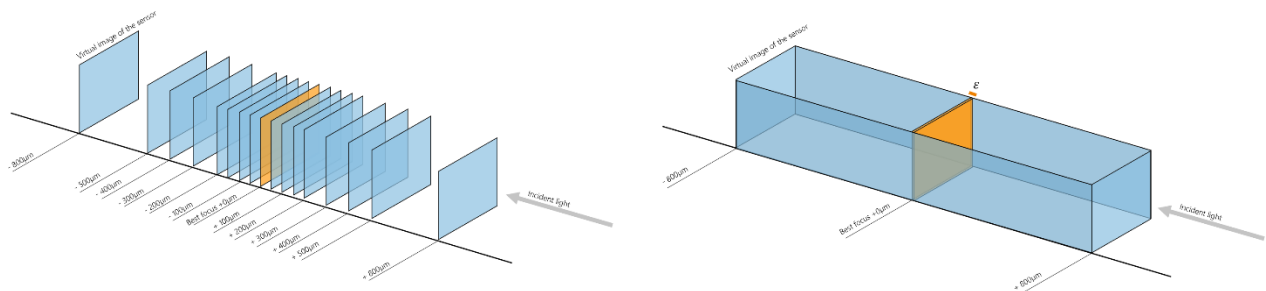
Figure 5 Comparison of the static and dynamic modes used to record the sequences

4   RESULTS

PSF Calibration

We first measured the PSF Full Width at Half Maximum (FWHM) for different position of the stage on static recordings and compared it to the FWHM measured on all frames of a dynamic recording (at 2kHz). Figure 6 shows the result.

Figure 6: Comparison of the PSF FWHM measured on the static sequences compared to the PSF FWHM measured on a dynamic sequence at 2kHz

One can observe that the sharpness measured on the dynamic sequences (FWHM 14µm) is slightly lower than the sharpness measured on the static sequences (FWHM 11µm). The lower sharpness of the dynamic sequences can partially be explained by the stage displacement between t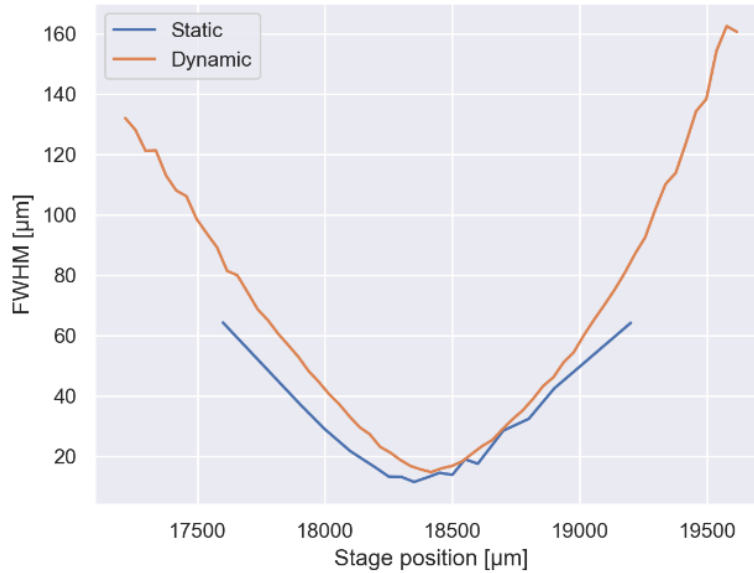he begin and the end of the integration of each frame. With a speed of 80mm/sec, for a dynamic recording made at 2kHz with an integration time of 0.5msec, the sensor has virtually moved by 40µm during the integration of each frame which is not negligible when looking at the differences between the simulated PSFs shown in Figure 1. One can also not rule out that by using stacked stages, low amplitude high frequent vibrations are occurring. Vibrations may bring an additional degradation of the sharpness when compared to static recordings.

It is also worth noting that even with static measurement, the system sharpness differs from the one predicted by the simulation. When comparing those results to the diffraction limit, the FWHM of the airy disk corresponding to our system is

$$FWHM_{airy} = 1.025 \frac{\lambda f}{D}$$

With $\lambda = 550$nm, D =0.15m and f=1.0m, one gets 3.8µm for the FWHM of the airy disk, much smaller than the 11µm measured on the static sequences. Therefore, we conclude that the current setup is not limited by the diffraction but by its own aberrations.

Influence of the defocus distance

Next, we compared the reconstruction quality by varying the defocus distance used to create the aberrated image. The tunnel setup has a single source of light, the lighting conditions are much darker (below 5000 lux) when compared to outdoor conditions (>50000lux). In order to minimize the effect of noise and the effect of the integration while the stage is in motion, conditions specific to our optical corridor, we compared the effect of the defocus distance using the images from the static recordings.

We select the sharpest image and measured the FWHM of the corresponding PSF, which was 11,42µm. We processed the sharp image (called PD0) together with the image (PD1) corresponding to the defocus distance for which we want to evaluate the quality. When comparing the FWHM of the PSF corresponding to the corrected image to the FWHM of the PSF corresponding to both input images, we notice that the sharpness of the output image is in all cases higher than the one of both input images. The sharpness in the output image does not strongly depend on the defocus distance that has

been chosen even though we notice a slight sharpness increase when using longer defocus distances. The FWHM of the PSF corresponding to the corrected image is still 1,8x wider than the FWHM that could be expected from the Airy pattern corresponding to the diffraction limit of the system.

Table 7: FWHM of the corrected image for different defocus distances and detail crops of the sharp, the aberrated and the corrected image.

Defocus distance [μm]	FWHM of the aberrated image [μm]	FWHM of the corrected image [μm]	Sharp image	Aberrated image	Corrected image
100	14,45	7,62			
150	13,77	6,94			
200	18,92	7,17			
250	17,47	6,50			
350	28,45	6,72			
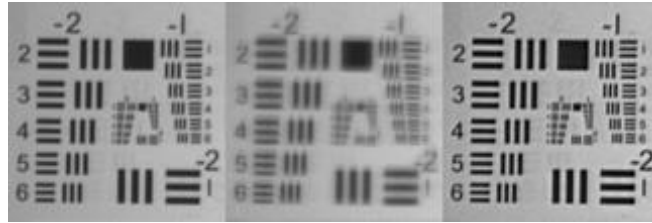
450	32,36	6,49
550	42,45	6,04

5 CONCLUSIONS, DISCUSSION AND FUTURE WORK

In this paper we explored some of the advantages of having a time-varying phase diversity setup for turbulence mitigation. The principle is that we record two images within a short time interval that the turbulence fluctuations are limited. The recording is made while translating the sensor focal plane to record multiple images known aberrations between them. In this paper, the designed setup, where the aberration is inserted by a moving mirror, is described along with experiments in an optical corridor. It is demonstrated that the setup can indeed be used to sharpen the images. However, it can be seen that the movement of the stage introduces more aberrations than originally expected in the simulations. A next step for this work is to test the setup with images recorded over an optical path including turbulence. This can be simulated turbulence by placing heaters in the optical corridor or measurements in the real world. We also plan to compare the time varying phase diversity prototype to the classical single aperture /dual sensor approach.

6 ACKNOWLEDGEMENTS

7 REFERENCES

[1] M. C. Roggemann, B. M. Welsh and B. R. Hunt, Imaging through turbulence, CRC press, 2018.

[2] A. W. M. Eekeren, K. Schutte, J. Dijk, P. B. W. Schwering, M. Iersel and N. J. Doelman, "Turbulence compensation: an overview," in *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIII*, 2012.

[3] R. Gerchberg and W. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik,* vol. 35, pp. 227-246, 1972.

[4] D. Misell, "A method for the solution of the phase problem in electron microscopy," *Journal of Applied Physics,* vol. 6, no. 1, pp. L6-L9, 1973.

[5] R. A. Gonsalves, "Phase Retrieval And Diversity In Adaptive Optics," *Optical Engineering,* vol. 21, no. 5, 1982.

[6] A. Eekeren, K. Schutte, J. Dijk and P. Schwering, "Time-Varying Phase Diversity Turbulence Compensation," *Proceedings of SPIE - The International Society for Optical Engineering,* vol. 8012, no. 05, 05 2011.

[7] R. J. Noll, "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am.,* vol. 66, no. Mar, pp. 207-211, 1976.

[8] R. L. Lucke and C. Y. Young, "Theoretical wave structure function when the effect of the outer scale is significant," *Appl. Opt.,* vol. 46, p. 559–569, February 2007.

[9] L. C. Andrews and R. L. Phillips, Laser Beam Propagation Through Random Media, SPIE Optical Engineering Press, 1998.

[10] D. J. Lee, M. C. Roggemann and B. M. Welsh, "Cramér-Rao analysis of phase-diverse wave-front sensing," *J. Opt. Soc. Am. A,* vol. 16, pp. 1005-1015, 1999.

[11] Zemax, Zemax, [Online]. Available: https://www.zemax.com/pages/try-opticstudio-for-free?utm_source=google&utm_medium=ppc&utm_term=engine:google|campaignid:9218530 070|adid:487119654416|gclid:EAIaIQobChMImdzA6LaC-gIVbI9oCR1LawM4EAAYASAAEgL8o_D_BwE&utm_campaign=G:S:BR:Root:EUR_EN: E:&utm_conte. [Accessed 09 2022].

[12] Phantom, "Miro C210," 02 10 2022. [Online]. Available: https://www.phantomhighspeed.com/products/cameras/mirocnn/c210.

The Practicalities of a Spatial Coherence Filter in free-space LIDAR Environments

Paul Hawthorne*[ab], Mark W. McDonald[a] ,Ian Park[b], Duncan P. Hand[a]

[a] Institute of Photonics and Quantum Sciences, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom; [b] MBDA Bristol, Golf Course Lane,, Bristol BS34 7QS

ABSTRACT

A scanning LIDAR system that uses a single pixel detector can be highly attractive, with simple data processing coupled with low cost and complexity. However, the impact of ambient light noise is much greater than with a multiple pixel system. A potential means of overcoming this is to filter for transverse spatial coherence. Such filters have been discussed and evaluated in the literature, typically based on an axicon or a spiral phase plate that creates a ring with coherent light. Incoherent light, in contrast, smears the light out diffusely, allowing for spatial separation and thus, filtering. The focus of the existing literature tends to be in optical communication or underwater ranging, whereas a free-space LIDAR environment has distinct issues that inhibit the practicality of the filter if a direct replication is performed. This work thus focusses on exploring the practical implementation of these filters in a free-space LIDAR environment.

Keywords: Times Roman, image area, acronyms, references

1. INTRODUCTION

Achieving a long-range LIDAR system is a challenging task, primarily complicated by the large amounts of background light contributed by the sun across a wide spectrum. As such, minimising the amount of background light picked up is essential. This can be accomplished in a variety of ways, such as reducing the field-of-view (FOV) observed (either by compromising on the system's total FOV or by dynamically steering it), using a detector array, range-gating or by taking a frequency-modulated continuous-wave (FMCW) approach. Such methods come with downsides, such as the need for additional electronics with an actively steered FOV approach, the expense of sensitive detectors with an array approach, a loss of dynamic range with range-gating and the need for highly expensive, long-coherence length lasers for FMCW.

A potential method for separating signal from noise that would come with relatively minimal downsides is to employ a spatial filter that exploits spatial coherence. Such a system would take advantage of the higher transverse spatial coherence of the reflected speckle, separating it from the background noise using interference, an approach that would be entirely passive. The noise light could then be blocked, and the signal light reconverted to a form that would allow for a single-pixel detector. This technique has been discussed in various works, exploiting axicons [1], SLMs [2][3], and spiral phase plates [4][5][6] as the optical elements in the filter. The goal of the work presented in this paper is to explore whether these filters can be employed in a free-space LIDAR environment, and whether they offer any improvement over a simple lens and iris.

The transverse spatial coherence length, d_c, describes the spatial extent of a light source [7], specifically its angular extent θ relative to the wavelength λ, as shown in equation (1) [8]:

$$d_c = 1.22 \frac{\lambda}{\theta} \tag{1}$$

The above describes light sources where the points of illumination are spatially uncorrelated with each other (e.g. a bulb filament, speckle [9][10]).

To take advantage of a spatial coherence filter, the signal FOV must be notably smaller than the noise FOV. A scanning LIDAR system describes just such a situation.

*ph2001@hw.ac.uk

2. LITERATURE

There have been a number of papers published in recent years that discuss different methods of employing spatial coherence filtering.

One such paper, "Enhanced underwater ranging using an optical vortex" by Jantzi et al. [4] employed a spiral phase plate to separate coherent backscatter from an underwater target from incoherent forward scatter. A spiral phase plate imparts orbital angular momentum [11] onto a beam (i.e. the wavefront becomes helical through space, with the number of helices per wavelength called the topological charge), generating a ring-shaped Laguerre-Gaussian (LG) output beam from a gaussian input. As it is an interference phenomenon, the formation is coherence sensitive.

The filter thus operates by having coherent backscatter (which originates from a small part of the FOV) form a ring shape, where incoherent forward scatter (originating from a larger part of the FOV) forms diffusely. A streak camera then records both the intensity pattern and the spread of arrival times of the photons. The filter allowed them to reduce the timing error in their measurements, particularly in turbid water.

Another paper that describes a spatial coherence filter is "Numerical simulation model of an optical filter using an optical vortex" by Zhou et al. [3]. In their work, they utilise a spatial light modulator (SLM) with a phase map that will impart a topological charge onto the beam, forming a ring-shaped Laguerre-Gaussian (LG) beam if the incident light is coherent, and experiencing no significant change if the light is incoherent.
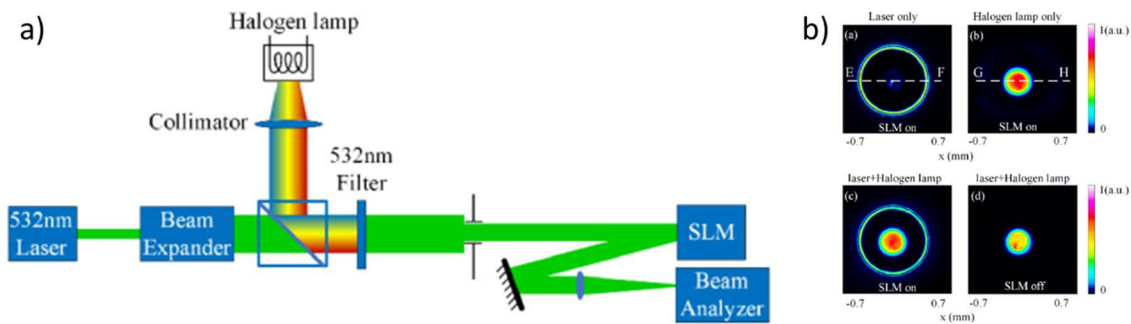
Figure 1. Taken from [3]. a) shows the setup used to demonstrate spatial coherence filtering of a laser and a halogen lamp. The coherent laser light will form a ring after interacting with the phase map of the SLM, where the incoherent light will not interact meaningfully with the SLM. b) shows the spatial separation achieved by the filter.

To further increase the spacing, a Fresnel lens mask was combined with the LG mask, with their results showing that the incoherent light does not strongly interact with this, but the coherent light will defocus to increase the ring radius. Their system is shown in Figure 1, alongside images of their filter.

With these papers as a basis, an exploration of how a spatial coherence filter could be employed in free-space LIDAR was performed.

3. SETUP & METHOD

To explore the implementation of an interferometric spatial coherence filter, the setup in figure 2 was used. Light from a 532nm laser is passed through a ground glass plate (GGP) with a diverging lens placed at different distances to vary the illumination area on the GGP, which in turn varies the FOV of the light scattered by the GGP and thus, the spatial coherence. This light then passes through a linear polariser before interacting with an SLM.

Figure 2. Setup for spatial coherence experiment. Laser light is expanded by a lens before being incident on a ground glass plate and passing through a linear polariser. The light is then incident on an SLM which generates a Laguerre-Gaussian b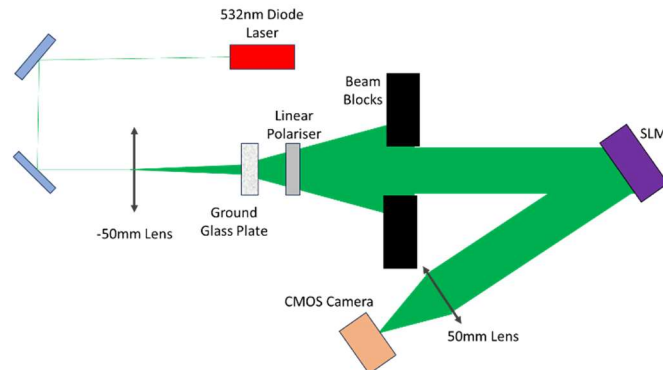eam of varying quality dependent on the spatial coherence length of the incident light, which is later imaged onto a camera.

The SLM adds an LG phase mask to the light incident on it. The linear polariser ensures that the incident light is at the optimal polarisation for the SLM. The filter would take a similar form to Zhou et al's, where a central blockage could filter out the low coherence length light from the higher coherence length light. With a scanning LIDAR, the high coherent length light can originate from different parts of the FOV. Thus, for the filter to operate successfully, the high coherent signal would need to pass the central blockage with low loss in all positions in the FOV, compared to the low coherent "noise".

4. RESULTS

To analyse this, a MATLAB code was developed (adapted from [12]) that examined the total radial power distribution for the image. The code creates pixel bins with different radial bands. The total power within one of these bins is then calculated by simply summing the pixels within it. It is important to look at the total since a single line of pixels would not give an accurate impression due to the speckle distribution.

With the radial power thus measured, this is then translated into a cumulative power plot, to show how the total radial power varies with radii. Finally, this is normalised and subtracted from 1, to show remaining radial power at different radii. This was analysed for light over a large FOV (d_c of 18.9μm), light from a small FOV (d_c of 233μm) in the centre of the large FOV; light from a small FOV at the edge of the large FOV; and light from a small FOV midway between the centre and the edge of the large FOV. For the filter to operate, all 3 small FOV positions would need to simultaneously pass more light than the large FOV at a particular radius. The results are shown in figure 3.
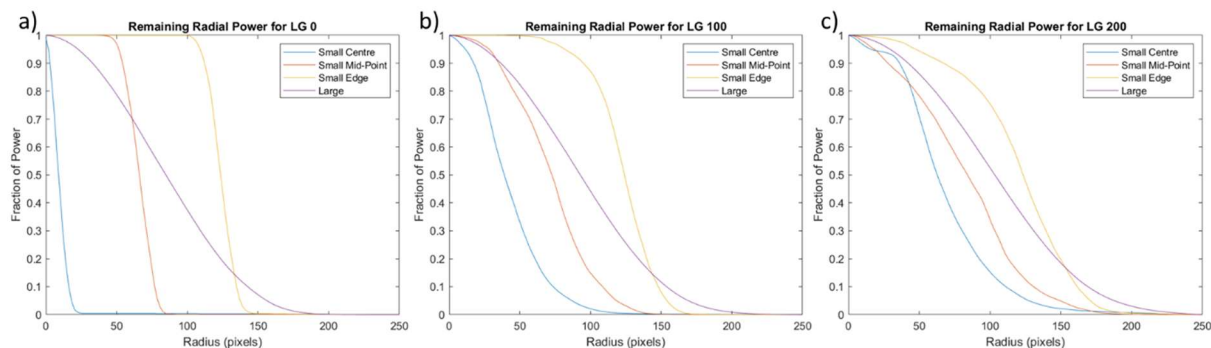
Figure 3. The remaining radial power for light of different spatial coherences (233μm for small and 18.9μm for large) and different positions in the FOV for a) LG 0, b) LG 100, c) LG 200

It can be seen that the spatial coherence filter does not work as there is no pixel radius value in any of the graphs where all 3 positions of the small spot are able to get above the large FOVs value. This was true when tested with topological charges of 0, 10, 20, 50, 100, 200 and 400. One reason for this is that low spatial coherence light also becomes ring-like, particularly for higher topological charges. The "noise" thus tends to chase the signal, reducing the effectiveness of such an approach. A second reason is that the signal cannot get far enough outside the central noise spot, resulting in only a small fraction of the total noise FOV being blocked. Attempts were made to improve the size of the ring using a Fresnel lens mask, in line with the approach reported by Zhou et al, but we found that this lens interacted to the same extent for light regardless of its coherence (which is expected given it is a refractive process rather than an interference based one).

5. DISCUSSION

There are various reasons why the spatial coherence filter reported in this paper fails in comparison with the previous. In [1], "Enhanced underwater ranging using an optical vortex", the key benefit of the system is in reducing the return time error, rather than specifically as an SNR improvement. Such an application would be unlikely to be beneficial in most free-space LIDAR applications, where forward scatter makes up a relatively small fraction of the noise.

In [3], "Numerical simulation model of an optical filter using an optical vortex", it is unclear how they were able to achieve the improvement using a Fresnel lens mask on the SLM. As discussed earlier, a Fresnel lens exploits refraction rather than interference to produce its focussing effect.

In other works [2][4], the spatial coherence filter ultimately acts as an FOV filter, where the position of the object is known. It is unclear what the benefits are of such an approach compared to a simple iris and lens.

There could be some viability with this spatial coherence filter if it could be paired with a circle detection scheme. This would necessitate computationally heavy algorithms however, and the benefits in terms of the SNR are unclear (e.g. the SNR is better with a single strong measurement, rather than multiple weak measurements with the same total).

6. CONCLUSION

Filtering the strong background noise is essential in long-range scanning LIDAR. A potential means of doing this is to exploit spatial coherence to enable spatial filtering, as described in publications by other authors. This approach exploits the higher spatial coherence of a reflected signal relative to the background noise that would make it prone to experience interference effects.

The free-space scanning LIDAR arrangement explored in this paper is however different from the LIDAR arrangements and applications of the earlier authors, and so we have aimed to explore how such an approach can be adapted to the new environment. The work described here used an SLM to generate an LG ring, with the idea being that the low coherent noise light would focus to a central spot, whilst the high coherent signal light would form a ring outside this. The noise would thus be blocked, whilst the signal light would be able to pass.

However, our results show that such a filter does not work with the free-space scanning LIDAR arrangement. We showed that the incoherent noise was affected similarly to the coherent signal in that both generated an LG ring in the far field thus making it impossible to spatially filter out the noise.

7. REFERENCES

[1] Jantzi, A. W., Jemison, W. D., Illig, D. W., & Mullen, L. (2021). Spatial and temporal domain filtering for underwater lidar. Journal of the Optical Society of America, 38(10), B10. https://doi.org/10.1364/josaa.430543

[2] Zhou, Y., Li, X., Yin, Z., Yi, Y., Wang, L., Wang, A., Mao, S., & Wang, X. (2022). Numerical simulation model of an optical filter using an optical vortex. Optics Express, 30(20), 36235. https://doi.org/10.1364/oe.466181

[3] Li, X., Wang, X., Yi, Y., Zhou, Y., Chen, Q., Wang, A., Mao, S., & Yan, Y. (2023). Separation of coherent and incoherent light by using optical vortex via spatial mode projection. Optics Communications, 527, 128986. https://doi.org/10.1016/j.optcom.2022.128986

[4] Jantzi, A. W., Jemison, W. D., Laux, A., Mullen, L., & Cochenour, B. (2018). Enhanced underwater ranging using an optical vortex. Optics Express, 26(3), 2668. https://doi.org/10.1364/oe.26.002668

[5] Arnone, R.A., et al. (2018), An optical vortex transmissometer, in *Ocean Sensing and Monitoring X*.

[6] Zhou, Y., Li, X., Yin, Z., Yi, Y., Wang, L., Wang, A., Mao, S., & Wang, X. (2022). Numerical simulation model of an optical filter using an optical vortex. Optics Express, 30(20), 36235. https://doi.org/10.1364/oe.466181

[7] Hecht, E. (1998). *Optics (5th Edition)*, 2022, Pearson+, ISBN 978-0-137-52642-0

[8] Lahiri, A., *Basic Optics*. 2016, Elsevier. p. 309-384, ISBN 978-0-128-093078

[9] Francon, M., *Laser Speckle and Applications in Optics*. 1979, Academic Press: New York. p. 10-16. ISBN 0-12-265760-8

[10] Dainty, C*., ed. (1984). Laser Speckle and Related Phenomena (2nd ed.). Springer-Verlag. ISBN 978-0-387-13169-6*.

[11] Shen, Y., Wang, X., Xie, Z., Min, C., Fu, X., Liu, Q., Gong, M., & Yuan, X. (2019). Optical vortices 30 years on: OAM manipulation from topological charge to multiple singularities. Light-Science & Applications, 8(1). https://doi.org/10.1038/s41377-019-0194-2

[12] David Fischer (2023). radialavg.zip (https://www.mathworks.com/matlabcentral/fileexchange/46468-radialavg-zip), MATLAB Central File Exchange. Retrieved August 15, 2023.

Gated Viewing at 2.09 µm laser wavelength – Experimental system assessment and comparison to 1.57 µm

Benjamin Göhler*[a], Peter Lutzmann[a], Dominik Walter[a], Helge Bürsing[a],
Johannes Deutsch[a,c], Marc Eichhorn[a,c], Christelle Kieleck[a],
Alexander Sieck[b], Robert Wiegleb[b]

[a]Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB),
Gutleuthausstraße 1, 76275 Ettlingen, Germany;
[b]AIM Infrarot-Module GmbH, Theresienstraße 2, 74072 Heilbronn, Germany;
[c]Institute of Control Systems (IRS), Karlsruhe Institute of Technology (KIT),
Fritz-Haber-Weg 1, 76131 Karlsruhe, Germany

ABSTRACT

We have set up a Gated Viewing (GV) system operating at a laser wavelength of 2.09 µm in the short-wave infrared (SWIR) spectral range to experimentally assess the potential of such a system for security and military applications like long-range target identification and intelligence, surveillance and reconnaissance (ISR) in low visibility conditions. In particular, we compare this system with GV systems operating at the widely used SWIR wavelength of 1.57 µm. Our focus is on examining physical effects such as laser reflection and speckles at the target surface as well as atmospheric impacts like transmission and turbulence. Finally, estimates of system ranges are made.

The gated viewing camera is based on an array of 640×512 mercury cadmium telluride (MCT) avalanche photodiodes (APD) with a pitch of 15 µm. The cut-on and cut-off wavelengths are 0.9 µm and 2.55 µm, respectively, providing sensitivity in the extended SWIR (eSWIR) spectral range. This allows to capture both laser wavelengths 1.57 µm and 2.09 µm with the same GV camera. The camera is equipped with an aspherical F/3 lens with a focal length of 600 mm, resulting in a field-of-view (FOV) of $0.92° \times 0.73°$.

The 1.57 µm laser is based on a commercial flashlamp-pumped Nd:YAG laser combined with an optical parametric oscillator (OPO) with a maximal pulse energy of 65 mJ at 20 Hz pulse repetition frequency (PRF) and a pulse width of $\tau = 11$ ns. The 2.09 µm laser is an in-house developed solution with approximately 20 mJ at 20 Hz PRF and $\tau = 12$ ns.

Keywords: Laser Gated Viewing, Range-Gated Imaging, Active Imaging, Laser Wavelength, Short-Wave Infrared (SWIR), Poor Visibility Condition, Contrast Enhancement, 3D Imaging

1. INTRODUCTION

In recent years, a lot of research has been conducted and published on the topic of gated viewing (GV) for long-range reconnaissance in military and security applications ([1]-[5]). Typically, the laser wavelength used for such GV systems is around $\lambda = 1.57$ µm in the short-wave infrared (SWIR) ranging from 1 to 3 µm. The two main reasons for this are that, on the one hand, this wavelength can be utilized to achieve long system ranges of several kilometers despite compliance with laser safety regulations ($\lambda > 1.4$ µm) and, on the other hand, high-power laser sources at this wavelength are easily available due to a variety of compact and efficient technical solutions such as flashlamp- or diode-pumped solid-state Nd:YAG lasers, Er fiber lasers and laser diodes.

*benjamin.goehler@iosb.fraunhofer.de; phone +49 7243 992-260; fax +49 7243 992-299; www.iosb.fraunhofer.de

Considering the detector side, there are two main technical concepts providing sensitivity for this laser wavelength. Gated SWIR cameras based on Indium Gallium Arsenide (InGaAs) detector arrays are available with up to 640×512 elements currently ([6]) and EBCMOS GV cameras based on InGaAs photocathodes with internal gain provide high sensitivity in the SWIR with a quantum efficiency above 25% at 1.55 µm ([7]). InGaAs cameras with internal gain forming an avalanche photodiode (APD) detector array are also suitable for this purpose ([8]).

We investigated alternative wavelengths with comparable properties that are also feasible for a GV system. With respect to the atmospheric transmittance and the possibility of technical realization of laser system, the laser wavelength λ = 2.09 µm was chosen in this research. It offers the following advantages:

- Comparable atmospheric transmittance to 1.57 µm (approximately 96% for a standard atmosphere and a horizontal path with 1 km length in Figure 1)

- Possibility of technical realization by pumping a Ho^{3+}:YAG crystal with a continuous-wave (cw) Tm^{3+}:fiber laser and combining it with an acousto-optic modulator (AOM) for pulsed operation (see section 2.2)

FASCODE / HITRAN

Model Atmosphere = "MidLatitude Summer"
Aerosol Model = "Rural - VIS=23km"
Path Length = 1 km (horizontal path 1 m above ground)
Mixing Ratio for CO2 = 420ppmv
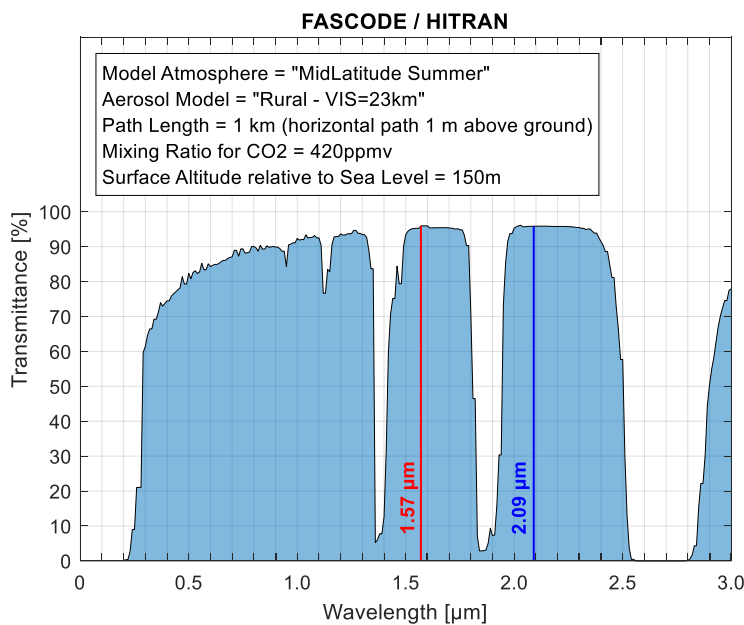Surface Altitude relative to Sea Level = 150m

Figure 1: Atmospheric transmittance as a function of the wavelength for a model standard atmosphere and a horizontal path with length 1 km calculated with the software FASCODE from the database HITRAN (Ontar Corp.)

InGaAs based cameras are not sensitive for a laser wavelength of 2.09 µm due to their typical cutoff wavelength of about 1.7 µm. Therefore, we have utilized a SWIR GV camera developed at AIM in Germany based on mercury cadmium telluride (MCT or HgCdTe) APDs ([9]) identical to the InGaAs based one in [8] mentioned above except for the detector material. Thanks to a suited mercury-to-cadmium ratio, the photodiodes of this detector provide a cutoff wavelength of 2.55 µm, thus allowing the detection of both wavelengths 1.57 µm and 2.09 µm. In previous work, we have already used this MCT-APD based SWIR GV camera ([10]-[12]).

For the purposes of experimental system assessment and investigation of wavelength dependent physical effects, we have set up a SWIR GV system consisting of the MCT-APD based GV camera and two lasers at 1.57 µm and 2.09 µm wavelength for sequential illumination of the scene.

In chapter 2, the components and the set-up of this SWIR GV system are described and their specification data are given – in sections 2.1 and 2.2 for the illuminations lasers, in section 2.3 for the SWIR GV camera and in section 2.4 for the entire integrated SWIR GV system.

2. SYSTEM SET-UP AND SPECIFICATIONS

2.1 Illumination laser at 1.57 µm wavelength

The laser illuminator at 1.57 µm wavelength is based on the commercially available pulsed lamp-pumped Nd:YAG laser CFR400 from Lumibird Quantel Lasers modified by the technique for homogeneous speckle-reduced illumination presented in [13]. The main laser parameters are listed in Table 1.

Table 1: Relevant parameter of the illumination laser at 1.57 µm wavelength

Parameter	Value / Property	Remark(s)
Wavelength	1.57 µm	-
Pulse Energy	33 mJ	measured after beam shaping, originally 65 mJ
Repetition Rate	20 Hz	max.
Pulse Width	~7 ns	measured FWHM
Beam Diameter	< 7 mm	at laser output
Beam Divergence	$1° \times 0.8°$	after homogenization waveguide, adapted to the camera field-of-view (see section 2.3)
Beam Profile	Rectangular Flat-top	
Cooling Type	Water-to-Air	-
Laser Class	4	calculated with software "LaserSafe PC Professional" (Ver. 5.1) for the above parameters
N.O.H.D.	1.2 m	
Extended N.O.H.D.	11.1 m	

The laser is mounted next to the GV camera on a common tripod. They were aligned with each other so that the illuminated area matches to the field-of-view of the camera at the measuring distance.

2.2 Illumination laser at 2.09 µm wavelength

The 2.09 µm illumination laser is an in-house development at IOSB. It is realized on a 600 mm × 400 mm breadboard which is encapsulated from the environment by a custom manufactured housing, providing protection against environmental influences. The board is mounted on a separate tripod allowing the illumination laser to be lined up according to the GV camera. Including housing and components, the overall weight is below 46 kg. To maintain a constant temperature in sealed operation, all beam dumps are actively cooled. Output coupling from the housing is realized by an anti-reflection (AR) coated window.

The 2.09 µm illumination laser is realized as an actively Q-switched Ho^{3+}:YAG laser, optimized for short, high-energy pulses. A schematic drawing of the laser is shown in Figure 2. For pumping, a commercially available Tm^{3+}-doped fiber laser with a beam quality of $M^2 < 1.1$ and a spectral emission at 1908 nm is used. The pump laser is equipped with a collimator and the pump beam is focused by a telescope consisting from the lenses L_1, L_2 and L_3 into the Ho^{3+}:YAG crystal with a beam diameter of 1.1 mm. The pump beam is coupled into the cavity with a 45° dichroic mirror which is highly transmissive for the p-polarized 2090 nm light. This ensures the linearly polarized output.
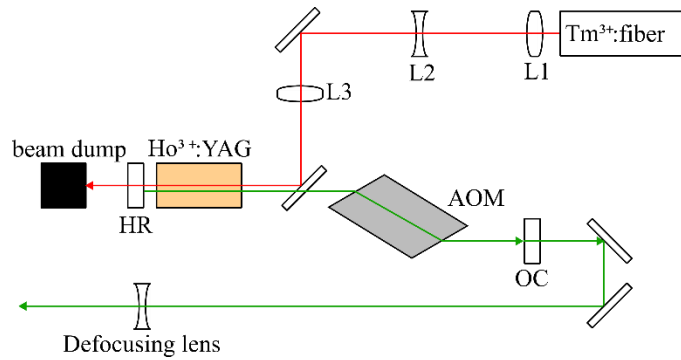
Figure 2: Setup of the 2.09 µm illumination laser

The cavity is built by a dichroic highly-reflective (HR) mirror next to the Ho^{3+}:YAG crystal that transmits the residual pump light into a beam dump. An output-coupler (OC) with a relatively low reflectivity of 30% is used to reduce photon density inside the cavity. Q-switching is realized by a Brewster-cut AOM. The beam divergence is adjusted to 2° (1/e^2 beam diameter) by an additional defocusing lens.

With the cavity primarily designed for high pulse energy, the set point of operation for the 2.09 µm illumination laser provides Gaussian shaped pulses with a repetition rate of 20 Hz and a pulse energy of 19.9 mJ. The laser pulses are characterized at a pump power of 32.4 W and 20.9 W. For illumination, the laser was optimized and operated at the high-energy setting. In both points of operation, the beam quality factor is below $M^2 < 1.2$ for both axes. Due to a higher inversion at a pump power of 32.4 W, the pulses were considerably shorter and the pulse delay was decreased compared to the operation point of 20.9 W of pump power (see Table 2).

Table 2: Characteristics of the 2.09µm illumination laser at different points of operation

Pump power:	32.4 W	20.9 W
Pulse energy:	19.9 mJ	5.7 mJ
Pulse width (FWHM):	12 ns	54 ns

For measuring these characteristics, the pulses were recorded by a fast photodetector and a fast oscilloscope was used to measure the trigger and pulse signal.

Since the excited state lifetime of $\tau_f \approx 7.9$ ms of the Ho^{3+}-ions ([14]) is shorter than a pump interval at the repetition rate of 20 Hz, the gain medium is saturated prior to every pulse. Consequently, reductions of the repetition rate will maintain pulse energy, peak power and pulse width while the average power scales proportionally to the repetition rate.

In the opposite regime, $f_{rep} \gg 1/\tau_f \approx 127$ Hz holds and it is characterized by a constant average power, a proportionality of the pulse duration with the repetition rate and an inverse proportionality of the pulse energy with the repetition rate. For Ho^{3+}:YAG lasers, comparable designs show a transition into the quasi-continuous regime at around 2 kHz ([15]). Due to then different thermal lensing power inside the Ho^{3+}:YAG crystal and shorter pump phases at higher repetition rates, the laser design needs some adjustments in order to provide good mode matching and high pulse energies. Although the highest pulse energies being more easily generated at lower repetition rates, as longer pump times allow for higher inversions, high multi-mJ output pulse energies are possible also at high repetition rates when needed for a given application.

2.3 SWIR GV camera based on MCT APD array

GV Focal-Plane-Array (FPA)

The core of the GV active imaging system is the detector module providing sufficient sensitivity and precise gating capability. The Readout-Integrated-Circuit (ROIC) developed at AIM features a capacitive-transimpedance-amplifier (CTIA) in the input stage and provides four analog output channels with a clock rate of 10 MHz per channel. The format of the GV-ROIC is 640×512 pixels with 15 µm pixel pitch. The charge handling capacity (CHC) is about 120.000 e- and the readout noise with no correlated double sampling (CDS) is about 80 e- ([16]).

In GV mode the integration phase is executed first and the pixels are read out afterwards. The integration time is synchronized to an external gating signal by sampling the gating signal with the internal clock. The internal clock is generated by an on-chip phase-locked-loop (PLL) derived from an external master clock. The maximum frequency of the internal clock is 200 MHz providing a time base of 5 ns for internal gate control of delay and integration, corresponding to a depth control in steps of 75 cm. The pixel clocks controlling the integration time are synchronized to the internal clock. Gate control can either be done internally by programmable registers or by external signals. In GV mode a maximum frame rate of 100 Hz, if only the signal level is read out, or of 50 Hz, if also the reset level is read out for CDS, can be achieved.

The avalanche photodiodes (APD) in the array are based on AIM's standard n-on-p HgCdTe based technology for SWIR detectors. The doping profile has been optimized to enable high reverse bias voltages. For HgCdTe-based photodiodes with a cutoff wavelength of 2.55 µm an avalanche multiplication factor of up to 20 at a reverse bias of 14 V can be achieved ([9], [16]). With longer cutoff HgCdTe-material even higher multiplication factors are possible at a comparable reverse bias but lower operating temperature. Due to the specific bandstructure of HgCdTe the APDs exhibit a close to constant exponential gain increase with respect to the reverse bias and the excess noise factor is close to unity and almost independent of the gain. For 2.55 µm cutoff wavelength the dark signal limited operation temperature is about 170 K.

The GV FPA is built into a standard integrated dewar/cooler assembly (IDCA) with an AIM SX040 split linear microcooler and a dedicated command and control electronics (CCE). A separate trigger board allows synchronisation with the laser illuminator. The quantum efficiency η of the IDCA is relatively constant at a value of 75 % ± 5 % over the whole spectral range from 900 nm to 2400 nm. The quantum efficiency for the two laser wavelengths are nearly equal with values of approximately $\eta_{1.57\mu m} = 74$ % and $\eta_{2.09\mu m} = 75$ %. The small difference is due to the design of the AR-coating on the FPA. The reflection is slightly different at the two wavelengths.

GV Field-Demonstrator

To demonstrate the long-range reconnaissance capability of the GV detector a field demonstrator has been developed. The design goal was a spatial resolution of 2.5 cm on objects in a distance of 1000 m. A 3D CAD sketch of the SWIR GV camera head is depicted in Figure 3.
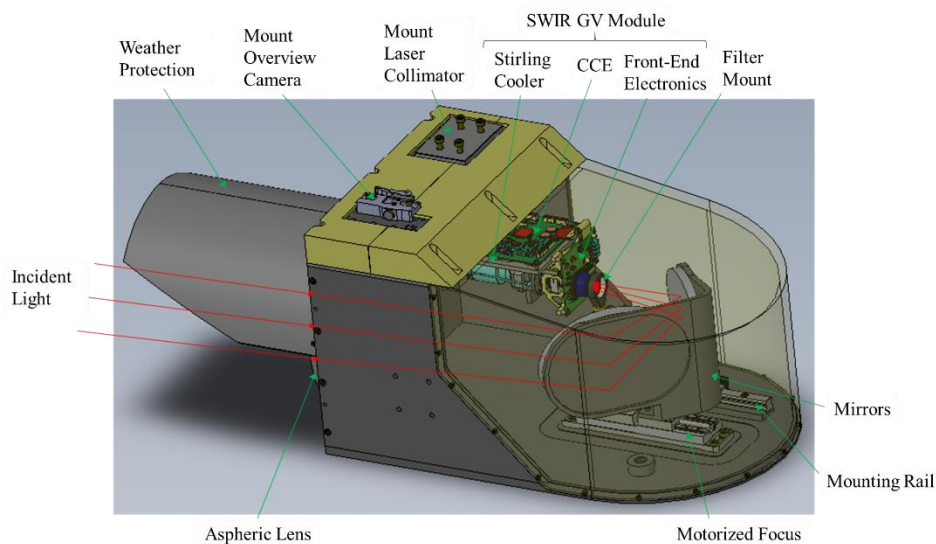

Figure 3. 3D CAD sketch of the camera head of the developed SWIR GV field demonstrator

It consists of one aspheric lens with F# 3 and focal length f = 600 mm. The lens focusses the incident light by the two mirrors onto the SWIR GV module. The folded optical path assures a rather compact housing. A filter mount for a band pass filter is placed in front of the SWIR GV module in order to block any light with a wavelength different from the selected laser wavelength. The two mirrors are attached to a mounting rail, which can be moved over a motorized stage to focus to distances up to infinity.

The given detector format and pitch size of 640×512 and 15 µm and the focal length of f = 600 mm result in a narrow field of view (NFOV) of approximately $0.92° \times 0.73°$. At the top of the housing, a mount for the laser collimator optics is available. Optionally, an overview camera, e.g. with a wider field of view than that of the GV camera, can be placed on top of the GV camera.

2.4 Integrated SWIR GV system

Figure 4 gives an impression of the experimental set-up of the entire SWIR GV system with the two lasers at 1.57 µm and 2.09 µm wavelength and the SWIR GV field demonstrator described in the previous sections.
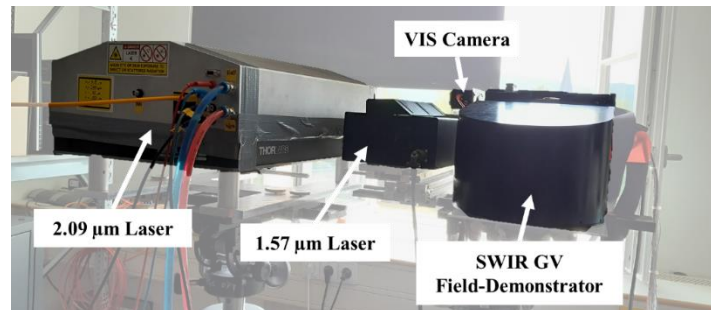

Figure 4. Image of the experimental set-up of the SWIR GV system mounted on two tripods and directed towards the measurement site through an open window in our lab

The two different laser wavelengths have been subsequently used for scene illumination with a time interval of some few minutes (~2-3) in between for the following system modifications:

1. **Adapting the trigger wiring, settings and modes:**

   a. For 1.57 µm, the GV camera as a master triggers the flashlamp of the laser, which in turn triggers internally the Q-switch after 168 µs delay time. Then the Q-switch OUT trigger of the laser is used for the GV camera as synchronization signal to provide a minimal jitter of only about 1 ns.

   b. For 2.09 µm, the GV camera also as a master triggers the AOM of the laser for pulsed operation. Then, the GV camera is switched active after a fixed time interval which is based on the low-jitter delay time of the laser of approximately 1 µs until the laser pulse is emitted.

2. **Changing the bandpass filters in front of the SWIR GV module:**

   a. For 1.57 µm, two bandpass filters are used simultaneously. First, a narrow spectral filter at 1575 nm with a FWHM of 50 nm and a transmission of almost 100 % is used to block any other wavelength down to 200 nm and up to 1800 nm. Second, a bandpass filter at 1550 nm with a FWHM of 255 nm and a transmission of 95 % is used to block wavelengths up to 2.6 µm above the detector cutoff wavelength.

   b. For 2.09 µm, a bandpass filter with a window at 2-2.5 µm and a transmission of 90 % is used to block any other wavelength of the detector sensitivity range. Unfortunately, a filter with a narrower transmission window was not available at the time of the measurements.

3. **Refocusing of the receiver optics to account for the different laser wavelengths**

3. RESULTS

3.1 Church tower at 700 m

We performed the first measurements with the built-up SWIR GV system on a church tower at a distance of about 700 m. The tower at this distance fits quite well in the field-of-view and offers differently reflecting building parts and small details to be resolved. Further the laser beam path length of 1400 m allows investigating the influence of the laser wavelength on the intensity scintillations caused by atmospheric turbulence.

The gate length was set to be 30 m and the frame rate was 20 Hz for all these measurements. The maximal laser pulse energy for each wavelength were used, i.e. E = 33 mJ for $\lambda = 1.57$ µm and E = 20 mJ for $\lambda = 2.09$ µm. The APD gain M of the SWIR GV field demonstrator was set to maximal values under the constraint that the camera image pixels are not saturated, resulting in M = 6 and M = 9, respectively. In total 200 frames were recorded for each wavelength to be able to perform frame averaging for turbulence mitigation and to evaluate statistics with respect to intensity scintillations. In Figure 5 the first frames of these sequences are shown in the top row with corresponding average images of the first 20 frames in the bottom row.
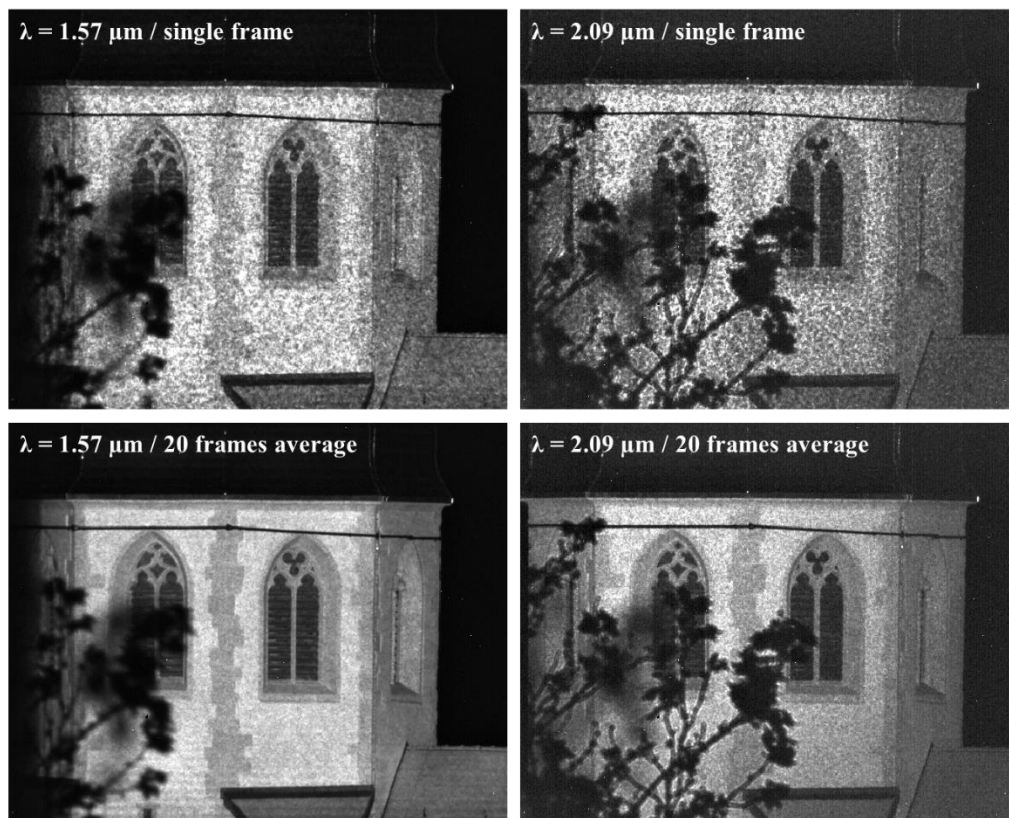
Figure 5. Top row: Single frames of the church tower at 700 m distance captured with the SWIR GV system for the laser wavelengths $\lambda = 1.57$ µm (left) and $\lambda = 2.09$ µm (right). Bottom row: Resulting images after performing the pixelwise average of 20 subsequent frames from the captured sequences to reduce the strong scintillations caused by turbulence.

The single frames in the top row of Figure 5 are strongly affected by turbulence induced scintillations for both wavelengths. By averaging the first 20 frames of the recorded sequences this effect can be significantly reduced resulting in much more homogeneous images in the bottom row. The image quality for $\lambda = 2.09$ µm with respect to the residual speckle pattern is slightly worse due to the speckle-reducing illumination technique used for $\lambda = 1.57$ µm. However, the quantitative image impressions for the two wavelengths are comparable. So, $\lambda = 2.09$ µm is obviously are promising alternative illumination wavelength to $\lambda = 1.57$ µm with this respect.

The study the temporal intensity fluctuations, i.e. the turbulence induced scintillations, image pixels were selected and the gray values of these pixels over the frame number in the recorded sequences were analyzed. Exemplarily in Figure 6, the results for pixel (425|425) which is below the right window of the church tower are depicted.
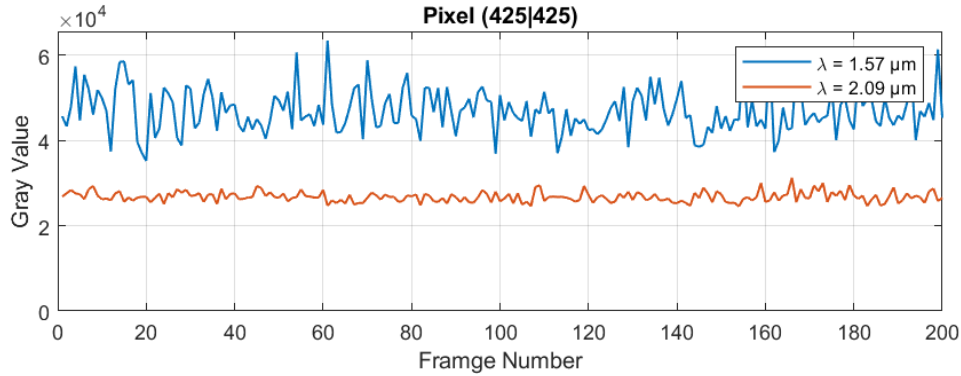

Figure 6. Gray values as functions of the frame number in the recorded sequences for the two laser wavelengths $\lambda = 1.57$ µm (Blue) and $\lambda = 2.09$ µm (Orange).

In the graph in Figure 6, significantly smaller absolute fluctuations can be seen for $\lambda = 2.09$ µm compared to $\lambda = 1.57$ µm. For a relative evaluation, the variances normalized to their corresponding mean values are calculated in Table 3.

Table 3: Temporal statistics for the data in Figure 6

Pixel (425|425)	$\lambda = 1.57$ µm	$\lambda = 2.09$ µm
Standard Deviation (σ)	5179	1216
Mean Value (μ)	46901	26732
Normalized Variance (σ^2 / μ^2)	**0.0122**	**0.0021**

According to the results shown in Table 3, the normalized scintillation variances confirm the trend of a decrease of the scintillation for longer wavelengths as well. For a more reliable result, this evaluation is performed for all pixels in a region-of-interest (ROI) of size 50×50 below the right window of the church tower. The results are shown in the histogram in Figure 7.
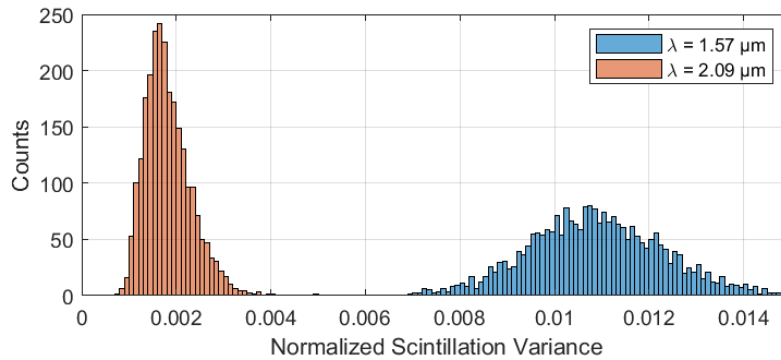
Figure 7. Histogram of the normalized scintillation variance for all ROI pixels for the two laser wavelengths $\lambda = 1.57$ µm (Blue) and $\lambda = 2.09$ µm (Orange).

The mean values of the normalized scintillations variances in Figure 7 over all ROI pixels are:

$$\sigma^2_{1.57\,\mu m} = 0.0108 \quad \text{and} \quad \sigma^2_{2.09\,\mu m} = 0.0017$$

This trend is also consistent with the expected behavior from theory, i.e. $\sigma^2 \propto \lambda^{-7/6}$ (section 2.3.4 in [17]), although the numerical extend of the decrease is slightly larger than expected.

3.2 Laser reflectance of persons – Human skin and clothes

In this section, investigations on the laser reflectance of persons with respect to the human skin and some cloth materials are made. Therefore, we have used four of the six reference targets shown in Figure 8. They have diffusely reflecting surfaces with predetermined hemispherical reflectance values for the calibration of measured pixel gray values.

Figure 8. Diffuse reflectance targets with nominal reflectance values used for calibration

In Table 4, the actual 8° hemispherical reflectance values of the calibration targets are listed for the two laser wavelengths $\lambda = 1.57$ µm and $\lambda = 2.09$ µm according to their calibration certificates.

Table 4: 8° hemispherical reflectance values of the used reference targets for the two laser wavelengths

Target	$\lambda = 1.57$ µm	$\lambda = 2.09$ µm	Target	$\lambda = 1.57$ µm	$\lambda = 2.09$ µm
5%	4.688%	4.720%	50%	47.053%	45.453%
10%	9.196%	9.202%	90%	(not used)	(not used)
20%	24.614%	25.196%	95%	(not used)	(not used)

The reference targets are placed at a distance of 70 m. The laser pulse energy is set to the maximal value of 19.9 mJ for $\lambda = 2.09$ µm and to an attenuated value of 3.3 mJ (10%) for $\lambda = 1.57$ µm for optimal use of the dynamic range of the GV camera. The gate length is set to 30 m and the camera gain is off (M = 1). For each wavelength, a GV image of the reference targets is recorded and all pixels in ROIs of the targets indicated by yellow rectangles in Figure 10 are averaged. In the graph in Figure 9, the resulting mean values are plotted as circle markers vs. the reflectance values in Table 4. Then linear fits are calculated for each laser wavelength for calibration of the measured pixel gray values to the actual laser reflectance.
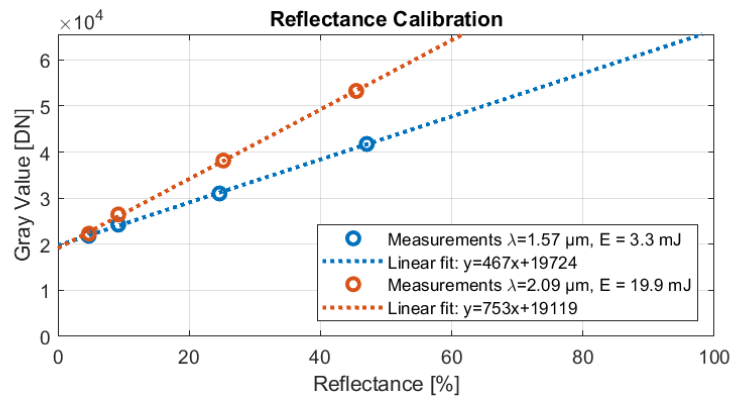

Figure 9. Measurements of the mean pixel gray values (O) in the ROIs of the reference targets and linear fits for calibration

After application of the reflectance calibration, the calibrated images in Figure 10 both scaled to the same reflectance range 0-50% are obtained. In these images, a person is sitting next to the reference targets. The reflectance values averaged in ROIs indicated by yellow rectangles are given for the reference targets as well as for the person's head and clothes.
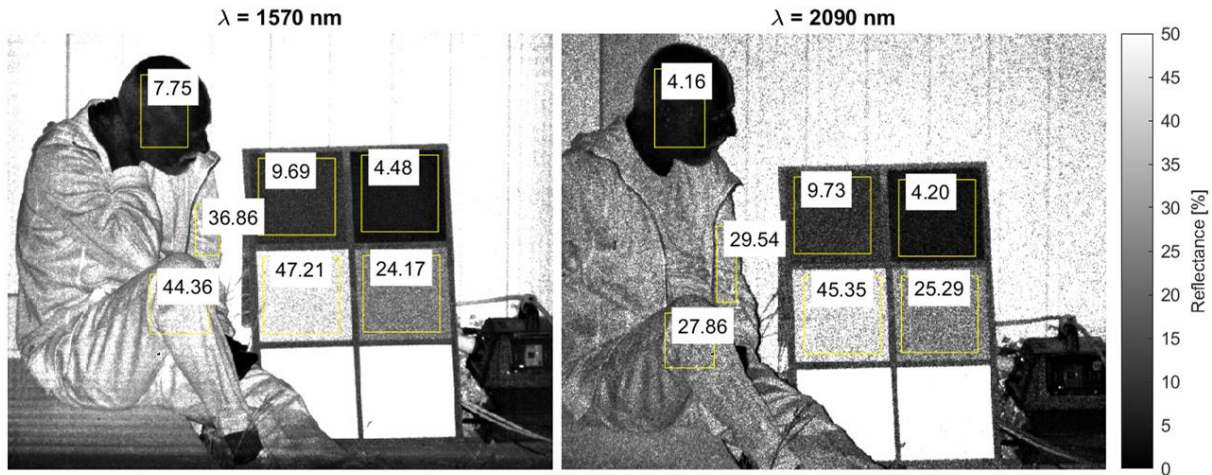
Figure 10. Calibrated GV images for the laser wavelength λ = 1.57 µm (left) and λ = 2.09 µm (right) displayed for reflectance values in the range 0-50% showing a person next to the reference targets with resulting ROI reflectance values

By comparing the averaged ROI reflectance values for the reference targets in Figure 10 with the actual values in Table 4, a very good agreement can be seen, verifying the correctness of the calibration process. Further, significantly lower reflectance values of the human skin at the person's head (~4% vs. ~8%) and the person's clothes (~29% vs. ~40%) for the laser wavelength λ = 2.09 µm compared to λ = 1.57 µm can be observed. The relative decrease is approximately 46% and 20-37%, respectively. Figure 11 shows enlargements of the upper left sections of the above images both with enhanced contrast to reflectance values in the range 0-25% providing more image details.
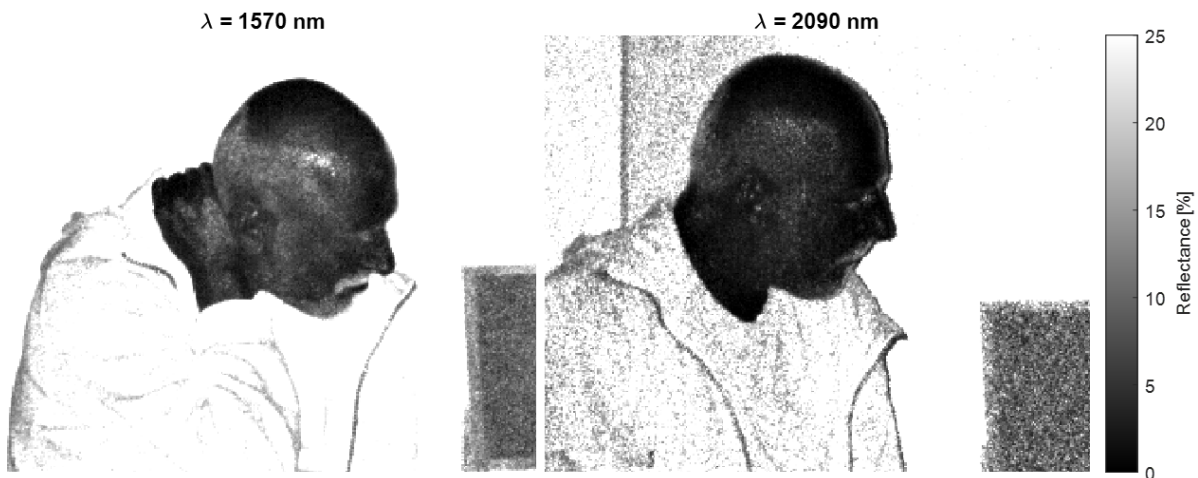
Figure 11. Enlarged sections of the images in Figure 10 with reflectance values in the range 0-25%

The reason for the lower reflectance of the human skin for the laser wavelength λ = 2.09 µm compared to λ = 1.57 µm is the strong absorption by water near 1.9 µm wavelength (see also Figure 13 in the next section on reflectance of vegetation).

3.3 Laser reflectance of vegetation – Tree trunks and leaves

In this section, the laser reflectance of vegetation for the laser wavelength λ = 2.09 µm compared to λ = 1.57 µm is studied. Therefore, we have recorded a chestnut tree at a distance of 100 m with the SWIR GV system operated with the same laser pulse energies as in the previous section. So, the same calibration curves are valid and can be used for reflectance calibration because the difference in target distances of 30 m (100 m vs. 70 m) is neglectable. Again, the gate length is set to 30 m and the camera gain is turned off (M = 1). The results are shown in Figure 12. For the reason of a better comparison, both reflectance images are scaled to the same range 0-30%.
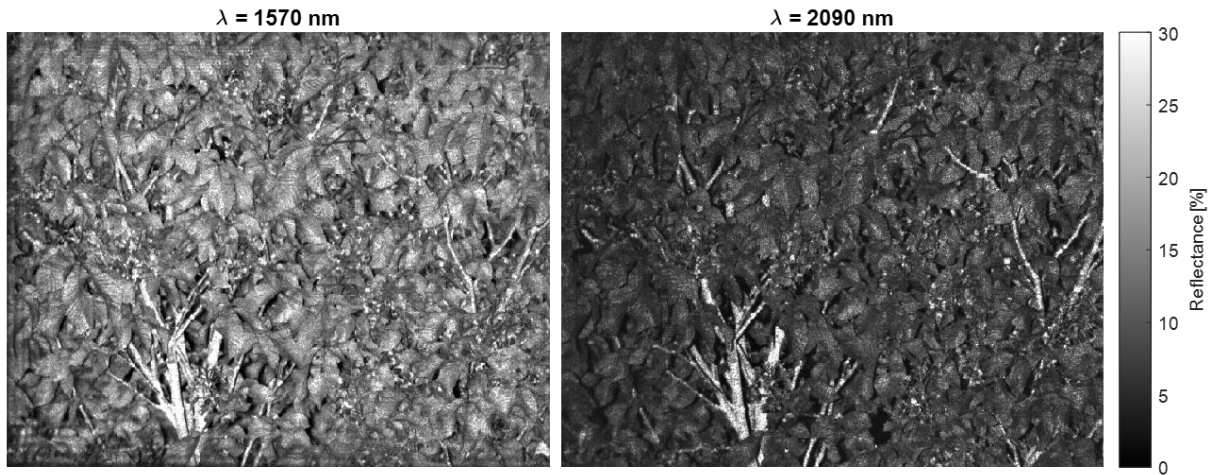
Figure 12. Calibrated GV images of a chestnut tree at a distance of 100 m for the illumination laser wavelength λ = 1.57 µm (left) and λ = 2.09 µm (right) with the same scaling range of 0-30% for the reflectance values

Immediately, a clear difference in the images for the two laser wavelengths can be seen in Figure 12. The overall reflectance in the image for λ = 2.09 µm is significantly lower compared to λ = 1.57 µm. To quantify this, two ROIs on the tree trunk and on a leaf with sizes of 8 × 12 pixels and 21 × 18 pixels, respectively, are defined and again, all pixel gray values in these ROIs are averaged. The results are listed in Table 5.

Table 5: Vegetation reflectance as mean ROI values from the images in Figure 12 for the two wavelengths with statistics

Target	λ = 1.57 µm	λ = 2.09 µm	Difference
Trunk	34.4%	24.8%	-9.6 pp
Leaf	20.9%	10.4%	-10.5 pp
Ratio	1.65	2.38	-

From the statistics in Table 5, it can be seen that the longer laser wavelength results in an absolute decrease of reflectance for both the tree trunk and the leaf by approximately 10 percentage points (pp) from 34.4% to 24.8% and 20.9% to 10.4%, respectively. Further, the trunk-to-leaf ratio which is measure for the contrast is increased by around 44% from 1.65 to 2.38 which is obvious in Figure 12 as well.

An open database of typical vegetation types was used to verify the reflectance results of leaves. Figure 13 shows two curves of leaf reflectance of deciduous and coniferous trees as a function of wavelength. In addition, the above results for λ = 1.57 µm (20.9%) and λ = 2.09 µm (10.4%) are plotted as black markers showing very good agreement with the typical reflectance curves with a slight deviation toward conifers instead of deciduous trees as would be expected.
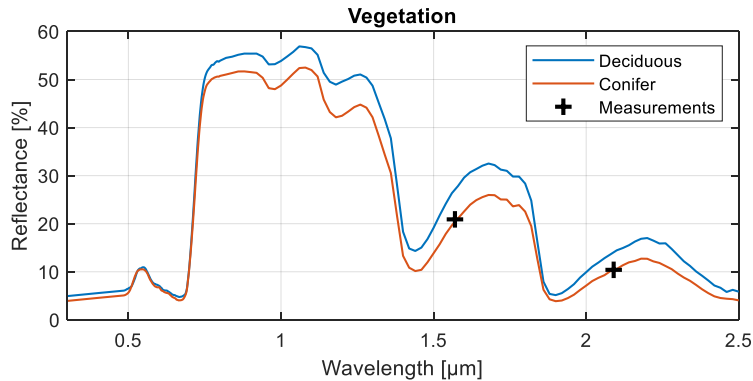
Figure 13. Typical reflectance curves of deciduous and coniferous trees as a function of wavelength over the visible and short-wave infrared spectral range. The above reflectance results of the leaves of a chestnut tree are plotted as black markers. (typical reflectance data for deciduous and coniferous trees were taken from [18], own illustration)

The significant drops in the reflection curves at wavelengths around 1.45 µm, 1.9 µm and 2.5 µm are due to the strong absorption of water and carbon dioxide in the atmosphere as already evident from the atmospheric transmittance curve in Figure 1.

3.4 Long-range tests at 2 km distance

Finally, the built-up SWIR GV system was tested for the use at long distances. A tree at a distance of approximately 2 km was selected as target. For both laser wavelengths, the gate length is set to 37.5 m, a camera gain of $M = 9$ is used and 20 frames are recorded for frame averaging. Due to the broadness of the used bandpass filter especially for $\lambda = 2.09$ µm with a bandwidth of 500 nm, strong proportion of passive radiations compared to the low laser signal from the target at this distance is expected. Therefore, a dark frame with no laser illumination is pixelwise subtracted from the actual target images. Figure 14 shows the averaged and offset-corrected image results of the tree. The scaling of the displayed gray values is different for the two wavelengths.
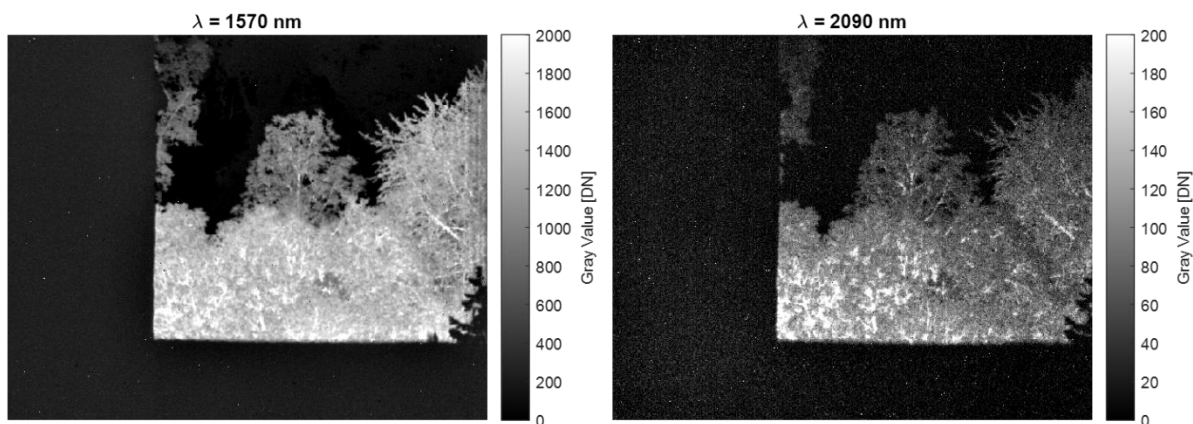
Figure 14. Averaged (20 frames) and offset-corrected GV images of a tree at a distance of 2 km for the laser wavelengths $\lambda = 1.57$ µm (left) and $\lambda = 2.09$ µm (right)

In both GV images of Figure 14, the tree at 2 km distance and details such as the tree trunk and the leaves are clearly recognizable. However, for the laser wavelength $\lambda = 2.09$ µm, the received target signal is significantly lower than for $\lambda = 1.57$ µm. Considering the maximal gray values of the displayed ranges of 2000 compared to 200 and the target

brightness in the images, it is obvious that the signals differ relatively by a factor of >10. This high difference needs further explanation because it is noticeably greater than expected:

- factor of 1.7 due to the pulse energy (33 mJ vs. 19.9 mJ)

- × factor of 1.1 due to the filter transmission (95% vs. 90%)

- × factor of 1.0 due to the quantum efficiency (74% vs. 75%)

- × factor of 2.0 due to the leave reflectance (20.9% vs. 10.4%)

- = total factor of **3.7**

Concerning this discrepancy and the significantly lower target signal in Figure 14, it is expected that the range of a SWIR GV system operating at λ = 2.09 µm is noticeably shorter compared to λ = 1.57 µm for comparable system parameters.

4. CONCLUSIONS

To the best of our knowledge, this is the first time that an MCT-APD based SWIR GV system with a resolution of 640×512 pixels operating at both laser wavelengths 1.57 µm and 2.09 µm has been set up, and that a proof-of-principle for field measurements at distances up to 2 km for direct comparisons of the two wavelengths has been made.

The results show that λ = 2.09 µm is a promising alternative to the widely used laser wavelength λ = 1.57 µm for active SWIR imaging systems.

Further, we experimentally confirmed known theoretical impacts of the longer wavelength on physical effects such as turbulence-induced scintillation or vegetation reflection in their tendency, although the absolute numerical values still require explanations in some cases in future work.

ACKNOWLEDGEMENTS

REFERENCES

[1] B. Göhler, P. Lutzmann, "Review on short-wavelength infrared laser gated-viewing at Fraunhofer IOSB," Opt. Eng. 56(3), 031203 (2016); doi: 10.1117/1.OE.56.3.031203

[2] B. Göhler, P. Lutzmann, "SWIR laser gated-viewing at Fraunhofer IOSB," Proc. SPIE 9649, Electro-Optical Remote Sensing, Photonic Technologies, and Applications IX, 964902 (2015); doi: 10.1117/12.2195932

[3] E. Repasi, P. Lutzmann, O. Steinvall, M. Elmqvist, B. Göhler, G. Anstett, "Advanced short-wavelength infrared range-gated imaging for ground applications in monostatic and bistatic configurations," Applied Optics Vol. 48, No. 31 (2009)

[4] M. Mischung, J. Schmidt, E. Peters, M. Berger, M. Anders, M. Stephan, "Development and characterisation of a portable, active short-wave infrared camera system for vision enhancement through smoke and fog," Proc. SPIE 12271, Electro-optical and Infrared Systems: Technology and Applications XIX, 122710M (2022); doi: 10.1117/12.2636216

[5] F. Christnacher, M. Laurenzis, Y. Lutz, A. Matwyschuk, "60 years of advanced imaging at ISL: from the Cranz-Schardin camera to non-line-of-sight imaging," Proc. SPIE 11160, Electro-Optical Remote Sensing XIII, 1116007 (2019); doi: 10.1117/12.2533902

[6]   New Imaging Technologies (FR), WiDy SenS 640, https://new-imaging-technologies.com/product/widy-sens-640/ (Accessed: August 16, 2023)

[7]   EOTECH, LLC. (US), LIVAR® M506 Gated SWIR Camera System (2022), https://www.eotechinc.com/media/documents/spec_sheets/2022_EOTECH_LIVAR_M506.pdf (Accessed: August 16, 2023)

[8]   F. Rutz, R. Aidam, H. Heußen, W. Bronner, R. Rehm, M. Benecke, A. Sieck, S. Brunner, B. Göhler, P. Lutzmann, "InGaAs APD matrix sensors for SWIR gated viewing," Adv. Opt. Techn. 2019, 8(6), pp. 445–450 (2019); doi:10.1515/aot-2019-0039

[9]   R. Breiter, M. Benecke, D. Eich, H. Figgemeier, H. Lutz, A. Sieck, A. Weber, R. Wiegleb, "MCT SWIR modules for active imaging," Proc. SPIE 11002, Infrared Technology and Applications XLV, 1100212 (2019); doi:10.1117/12.2519891

[10]   B. Göhler, P. Lutzmann, "Super-resolution depth information from a short-wave infrared laser gated-viewing system by using correlated double sampling," Proc. SPIE 10434, Electro-Optical Remote Sensing XI, 104340M (2017); doi: 10.1117/12.2278431

[11]   B. Göhler, P. Lutzmann, "Extending the 3D range of a short-wave infrared laser-gated viewing system capable of correlated double sampling," Proc. SPIE 10796, Electro-Optical Remote Sensing XII, 107960D (9 October 2018); doi: 10.1117/12.2326916

[12]   B. Göhler, P. Lutzmann, "Range-intensity profile of a SWIR laser gated-viewing system for varying temporal response of in-pixel amplification," Proc. SPIE 11160, Electro-Optical Remote Sensing XIII, 1116003 (10 October 2019); doi: 10.1117/12.2535775

[13]   M. Laurenzis, Y. Lutz, F. Christnacher, A. Matwyschuk, J.-M. Poyet, "Homogeneous and speckle-free laser illumination for range-gated imaging and active polarimetry," Opt. Eng. 51(6), 061302 (2012)

[14]   M. Eichhorn, "Quasi-three-level solid-state lasers in the near and mid infrared based on trivalent rare earth ions," Appl. Phys. B 93, pp. 269–316 (2008); doi: 10.1007/s00340-008-3214-0

[15]   M. Griesbeck, H. Büker, M. Eitner, K. Goth, P. Braesicke, M. Eichhorn, C. Kieleck, "High-pulse-energy, Q-switched Ho^{3+}:YAG laser for mid-infrared optical parametric oscillator pumping," Proc. SPIE 11670, Nonlinear Frequency Generation and Conversion: Materials and Devices XX, 116700D (2021); doi: 10.1117/12.2582997

[16]   A. Sieck, M. Benecke, D. Eich, R. Oelmaier, J. Wendler, H. Figgemeier, "Short-Wave Infrared HgCdTe Electron Avalanche Photodiodes for Gated Viewing," U.S. Workshop on Physics and Chemistry of II-VI Materials 2017, J. Electron. Mater. 47, pp. 5705-5714 (2018); doi: 10.1007/s11664-018-6425-0

[17]   F. Smith, J. Accetta, D. Shumaker, "The Infrared & Electro-Optical Systems Handbook. Atmospheric Propagation of Radiation, Volume 2," Infrared Information Analysis Center, Ann Arbor, Michigan, USA & SPIE Optical Engineering Press, Bellingham, Washington, USA (1993)

[18]   https://gsp.humboldt.edu/olm/Courses/GSP_216/lessons/reflectance.html (Accessed: August 21, 2023), Humboldt State University, Arcata, California, USA

Development of a Long Wave InfraRed Photonic Doppler Velocimeter

G. Lefrère*[a,b], J-M. Goujon[b], L. Poffo[b], P. Besnard[b], R. Le Page[b], Y. Barbarin[a]
[a]CEA, DAM, GRAMAT, BP 80200, F-46500 Gramat, France
[b]CNRS, Institut FOTON, Université de Rennes, UMR 6082, F-22305 Lannion, France

ABSTRACT

Photonic Doppler Velocimetry (PDV) has become a gold standard technique in materials impact dynamic loading research offered by its high accuracy and resolution in determining the shock wave speed under extreme conditions (shock, explosion, high pressure, etc...). However, this technique is nowadays mostly restrained to surfaces velocities. On the opposite, Radio-Frequency systems may enhance penetration in specific materials, but at the expense of lower spatial and temporal resolutions. To reach adequate penetration depth at high-speed rate measurements, we propose an innovative long-wave (LWIR) infrared Doppler velocimeter architecture to measure shock waves inside a material, operating at a wavelength near 9.5 µm. The system is currently designed to measure velocities up to 4 km/s, with a 750 MHz bandwidth MCT photodetector. Moreover, the measurement is remotely done using a 300 µm diameter Hollow Core fiber with internal dielectric reflective layers. In order to optimize the signal penetration properties into different materials, a wide tunable quantum cascade infrared laser (IR-QCL) operating in the 8-12 µm region is used. As preliminary results, we present measurements at low-speed (<1 m/s) with different targets materials (copper, aluminum and diffuse reflector) in air and transparent medium, in which the sensitivity has been identified at 9.5 µm. Results show that, despite high attenuation components, the system is able to maintain a suitable fringe contrast to ensure the velocity measurement. Further investigation will concern high speed target measurement and wavelength penetration depth optimization for materials of interest.

Keywords: Photonic Doppler Velocimetry, shock, Quantum Cascade Laser, Long Wave Infrared, Hollow Core fiber

1. INTRODUCTION

In the field of shock compression of condensed matter, the characterization of dynamic mechanical properties of materials relies heavily on obtaining essential physical quantities, such as impact and shock velocities. To address this need, the first velocimeters were introduced by Barker in 1965, enabling the measurement of high-speed phenomena, including the VISAR[1,2]. Then, with the expansion of telecom, development of compact system using singlemode fiber at 1.55 µm appeared. This device is nowadays widely used in many experiments such as laser driven shock[3], particle ejecta[4], plasma detection and explosive detonation testing[5].

More recently, there has been a growing interest in measuring shock waves inside materials. To meet this demand, a radio-interferometer operating at 94 GHz has been developed[6] based on the results of Bel'skii[7]. This device operates in the W-Band where a lot of material and high energetic material are sufficiently transparent. However, the trade-off lies in the compromise on spatiotemporal resolution when compared to a PDV system, which is known for its high-resolution measurements. The signal processing of the radio-interferometer is based on phase in quadrature analysis, which does not allow measuring the high-speed velocity distribution of particles.

The primary objective of this article is to show the development of a system capable of measuring speed in the long wave infrared wavelength range (8-12 µm) through transparent mediums, while maintaining the ease of use associated with the Photonic Doppler Velocimetry (PDV) system. Our study is focused on typical materials used in compression dynamics research, with a particular emphasis on exploring their behavior at 9.5 µm wavelength.

Throughout this article, we present preliminary results obtained at low speeds by evaluation of the system's capabilities in measuring different targets in both air medium and transparent mediums. By demonstrating successful measurements in these scenarios, we aim to establish the feasibility and potential of our developed system for broader applications in shock compression dynamics research and beyond. The combination of long wave infrared wavelength and the ease of use of the

*gregory.lefrere@univ-rennes1.fr

PDV system opens up exciting possibilities for non-intrusive and accurate measurements through LWIR transparent mediums.

2. PRINCIPLE AND METHODS

2.1 Design of the system

The LWIR systems rely on a homodyne PDV system as illustrated in Figure 1. The laser is a tunable Quantum Cascade Laser that emits around 9.5 µm wavelength. The first coupler is an optical window in ZnSe that keeps 5% of the laser power to create the reference arm. The second output, with most of the optical power is injected into a Hollow Core Fiber (HCF). In order to optimize the injection, we use a parabolic mirror (not represented here) to avoid transmission loss. The probing signal is brought to the target by the HCF, which is for now 2 meters long. The amount of signal collected back by the collimator interferes with the reference arm through a 2nd beam splitter, which this time exhibit a 50/50 ratio. The electrical bandwidths of the photodetector at 9.5 µm can be selected in either [DC-260 Hz] or [300 Hz – 750 MHz]. The DC allows estimating the fringe contrast while measuring low velocities and the low cut-off frequency for the second electrical bandwidth suppresses any signal from unwanted interferences when measuring high velocities.
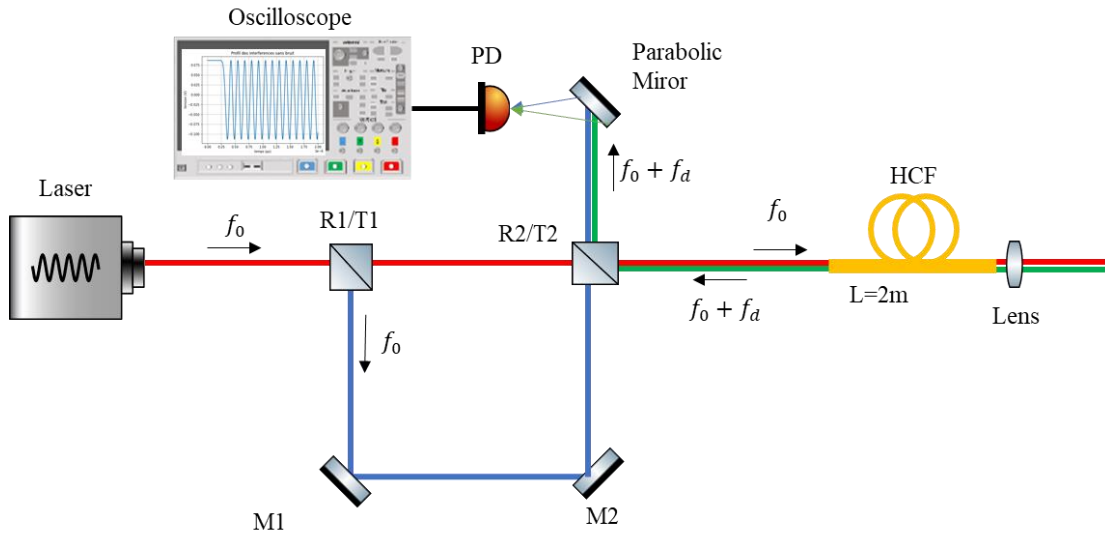
Figure 1. Architecture of the Long Wave Infrared PDV. The red path is the probing arm, the blue path is the reference arm and the green path is the backscattered signal, which carries the target speed information. R1/T1 and R2/T2 are the reflection and transmission coefficients of the two beam splitters. HCF is the Hollow Core Fiber from guiding photonics®. PD is the LWIR photodetector.

We used a QCL that operates throughout the infrared region, covering wavelengths from 3.9 µm to 11 µm. According to the manufacturer's specifications, this QCL exhibits of less than 1 cm^{-1} and a divergence of 4 mrad. Besides, to optimize the efficiency of the laser system, we employed gold-coated mirrors, specifically designed for the mid-IR region. These mirrors are chosen to ensure the highest possible reflection coefficient at these wavelengths. For our beam splitters, we opted for ZnSe material, known for its transparency in the infrared region. The beam splitter was designed with a wedge to prevent unwanted resonating cavity effects, ensuring accurate beam splitting.

Injecting the laser light into the fiber was achieved using a parabolic mirror, which allowed for precise control and direction of the laser beam. We selected a special hollow core waveguide[8] with an inner dielectric coating suitable for efficient light confinement for the LWIR region. The fiber has for now a maximum length of 2 meters and exhibits a theoretical attenuation coefficient of 1 dB/m, ensuring minimal signal loss during transmission. The signal that propagates in this fiber remains single mode as long as the wavelength exceeds 8 µm.

2.2 System calibration

The optical components utilized in the system face significant absorption challenges, which complicate the implementation process. Additionally, there are no fiber components available, such as circulators or couplers, which could offer convenient integration solutions.

Table 1. Optical budget of the Longwave infrared PDV, the backscattered signal depends of the target. The data are measured with a pyroelectric power meter.

Description	% passing	Optical power (mW)	Commentaries
Input system	100%	105.0	At maximum current
Probing arm			
Beam splitter #1	76%	79.6	ZnSe optical window
Beam splitter #2	55%	43.9	
Optical fiber	37%	16.1	Output fiber
Probe	94%	15.1	Incident power on target
Reference arm			
Beam splitter #1	1%	1.1	Fresnel reflection
Mirror #1	98%	1.0	Protected gold
Mirror #2	98%	1.0	Protected gold
Beam splitter #2	55%	0.3	

The combination of the shifted signal from the Doppler effect and the amount of signal from the reference arm will generate the interference signal at the photo receiver given by eq. (1).

$$I(t) = I_{ref} + I_{sig} + 2\sqrt{I_{sig}I_{ref}}\sin(2\pi f_d(t) + \varphi) \tag{1}$$

In this equation I_{ref} represents optical intensity from the reference arm, while I_{sig} represents the optical intensity collected by the probe from the moving target (surface, particles, shock wavefront), the term $f_d(t)$ is the Doppler frequency due to the Doppler Effect, φ is the phase of the signal related to the Doppler shifted light and the reference. In the configuration of our system, the two first terms are obtained when setting the photoreceiver to get the signal DC component. It is useful to estimate the fringe contrast. Finally, the third term includes the velocity information, it is directly linked to the target velocity by

$$f_d(t) = \frac{2}{\lambda} v(t)$$

Where λ is the wavelength of the laser that illuminates the target, and $v(t)$ is its speed.

2.3 Target measurement technique

The experimental setup in the laser lab consists of a target fixed on a motorized translation stage capable of achieving speeds up to 1.5 mm/s. To collimate the beam, we use an aspheric ZnSe AR-coated lens known for its excellent transmission at 9.5 µm. The probe is positioned at distances of 50 mm and 150 mm from the target. The incident power on the target is set at 15 mW.

The experiment involves investigating three different reflective materials: aluminum, copper, and a diffuse reflector with a gold-coated N-BK7 surface of 1500 grit (which is more diffuse). Initially, we conducted the experiments in air medium. For the second set of measurements, we introduce a representative transparent medium in polymer. After careful evaluation, we choose to start with high-density polyethylene (HDPE) material that displays significant transparencies at the 9.5 µm wavelength range[9] and is a material of interest in shock physic[10]. The HDPE sample used in the experiments has a thickness of 2 mm.

To estimate the fringe contrast, we use the [DC-260 Hz] bandwidth of the photodetector. We remove the initial offset from the detector and we use the equation (1) by searching the maximum and the minimum of I(t) The formula for the contrast can be expressed as:

$$c = \frac{V_{pp}}{2V_{mean}} \tag{2}$$

Where V_{pp} is the peak to peak voltage which represents the difference between the higher and the lower value of the voltage. Finally, V_{mean} is the mean voltage of the signal.
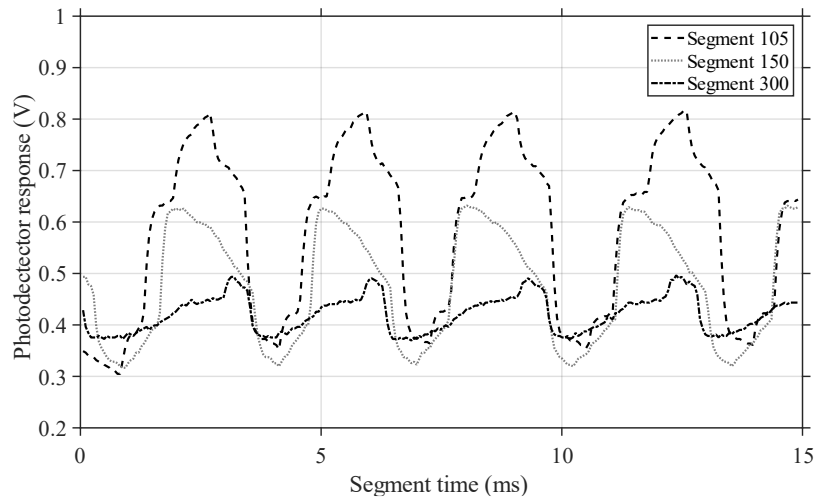
Figure 2. 3 Segments of 4 fringes from an aluminum target signal used to measure the fringe contrast.

The oscilloscope acquires the signal at a sampling frequency of 18.2 kHz, then we proceed with a short time Fourier transform processing to get the velocity information.

3. RESULTS

By utilizing equation (2), we accurately measure contrast across 300 segments to obtain precise results listed in Table (2). We consider the fringe contrast to be satisfactory when it allows us to track the velocity of the moving target. Both aluminum and copper, being reflective materials, exhibit some level of reflection. However, the diffuse reflective nature of these materials leads to lower contrast compared to other materials, but it remains at a satisfactory level.

In the air medium, collecting a significant amount of signal for aluminum and copper is relatively straightforward due to their inherent directivity with a contrast high contrast (>0.3). However, we noticed that introducing a 2 mm thick piece of HDPE results in a notable decrease in the signal, leading to reduced contrast in our measurements (<0.06). To get the velocity profile we use the [300 Hz – 750 MHz] bandwidth of the detector.

Table 2. Summary of the fringe contrasts obtained from the LWIR system for aluminum, copper and diffuse reflector target in air and through HDPE sample. The measurements were performed at 50 mm and 150 mm from the target. Crosses indicate that the measurement couldn't be perform.

Materials	Fringe contrast			
	WD = 50 mm	Expected value	WD =150 mm	Expected value
Aluminum	0.310 ± 0.038	0.42	0.388± 0.064	0.33
Copper	0.333 ± 0.044	0.40	0.375 ± 0.042	0.35
Diffuse reflector	0.254± 0.056	0.19	0.248± 0.054	0.25
Aluminum/HDPE	0.047± 0.006	x	0.020 ± 0.007	x
Copper/HDPE	0.055± 0.007	x	0.022± 0.002	x
Diffuse reflector/HDPE	x	x	x	x

We can notice that all measurements allowed us to track the surface velocity even through the HDPE sample except for the diffuse reflector through the HDPE sample due to the lack of backscattered power. In air medium, the measured contrasts are close to the expected value without taking into account the laser linewidth. Unfortunately, the expected value could not be obtained through the HDPE sample due to the minimal amount of signal collected, which was overwhelmed by environmental vibration noise.
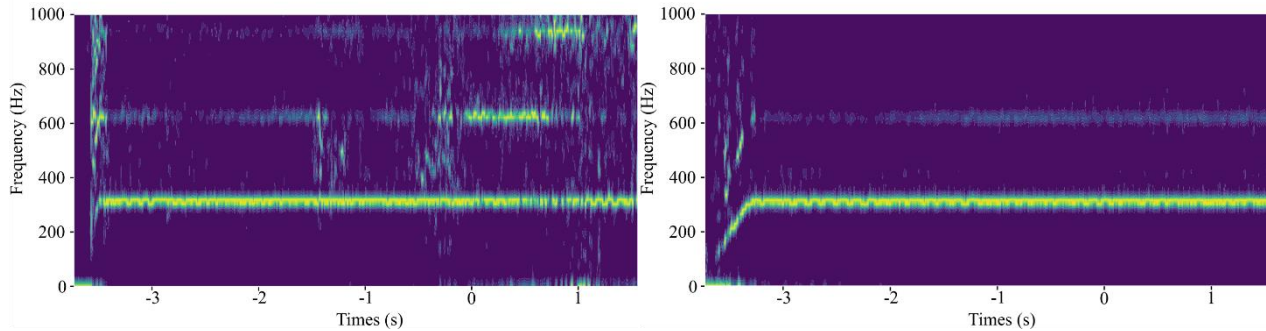

Figure 3. Spectrograms of the aluminum target at 150 mm in air (left) and through HDPE (right)

The spectrograms are clear (Figure 3), allowing us to track the velocity profile of the aluminum target. A real frequency shift of 320 Hz corresponds to a velocity of 1.5 mm/s at 9.5 µm. However, during measurements in air, multiple parasite profiles arise due to the lens reflection or local misalignment while the target is moving. Despite the challenge of low contrast in the HPDE measurements, the spectrogram remains highly clear and effectively suppresses contributions from vibration and multiple reflections.

4. CONCLUSION

In this conference paper, we present the first results of our developed long-wave infrared system with the aim to measure shock wave speed profiles into materials of study. The primary objective of this study was to assess the feasibility of conducting measurements using a non-conventional wavelength, specifically near 9.5 µm. We performed measurements in both an air medium and through a material transparent enough at 9.5 µm.

We observed high fringe contrast on diffuse reflective materials such as copper, aluminum and diffuse reflectors, even when measuring behind the HDPE sample. However, for diffuse reflectors, the collected signal proved to be insufficient for accurate velocity tracking. The limitation arises from the fact that the input laser power needs to be higher. The quantum cascade laser employed has inherent limitations in power output, the laser will be upgraded in the near future.

In conclusion, our study highlights the potential of the LWIR PDV system for measurements into various media. Nevertheless, future research should explore the system's performance at much higher speeds (> 100 m/s) and with a laser

with higher power and better linewidth. These enhancements could lead to further improvements in the system's capabilities and expand its applicability in high-speed measurement scenarios.

REFERENCES

[1] Barker, L. M. and Hollenbach, R. E., "Interferometer Technique for Measuring the Dynamic Mechanical Properties of Materials," Review of Scientific Instruments **36**(11), 1617–1620 (1965).

[2] Barker, L. M. and Hollenbach, R. E., "Laser interferometer for measuring high velocities of any reflecting surface," Journal of Applied Physics **43**(11), 4669–4675 (1972).

[3] Nissim, N., Greenberg, E., Werdiger, M., Horowitz, Y., Bakshi, L., Ferber, Y., Glam, B., Fedotov-Gefen, A., Perelmutter, L. and Eliezer, S., "Free-surface velocity measurements of opaque materials in laser-driven shock-wave experiments using photonic Doppler velocimetry," Matter and Radiation at Extremes **6**(4), 046902 (2021).

[4] Franzkowiak, J.-E., Prudhomme, G., Mercier, P., Lauriot, S., Dubreuil, E. and Berthe, L., "PDV-based estimation of high-speed ejecta particles density from shock-loaded tin plate," presented at SHOCK COMPRESSION OF CONDENSED MATTER - 2017: Proceedings of the Conference of the American Physical Society Topical Group on Shock Compression of Condensed Matter, 2018, St. Louis, MO, USA, 080004.

[5] Pei, H., Huang, W., Zhang, X. and Zheng, X., "Measuring detonation wave profiles in plastic-bonded explosives using PDV," AIP Advances **9**(1), 015306 (2019).

[6] Rougier, B., Aubert, H., Lefrançois, A., Barbarin, Y., Luc, J. and Osmont, A., "Reflection of Electromagnetic Waves From Moving Interfaces for Analyzing Shock Phenomenon in Solids," Radio Sci. **53**(7), 888–894 (2018).

[7] Bel'skii, V. M., Mikhailov, A. L., Rodionov, A. V. and Sedov, A. A., "Microwave diagnostics of shock-wave and detonation processes," Combust Explos Shock Waves **47**(6), 639–650 (2011).

[8] Patimisco, P., Sampaolo, A., Mihai, L., Giglio, M., Kriesel, J., Sporea, D., Scamarcio, G., Tittel, F. and Spagnolo, V., "Low-Loss Coupling of Quantum Cascade Lasers into Hollow-Core Waveguides with Single-Mode Output in the 3.7–7.6 µm Spectral Range," Sensors **16**(4), 533 (2016).

[9] Smith, B., "The Infrared Spectra of Polymers II: Polyethylene," Spectroscopy **36**(9), 24–29 (2021).

[10] Jordan, J. L., Dattelbaum, D. M., Schilling, B. F., Welch, C. F. and Stull, J. A., "Low pressure shock response and dynamic failure of high density polyethylene (HDPE)," AIP Conference Proceedings **1979**(1), 090006 (2018).

Design of a single optical bench for complete Performance characterization of night vision device

Morgane Lardennois[a], Romane Dorino[a], Adrien Bertaud[a], Marie d'Autume[a], Guillaume Arquetoux[a],
Arnaud Louboutin[a], Catherine Barrat*[a], Romain Guider[a]
[a]HGH Systèmes Infrarouges, 10 Rue Maryse Bastié, 91430 Igny, France

ABSTRACT

Low light night vision systems based on I² tubes have been expanding rapidly over the past few years, due to a combination of the growing advancement of this technology and the increased pressure in the current climate.

The design of a single optical bench able to fully characterize night vision devices is presented into this paper, focused more specifically on spot defects and goggle axes parallelism tests.

These criteria are indeed very important: misalignment between the two binocular images may be one source of visual fatigue and could degrade task performance of the night vision user, and spot defects can act as visual distractions and may be large enough to mask critical information pilots need to conduct normal night vision operations.

Thanks to HGH's IRCOL bench, these two tests are integrated on the same support. Spot defect measurement utilizes machine vision algorithms to determine the size and location of the defects, and the parallelism measurement identifies the angular misalignment between the two channels under test. The spot defect test has also been completely automatized compared to the only visible test previously available

All these results will be compiled and directly integrated into a computer-generated report that can be easily used for quality control or for maintenance applications.

Keywords: night vision system, spot defects, goggle axes parallelism, optical bench

1. INTRODUCTION

Night vision devices are electronic devices that allow the wearer to see better in the dark without using any external light source such as a torch or a lamp. Light amplification technology is not only the most widely used technology today, but also the most popular. Devices using this technology are commonly referred to as Night Vision Devices (NVDs) or Night Vision Goggles (NVGs). This type of device generally integrates a special tube, called an image intensifier tube or I² tube, to capture and amplify light, whether infrared or visible. This tube consisting of a photocathode, an anode in form of a phosphor screen, and other components intensifies an input low luminance image into a brighter image created on the anode (screen) [1].

These NVD systems can be used both in civilian life (animal observation or civil aviation, for example), and of course in the military field, to give significant supremacy in night-time operations. The improvements in night optics in the last decade have been immense. Drastic technological advancements have meant clearer images, improved light amplification, and unbelievable leaps forward for durability and portability. In the current context (terrorism, war in Ukraine), these systems are becoming crucial.

The purpose of this article is to present and document some of the tests that are now included in the HGH's IRCOL bench in order to fully characterize NVDs.

*Catherine.barrat@hgh-infrared.com; phone +33 1 69 35 47 70; https://hgh-infrared.com/

2. IRCOL TEST BENCH

2.1 Optical setup

The IRCOL benches are test equipment designed for the characterization and the performance validation of a wide array of electro-optical sensors, like visible cameras, NIR, SWIR, MWIR, LWIR imagers and laser rangefinders.

Figure 1: IRCOL Optical Test Bench

The core of the IRCOL system is the IRCOL collimator. The IRCOL collimator features an off-axis mirror-based projector with a wide clear aperture, a diffraction limited output beam and a high transmittance covering a wide spectral range from near UV to far IR.

In order to test and characterize any Electro-Optical system the IRCOL collimator projects to infinity optical stimuli, thanks to the specific targets or patterns located at its focus.

The bench is equipped with various sources such as blackbodies or visible to SWIR sources, depending on the Unit Under Test (UUT) type: thermal cameras, visible cameras, etc.

2.2 Infratest software

The INFRATEST software completes the IRCOL system, for the automated control of the bench and the computation of an exhaustive range of measurements: noise functions, NETD and other signal resolutions, bad pixel location and non-uniformity correction, MTF and spatial resolution data, distortion, MRTD, TOD, MRC and ranges calculation, multiple axes alignment control and laser rangefinder accuracy measurement and many other functions.

3. NVD TESTING CONFIGURATION

3.1 Light source specifications

The optical source is an integrating sphere-based source with a color temperature of 2856K +/- 50K . This value is a prerequisite for testing any NVD. The ISV range of sources of HGH's portfolio is precisely set in color temperature to the desired level using a calibrated colorimeter.

The ISV source is also equipped with a set of neutral densities enabling to adjust the irradiance of the NVD. Knowing the luminance of the sphere (adjustable between 1 and 3400 cd/m^2), the density used and the collimator geometrical parameters, the software automatically calculates the luminance setpoint of the ISV source as a function of the irradiance setpoint of the NVD. This configuration can simulate night levels down to level 5, i.e., an irradiance <0.7 mlux, essential for NVD tests.

3.2 Eye-camera

Night Vision Devices necessarily have an optical output through one or two eyepieces. Indeed, many of the tests on NVDs are nowadays still carried out thanks to the human eye of a technician, for example to count the number of spots defects. These procedures may lead to errors or subjective results depending on the operator's experience.

One of the major advances of our bench is the use of a camera collecting the image through the eyepiece to replace the human eye in order to analyze and validate all NVD characterization tests.

The main specifications of our Eye-camera are listed into the below table:

Parameter	Specification
Horizontal and Vertical Field of view	49 x 49 degrees
Input aperture	6 mm (similar to output aperture of the eyepiece)
Size of the smallest detectable spot defect for 16 mm diameter tubes	<75 µm[1]

Table 1: Specifications of the Eye-camera

The substitution of the operator by this monochrome camera allows to carry out all the tests usually done through the human eye. In addition to the removal of the subjectivity of the human eye for the tests, this feature allows to streamline and automate these tests integrated into Infratest software.

We integrated the ability to fully characterize NVDs into our existing software Infratest, focusing in more details on Spots Defects and Goggle axes parallelism tests. Thanks to the integration of these functions into the IRCOL optical bench, users are now be able to use the functions previously available in dedicated benches on their main bench, alongside the other applications already available on the IRCOL.

4. GOGGLE AXES PARALLELISM TEST

4.1 Principle

Goggles axes parallelism control is performed by measuring and comparing the positions of a pinhole image through each axis of the goggle. This function uses a high-resolution camera looking through the eyepiece. Figure 2 shows the setup used to carry out this test, consisting of our analysis camera, a translation stage and a collimator associated with a pinhole sight and a light source.

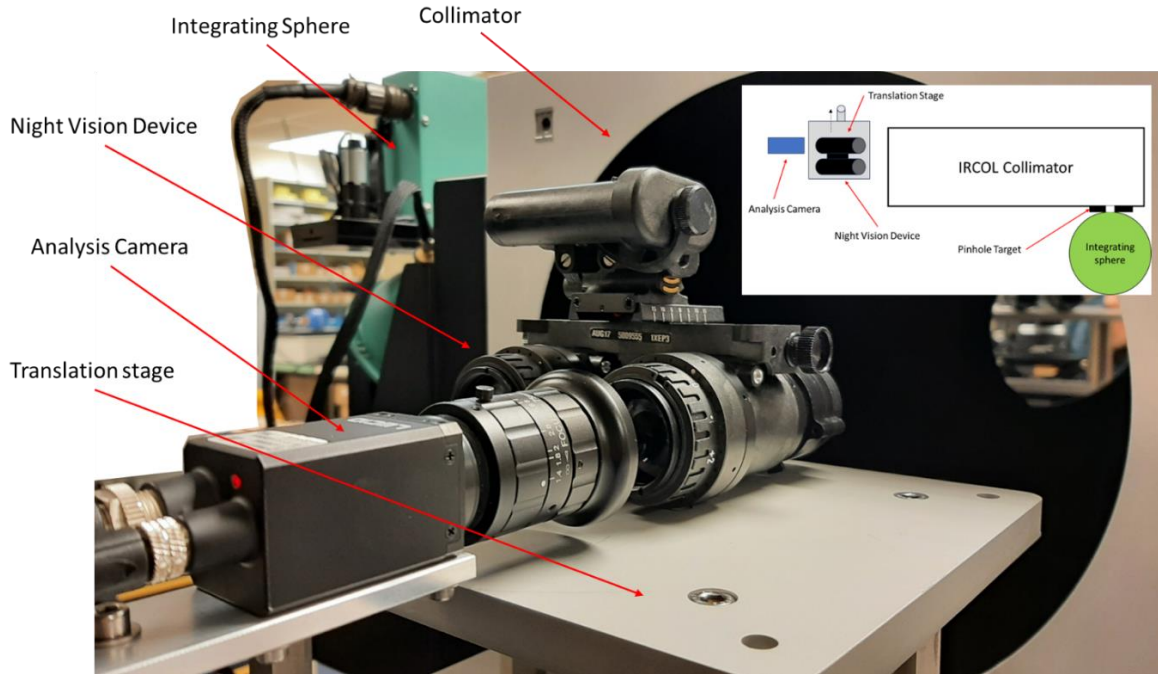[1] Spot size of category 1 according to [6]

Figure 2: Optical setup (photography and scheme) for Parallelism test.

The first test step is to align the Eye-camera with the pinhole test pattern, the NVD to be tested being removed. Once this is done, we insert and align the first channel of the NVD with the Eye-camera and pinhole. Having maximized the signal, we then move the NVD using the translation stage to align the second channel of the NVD.

The misalignment $\Delta\theta_x$ between the two channels can be expressed as follows:

$$\Delta\theta_x = \text{atan}\left(\frac{(x_2-x_0)\times p_x\times 10^{-3}}{f'}\right) - \text{atan}\left(\frac{(x_1-x_0)\times p_x\times 10^{-3}}{f'}\right) \tag{1}$$

Where $x_2, x_1\ and\ x_0$ are the barycentre positions of the pinhole images respectively on channels 2 and 1 of the goggle and on the camera alone. p_x corresponds to the horizontal pixel pitch (in µm) and f' to the focal length (in mm) of the Eye-camera.

4.2 Results

The procedure described above was carried out five times in succession, three times measuring first the right and then the left channel, and twice measuring first the left and then the right channel. Each time, the goggle was removed and then reinstalled. Figure 3 presents a screenshot of Infratest software during the measurement phase.
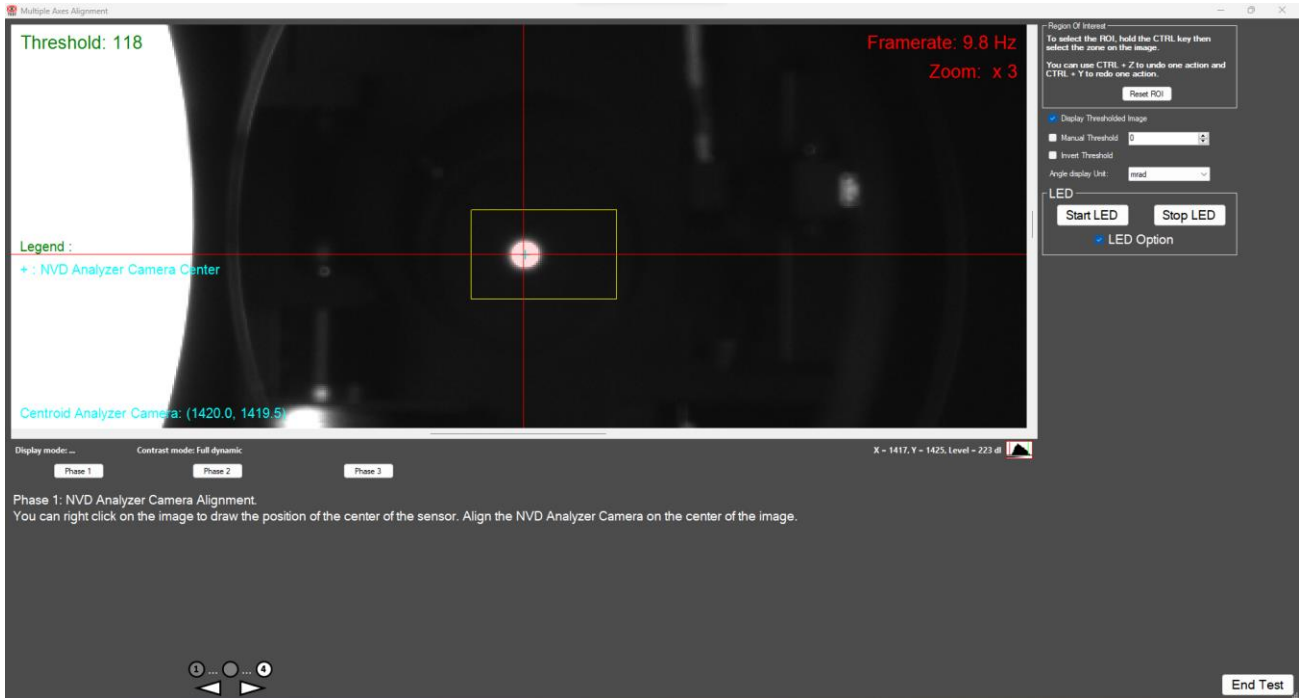
Figure 3: Infratest screenshot of Parallelism test.

The deviation found was 5.9 ± 0.50 mrad. As current market values are of the order of a few mrads [2], and as we were unable to compare the measured value with the supplier's reference data, further tests on the subject are planned, notably on other NVDs with factory-certified values.

5. SPOT DEFECTS MEASUREMENTS

Spots or Spot Defects are little black spots that can appear in an intensifier tube. They can happen for several reasons, but the most common are photocathode or phosphor burns, broken light fibres, bad channels in the microchannel plate and dust on the outside surfaces of the I^2 tube. These dark spots of various sizes are inevitable and can be located on any part of the field of view [3, 4].

5.1 Algorithm and setup definition

In order to provide the best possible quality products, standards have been defined for manufacturers of intensifier tubes to define the maximum number of spots and their size according to the area in which they are located on the image. Spots are always measured according to [6]. Figure 4 shows a graph identifying the three main zones of an 18 mm I^2 tube: Zone 1 correspond to a diameter below 5.6mm, Zone 2 of a diameter between 5.6 and 14.7mm and Zone 3 of a diameter between 14.7 and 17.5mm.
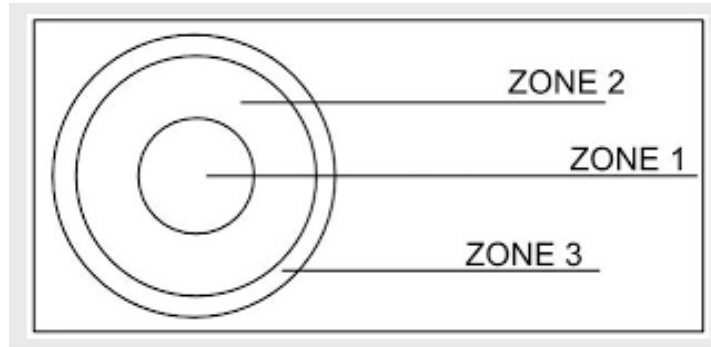
Figure 4: Zone chart

The presence of spot defects is obviously more critical in Zone 1 than in any other zone. The usual method for spot defect test is to locate them through visual inspection on a large target panel showing the above zones and covering the NVD field of view.

Figure 5: Example of spot defect panel used in visual inspection

The subjectivity of this method is obvious.

Our spot defect detection algorithm is based on the method described by Aghaziyarati et al. in [5] and uses absolute average difference weighted by cumulative directional derivatives.

From a hardware point of view, to integrate the characterization of these dark spots on our optical bench, we removed the integrating sphere source available on IRCOL bench coupled with our Eye-camera and a set of optical densities. A sleeve is also inserted to eliminate any stray light from the measurement. Figure 6 shows the setup for this measurement, integrated with the optical bench.
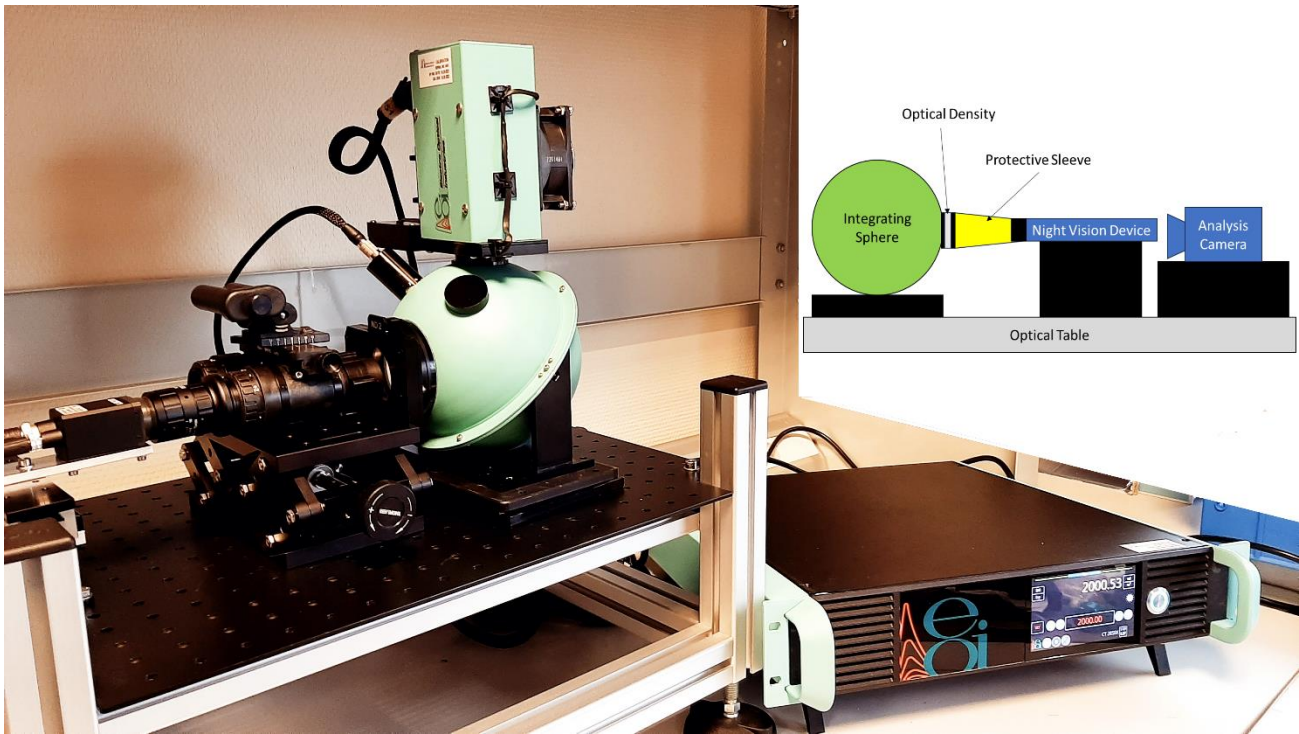
Figure 6: Optical setup and photography for Spot Defect Measurement

5.2 Tests presentation

Three tests were carried out to demonstrate the correct operation of this automated and objective method.

Test no 1: Spot detection. The objective of this test is to compare the number of spots found by our software with those identified by an experienced operator. In this test, the operator performed two counts: the first directly through the eyepiece of the NVD and the second by counting the spot defects saved by the camera image.

Test no 2: Repeatability. The second test is to demonstrate the repeatability of the bad pixel detection measurement by completely dismantling and reassembling five times the NVD on our measurement setup. We performed this test of both NVG Dual Tubes and NVG Mono Tube. As a reminder, Table 2 shows the categories defining spot sizes according to [6].
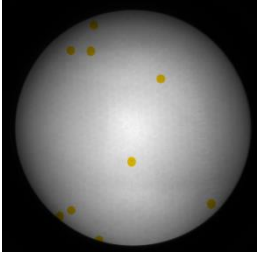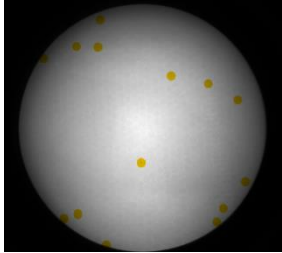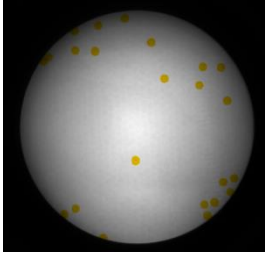
Table 2: Spot size classification

Spot Category	Spot Size (μm)
Category 1	< 75
Category 2	75-150
Category 3	151-230
Category 4	231-300
Category 5	301-380
Category 6	381-500

Test no 3 : correct classification. We compare the classification found by the software with the classification made by the operator and measured directly on the image, thanks to the knowledge of the diameter of the tube.

5.3 Test results

Table 3: Results of Test no 1

<table>
<tr><th colspan="3">Results of Test n°1</th></tr>
<tr><th colspan="2">Result found by operator</th><th>Result found by software</th></tr>
<tr>
<td>Observer looking directly out of binoculars</td>
<td>Observer looking at the raw image taken by the camera</td>
<td>Software result from the raw image taken by the camera</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 spots identified</td>
<td>14 spots identified</td>
<td>23 spots identified</td>
</tr>
</table>

At least all the spots identified by the operator are identified by the software. There is no missing spot through the software method.

Table 4: Results of Test no 2
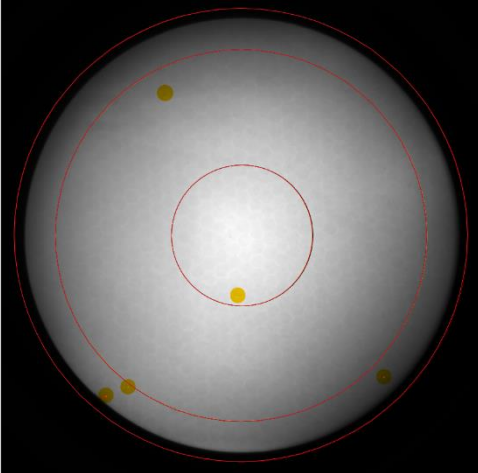
<table>
<tr><th colspan="4">Results of Test n°2</th></tr>
<tr><th colspan="4">2.1: Result found on NVG Dual Tubes</th></tr>
<tr>
<th></th><th></th>
<th>Nb of points detected on left optical path</th>
<th>Nb of points detected on right optical path</th>
</tr>
<tr><td rowspan="3">Zone 1 (Inner Circle)</td><td>Cat. 1</td><td>1.20 ± 0.45</td><td>0.13 ± 0.35</td></tr>
<tr><td>Cat. 2</td><td>No spots detected</td><td>1.00 ± 0.00</td></tr>
<tr><td>Cat 3 to 6</td><td>No spots detected</td><td>No spots detected</td></tr>
<tr><td rowspan="5">Zone 2 (Middle Circle)</td><td>Cat 1</td><td>12.40 ± 5.98</td><td>12.88 ± 8.54</td></tr>
<tr><td>Cat 2</td><td>1.00 ± 0.00</td><td>No spots detected</td></tr>
<tr><td>Cat 3</td><td>No spots detected</td><td>No spots detected</td></tr>
<tr><td>Cat 4</td><td>No spots detected</td><td>No spots detected</td></tr>
<tr><td>Cat 5 to 6</td><td>No spots detected</td><td>No spots detected</td></tr>
<tr><td rowspan="4">Zone 3 (Outer Circle)</td><td>Cat 1</td><td>12.40 ± 3.65</td><td>11.63 ± 6.93</td></tr>
<tr><td>Cat 2</td><td>No spots detected</td><td>No spots detected</td></tr>
<tr><td>Cat 3</td><td>1.00 ± 0.00</td><td>No spots detected</td></tr>
<tr><td>Cat 4 to 6</td><td>No spots detected</td><td>No spots detected</td></tr>
</table>

2.2: Result found on NVG Mono Tube			
		Nb of points detected on left optical path	Nb of points detected on right optical path
Zone 1 (Inner Circle)	Cat 1	61.60 ± 36.61	150.50 ± 16.26
	Cat 2	No spots detected	1.00 ± 0.00
	Cat 3 to 6	No spots detected	No spots detected
Zone 2 (Middle Circle)	Cat 1	239.20 ± 90.01	305.00 ± 14.14
	Cat 2	1.80 ± 0.84	2.00 ± 0.00
	Cat 3	No spots detected	No spots detected
	Cat 4	1.00 ± 0.00	1.00 ± 0.00
	Cat 5 to 6	No spots detected	No spots detected
Zone 3 (Outer Circle)	Cat 1	121.60 ± 32.12	42.00 ± 1.41
	Cat 2	2.00 ± 0.00	2.00 ± 0.00
	Cat 3	No spots detected	No spots detected
	Cat 4 to 6	No spots detected	No spots detected

Table 4 shows the results of spot detection as an average of the 5 successive measurements. It shows that the measurements are very repeatable for all spots of category ≥2 as the dispersion values are almost always 0.00.
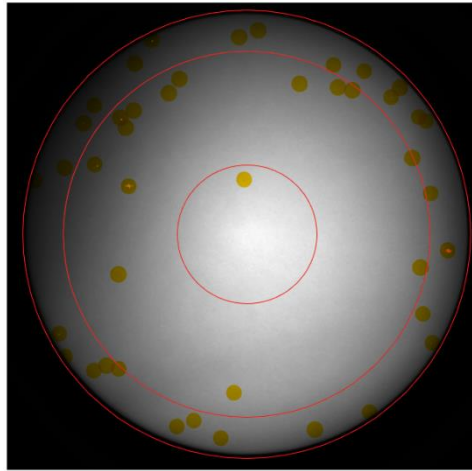
The repeatability of identification of spots of category 1 is much less repeatable but these spots are disregarded according to [6, 7].

Table 5: Results of Test no 3

Results for Test n°3
3.1: Result found on the left optical path of NVG Dual Tubes

	Values noted by the observer			Values found by software		
	Cat.1	Cat.2	Cat.3	Cat.1	Cat.2	Cat.3
Zone 1 (Inner Circle)	0	1	0	0	1	0
Zone 2 (Middle Circle)	1	0	0	1	0	0
Zone 3 (Outer Circle)	2	1	0	2	1	0

3.2: Result found on the right optical path of NVG Dual Tubes						
	Values noted by the observer			Values found by software		
	Cat.1	Cat.2	Cat.3	Cat.1	Cat.2	Cat.3
Zone 1 (Inner Circle)	1	0	0	1	0	0
Zone 2 (Middle Circle)	13	1	0	11	1	0
Zone 3 (Outer Circle)	27	0	1	13	0	1

Similarly, as category 1 spots are those of least importance for NVDs (<75µm), the fact that our operator measures more spots of this size than the software does not influence the overall performance of NVDs.

Finally, a check on spot size (linked to spot diameter) is carried out thanks to knowledge of tube diameter. Indeed, on two spots of categories 3 and 4, we measured diameters of 210µm and 250µm respectively, which correspond perfectly to the categories identified by the software for these spots.

5.4 Implementation of the test in the software

Figure 7 shows the interface for the Spot Defect test, integrated into the Infratest software. Its user-friendly interface includes full explanations of how to perform the test, output values and a complete report directly available. It is important to note that the diameter of the tube as well as its Field of View is a user selectable parameter of the software, which allows this test to be carried out for any NVD.
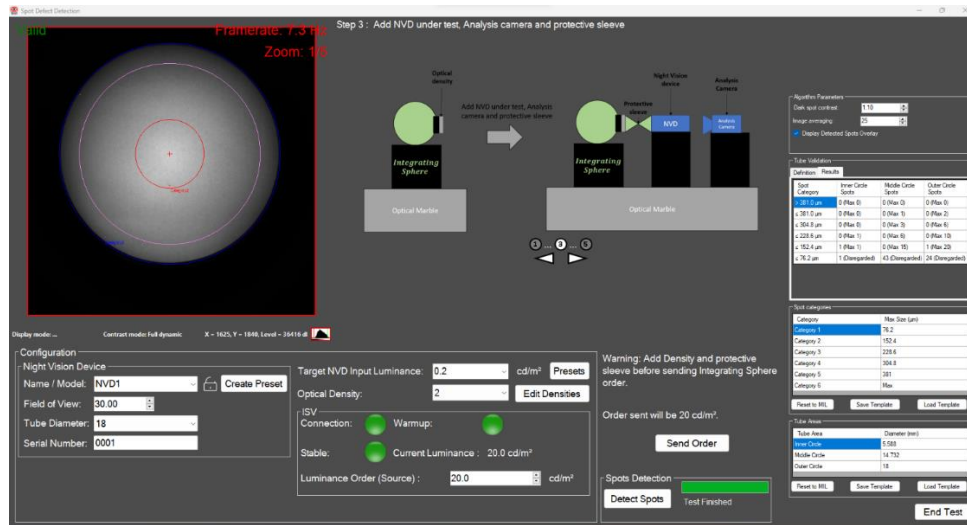
Figure 7: Spot Defects test with Infratest software

6. CONCLUSION

In conclusion, we have demonstrated in this article the integration in our Infratest software of several measurements enabling NVDs to be tested simply and efficiently in detail. In particular, we have detailed the parallelism and spot defect tests. We have also demonstrated the benefits of using an analysis camera instead of the human eye to eliminate any subjectivity in the analysis of the results obtained. Further measurements will enable us to improve and implement new tests in our software, and its combination with our range of source products (ISV) or collimators (IRCOL) will open up even further the possibility of measurements on the NVDs, which have become indispensable today.

REFERENCES

[1] Laurent, N., Bijl, P., & Deltel, G. Performance characterization of night vision equipment based on the triangle orientation discrimination methodology. Optical Engineering, 54(2), 023104-023104 (2015).

[2] Zhao, Z., Cheng, D., Yang, T., Wang, Q., Hou, Q., Gu, L., & Wang, Y. Design and evaluation of a binocular system. Applied Optics, 58(28), 7851-7857 (2019).

[3] Marasco, Peter L., Alan R. Pinkus, and H. Lee Task. "Photographic Assessment of Dark Spots in Night Vision Device Images." ASC98-1832 (1998).

[4] Scopatz, Stephen, et al. "Automated spot defect characterization in a field portable night vision goggle test set." Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXIX. Vol. 10625. SPIE (2018).

[5] Aghaziyarati, S., Moradi, S., & Talebi, H. "Small infrared target detection using absolute average difference weighted by cumulative directional derivatives," Infrared Physics & Technology, 101, 78-87 (2019).

[6] MIL−PRF−49052G Image intensifier assembly, 18-millimeter microchannel wafer, MX−9916/UV, (1999)

[7] MIL−PRF−49428 Image intensifier assembly, 18-millimeter microchannel wafer, MX−10160/AVS−6, (1995) .

Simulation of a Tomographic Single-pixel Imager Using Rosette Scanning

Armand Duvenage[a] and Warren du Plessis[a]

[a]University of Pretoria, South Africa

ABSTRACT

Spin-scan and conical-scan tomographic scanning (TOSCA) imagers have produced good-quality and cost-effective images and video in both the infrared (IR) and optical wavelengths. A novel rosette-scan implementation of TOSCA single-pixel imaging is presented below. Previous conical-scan TOSCA imagers implemented a reticle with a fixed number of thin slits. This resulted in a fixed angular resolution which implied a fixed image resolution. The feasibility of a rosette-scan implementation using similar processing techniques to conical-scan TOSCA imagers will be demonstrated. The rosette-scan implementation would only require a reticle with a single thin slit, instead of a reticle with a number of thin slits at fixed angles. The single thin-slit reticle can be rotated to be perpendicular to the line-scan angle of each rosette petal. The number of scan angles can be dynamically changed to achieve different trade-offs between resolution and frame rate by varying the rotational speeds of the prisms and the single thin slit reticle.

Keywords: Tomography, seekers, rosette scanning systems, single-pixel imaging (SPI), image sensors, image reconstruction, Image restoration.

1. INTRODUCTION

Developments in single-pixel imaging (SPI) have introduced inexpensive imaging over a range of wavelengths with exceptional resolutions and framerates.[1] High-resolution imagers in the infrared (IR) and ultra-violet (UV) spectra are extremely expensive due to the high cost of the sensor arrays. SPI attempts to address this issue by only making use of a single sensor to reduce the cost of such systems.[2,3] Single-pixel imagers have been implemented to create images and video at wavelengths from the IR to terahertz spectra.[4–6]

The majority of research in SPI focused on compressive sensing (CS) techniques to reconstruct images based on data captured by a single-pixel detector. A modulation device such as a digital micromirror device (DMD) or liquid-crystal display (LCD) is used to modulate the light that is detected by the single-pixel detector.[2] The scene is modulated in such a way as to sparsely sample the scene that is imaged. A rosette-scan pattern has also been used to sparsely sample the scene and reconstruct images using CS.[7–9] The image quality (for a fixed image size) for CS-based systems is associated with the number of samples taken. Since the development of high-specification DMD devices, high framerates were achievable in real-world applications.[1]

Image quality of tomographic scanning (TOSCA)-based systems is fundamentally not based on the number of samples taken, but on the number of line scans. The number of samples for each line scan can be varied to obtain a dynamic image size. Hovland introduced a TOSCA single-pixel imager and proved imaging capabilities at IR and optical wavelengths.[4,10] A spin-scan seeker-based mechanical assembly was used to implement the instantaneous field of view (IFOV) nutation required for the line scans required for tomographic reconstruction.[11] Two reticle types were introduced: the seeker-based knife-edge reticle and a modified reticle with a fixed number of thin slits.[12] The number of reticle knife edges or thin slits determined the number of line scans. A large number of line scans is required to obtain better image quality of the TOSCA reconstructed image.

Further author information: (Send correspondence to A. Duvenage)
A. Duvenage: E-mail: armand.duvenage@tuks.co.za
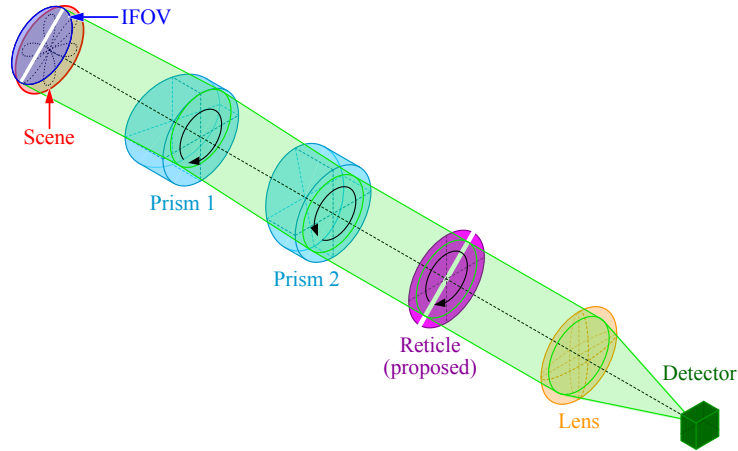W.P. du Plessis: E-mail: wduplessis@ieee.org

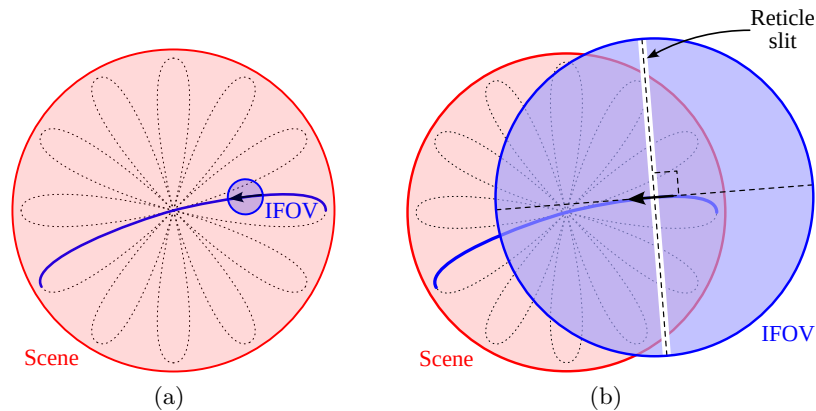Figure 1. Optics of the proposed rosette-scan TOSCA concept.

Figure 2. (a) Original rosette-scan pattern with a small IFOV and (b) rosette-scan TOSCA concept with a large IFOV and a reticle.

2. ROSETTE SCANNING OPTICS

A rosette pattern, like that in Figure 2(a), can be obtained with a dual-prism pair that is rotated in opposite directions and different rotational velocities, as illustrated in Figure 1. The relative rotational velocities of the two prisms determine the shape and number of the rosette petals.[13] This optical configuration originated from the rosette-scan seeker. A rosette-scan seeker system would ideally have a small IFOV with numerous rosette petals to have adequate coverage and be able to distinguish between real targets and countermeasure decoys.[13,14] The size of the IFOV is determined by the optical configuration of the system.

A rosette-scan TOSCA single-pixel imager can be realised by increasing the overall size of the IFOV and including a rotating reticle element. The rosette-scan TOSCA approach is illustrated in Figure 1. The reticle is placed after the prisms and before the final focusing optics and single-pixel detector. The reticle is rotated to be perpendicular to the direction that the IFOV is nutating across the scene to sweep the line across the scene, thereby realising line scanning. One line scan is completed from the outside of one petal to the outside of the next petal, as indicated by the solid rosette petal line in Figure 2(b). Figure 2(b) is drawn from the perspective of the observed scene and a light source placed at the detector's relative location. The reticle is included to limit the intensity data of the scene to the detector so that tomographic reconstruction can be implemented to reconstruct an image. The TOSCA reconstruction algorithms are described in detail by Hovland.[12]

3. RESULTS

Simulation results were generated using MATLAB 2023a along with the ISO 12233 Resolution Test Chart[15] as a test image. The ISO 12233 Resolution Test Chart vector image was converted to a bitmap image to conduct the
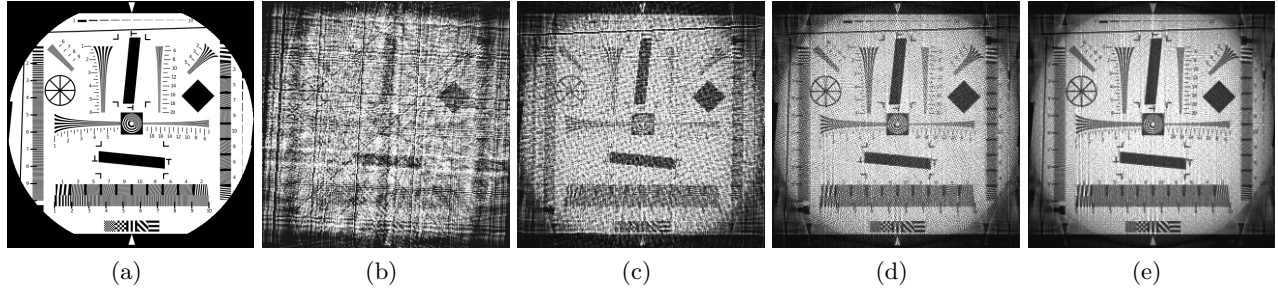
Figure 3. Reconstructed images using an ideal line-scan TOSCA approach with an increasing number of distinct scan angles. The (a) ISO 12233 Resolution Test Chart[15] was sampled at 256 samples per scan and a consistent reticle width of 1 pixel for all reconstructed images, with (b) 16, (c) 64, (d) 256, and (e) 1 024 scan angles.

simulations. The ISO 12233 Resolution Test Chart was cropped with only its central portion being retained, as shown in Figure 3(a), and various resolutions of this cropped test image were used to obtain the results. Most reconstructed images scanned a 512×512 test image, while one set of results scanned a 4 096×4 096 test image. A 512×512 test image was predominantly used because it exhibited low pixelation artefacts with adequate processing times. The 4 096×4 096 test image exhibited the lowest pixelation artefacts, however demanded substantial computational resources. The reconstructed image size was sampled such that a 256×256 size image was generated unless otherwise stated.

3.1 Ideal Line-scan TOSCA Approach

The number of distinct scan angles significantly increases the perceived resolution of the reconstructed image. A line scan was implemented to scan the scene and apply tomographic reconstruction to obtain an image. The reticle was always perpendicular to the scan angle. Furthermore, a thin-slit reticle with a width of 1 pixel was implemented.

As can be observed in Figure 3, the number of scan angles greatly increases the perceived resolution of the reconstructed images. The increased number of line scans also reduced the line artefacts inherent in tomographic reconstruction. From 256 scan angles onwards, the apparent resolution did not significantly improve. Perceptually, the best resolution of the reconstructed images with the fewest line scans was the reconstructed image with 256 line scans. Furthermore, artefacts due to interpolation errors can also be observed in all of the reconstructed images. The sampling space of the filtered back projection reconstruction process is non-Cartesian, while the displayed image is fundamentally in a Cartesian coordinate system so interpolation errors that also contribute to the artefacts arise.[11] The greater the number of distinct line-scan angles and the larger the reconstructed image, the fewer artefacts will appear.[4, 11]

3.2 Ideal Rosette-scan Pattern

The rosette-scan pattern can be altered to have overlapping petals by adjusting the ΔN variable, as described in Appendix A, where $\Delta N = N_1 - N_2$, N_1 and N_2 are the normalised rotational velocities of the first and second prisms, respectively. Three different rosette-scan patterns with 11 petals can be observed in Figures 4(a) to 4(c), with ΔN values of 2, 3 and 5, respectively. A larger ΔN value results in a larger overlap of the petals, with Figure 4(c) illustrating a petal overlap protruding to half of the adjoining petals.

To obtain a similar line scan to an ideal line scan, a very narrow petal is desirable. When $\Delta N = 2$, the IFOV would nutate across the centre of the scene in a relatively straight line, imitating an ideal line scan for the majority of the scan duration. The curvature of the petals at the centre of the scene is also reduced when a greater number of scan angles are used to scan the scene.

Images were generated using the same ΔN parameters. The reconstructed images were captured with 67 distinct scan angles, giving a rosette-scan pattern with 67 petals, with the images reconstructed using the TOSCA approach being shown in Figures 4(d) to 4(f). Although it is not clearly visible in the depicted figures, slight barrel distortion was noticeable with an increase of ΔN. The SSIM values of Figures 4(e) and 4(f), relative to
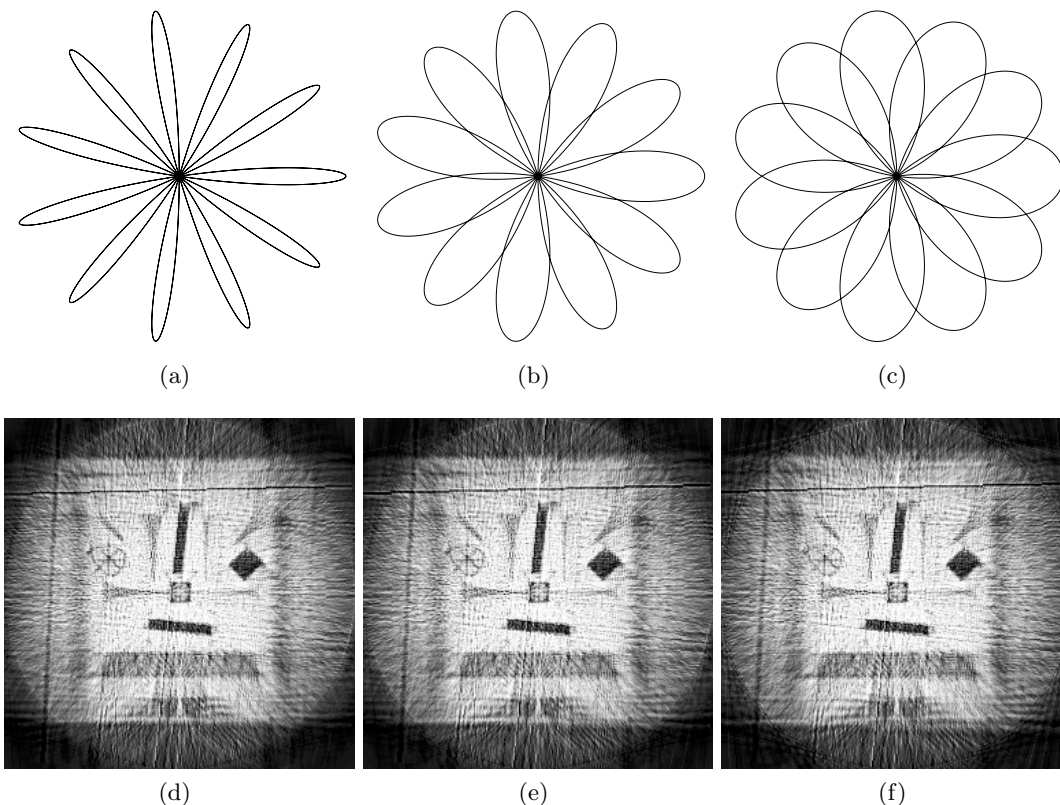
Figure 4. Rosette-pattern with different ΔN resulting in varying petal thickness. ΔN values of (a) 2, (b) 3 and (c) 5 are shown. Corresponding results of 67 distinct scan angles are also shown: ΔN values of (d) 2, (e) 3 and (f) 5. SSIM values of 0.8711 and 0.6378 were obtained for (e) and (f), respectively, relative to (d).
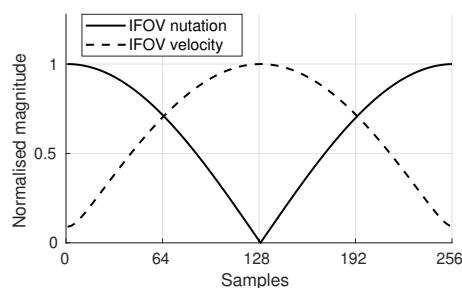

Figure 5. IFOV distance to scene centre and the IFOV relative velocity of one rosette-scan line scan from one petal maximum to the next.

Figure 4(d), were calculated to be $\text{SSIM}_{\Delta N=3} = 0.8711$ and $\text{SSIM}_{\Delta N=5} = 0.6378$. It is expected that fewer scan angles would result in an amplified barrel distortion effect, however, fewer scan angles did not produce adequate results to distinguish the finer elements of the ISO 12233 Resolution Test Chart.[15] Furthermore, it is expected that the barrel distortion will be minimised as more scan angles are incorporated into the reconstructed images, for $\Delta N > 2$, due to reduced curvature of the rosette petal near the centre of the scene.

Another noticeable distortion effect can be observed in the spoke wheel to the upper left region of the reconstructed image. This is an effect of the variation of the velocity of the movement of the rosette-scan pattern. As the IFOV approaches the outside of a rosette petal, the velocity of the IFOV is at its minimum, as shown in Figure 5. The combination of the changing IFOV nutation velocity and a constant sampling rate causes a smearing effect and distortion at the outer regions of the reconstructed images. The observed distortion is because the IFOV nutation velocity varies non-linearly, as depicted in Figure 5.
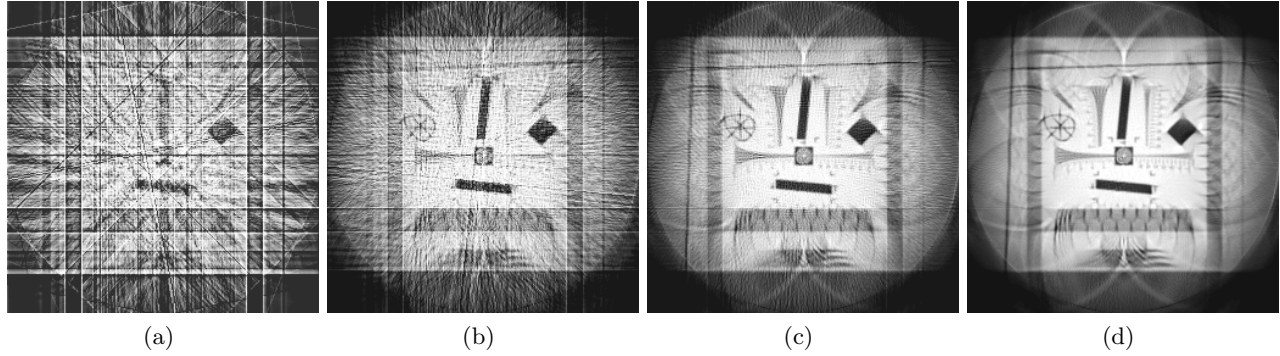
Figure 6. Reconstructed images using an ideal rosette-scan TOSCA approach with an increasing number of distinct scan angles. The centre of the ISO 12233 Resolution Test Chart[15] was sampled at 256 samples per scan and with a consistent reticle width of 1 pixel. (a) 16 scan angles, (b) 64 scan angles, (c) 256 scan angles and (d) 1 024 scan angles.
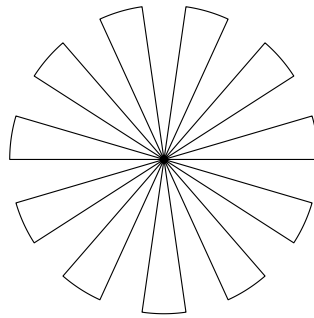

Figure 7. An illustration of a modified rosette-scan pattern for 11 line scans obtained with equal and constant rotational velocities of the two rotating prisms as depicted in Figure 1.

Figure 6 illustrates rosette-scan TOSCA reconstructed images with an increasing number of scan angles. It can be observed that the greater number of scan angles in Figures 6(c) and 6(d) results in reconstructed images that are clear of the line artefacts which are observed in Figure 6(a). Similar line artefacts were also observed in Figure 3(b). Large, distinct elements of the ISO 12233 Resolution Test Chart[15] result in line artefacts that span the entire image in Figure 6(a). However, increasing the number of scan angles from 64 to 1 024 allows the finer details to be distinguished, reducing these line artefacts. No significant resolution and clarity improvements can be observed between 256 in Figure 6(c) and 1 024 scan angles in 6(d). Smearing distortion can be observed in the outer regions of the reconstructed images when compared to the ideal line scan results in Figure 3. Furthermore, circular line artefacts can also be observed at the furthest regions of all the reconstructed images. Both the circular line artefacts and the observed smearing distortion are effects of the IFOV nutation velocity variation illustrated in Figure 5.

3.3 Modified Rosette-scan Pattern

In an attempt to minimise the inherent artefacts of the rosette-scan TOSCA approach, it is possible to rotate the prisms at equal rotational velocities to remove the petal curvature and to achieve a constant IFOV nutation velocity. One can engineer a system such that one scan angle be scanned with equal prism rotation velocities and nutate the IFOV to the next angular position by only rotating one prism, after which both prisms are utilised to nutate the IFOV with a constant velocity to the next angular position to complete a line scan. The proposed modified rosette-scan pattern is shown in Figure 7 for 11 distinct scan angles. The sampling method remains the same as depicted in Figure 2 but with the proposed modified rosette-scan pattern. It can be observed that the line scan from one angular position to the next is a straight line, similar to the ideal line scan. The IFOV nutation velocity will also vary similarly to the rosette-scan pattern but the velocity variation can be minimised by controlling the rotational velocities of the prisms.
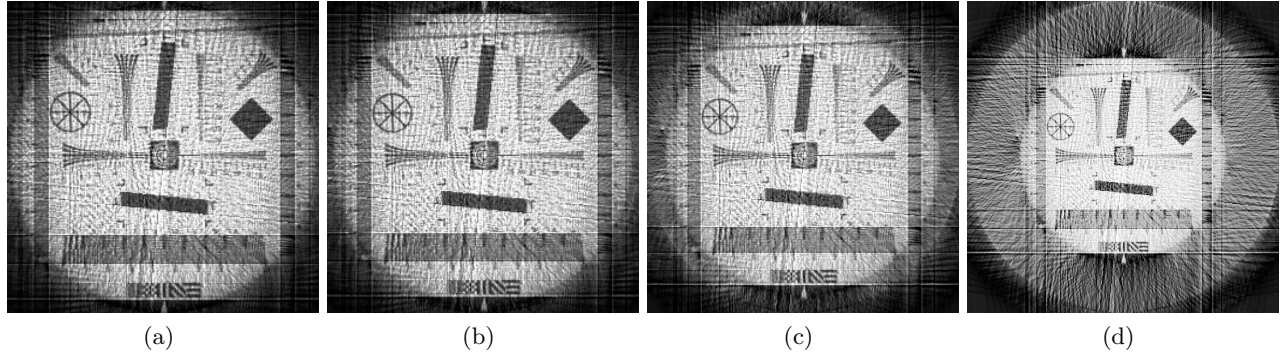
(a)   (b)   (c)   (d)

Figure 8. Reconstructed images using a modified rosette-scan TOSCA approach with an increasing number of samples from one scan angle to the next. 64 scan angles were implemented for all the depicted results. (a) 2 samples, (b) 8 samples, (c) 32 samples and (d) 128 samples.
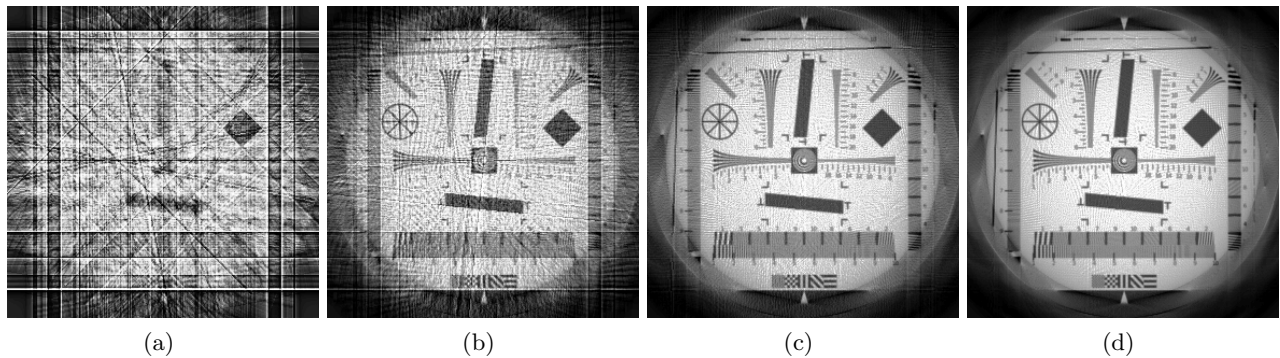

(a)   (b)   (c)   (d)

Figure 9. Reconstructed images using a modified rosette-scan TOSCA approach with an increasing number of distinct scan angles. The ISO 12233 Resolution Test Chart[15] was sampled at 256 samples per scan and with a constant reticle width of 1 pixel. (a) 16 scan angles, (b) 64 scan angles, (c) 256 scan angles and (d) 1 024 scan angles .

Additional samples will be recorded while moving from one angular position to the next at the outside of each petal of the modified rosette pattern. These samples are redundant and can be discarded in post-processing. Nevertheless, they are included in the simulation results to demonstrate the effects they have on reconstructed images. The number of samples at the outsides of the petals depends on the sampling rate and rotational velocity of the prism that rotates to cause the relevant movement (the other prism does not rotate during this radial movement). This enables the IFOV to nutate to the next angular position for the start of the next line scan.

The results for the modified rosette-scan TOSCA approach can be observed in Figure 8. The number of samples of each line scan was kept constant, but the number of samples at the outside of the petals was varied. Because the samples at the outside of the petals increased and the number of line scan samples remained the same, the total image size was enlarged from 260×260 (Figure 8(a)) to 386×386 (Figure 8(d)). What is apparent in Figure 8 is that the last sample taken from the test image is held for the number of petal maximum samples, creating a ring artefact on the outside region of the reconstructed image. Furthermore, it can be observed that the reconstructed images do not exhibit the same smearing distortion and circular artefacts observed in Figure 6. This confirms that the artefacts are inherent to the rosette-scan TOSCA approach as a result of the petal curvature and velocity variation at the centre of the scene.

Increasing the number of distinct scan angles increases the perceived clarity and resolution of the reconstructed image, as seen in Figure 9. The results obtained in Figure 9 are comparable to the results obtained from the ideal line-scan TOSCA approach in Figure 3. Minute details of the ISO 12233 Resolution Test Chart[15] are distinguishable from 64 angular scans onward (Figures 9(b) to 9(d)) show very low incidences of line artefacts. The ring artefacts in the reconstructed images in Figure 9 are result of the use of 32 samples at the outsides of the petals.
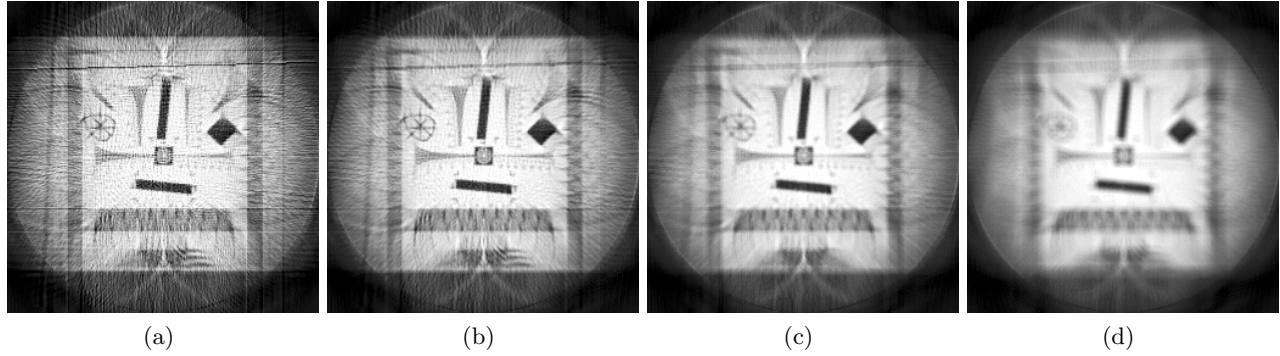
Figure 10. Simulation of ideal rosette-scan TOSCA reconstructed image of the ISO 12233 Resolution Test Chart[15] with different size thin-slit reticles. 128 scan angles were used for all reconstructed images. (a) 1 pixel, (b) 4 pixels, (c) 8 pixels and (d) 16 pixels sized thin-slit reticles.
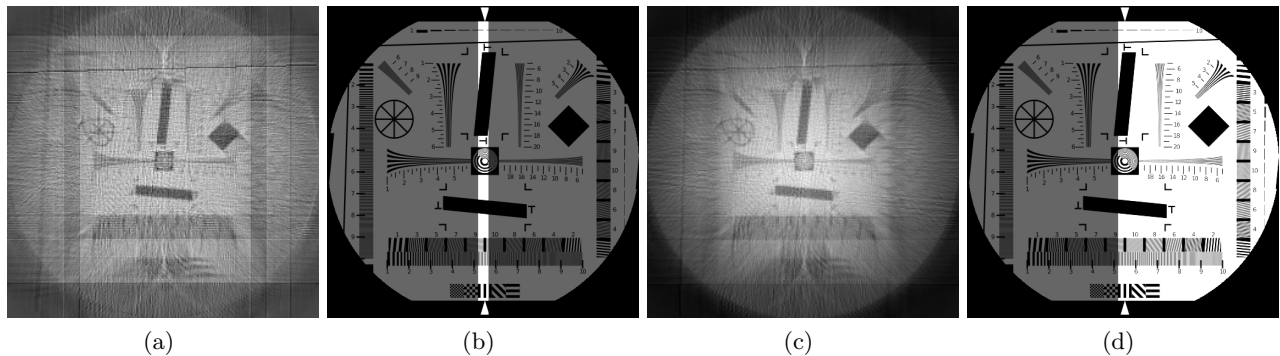

Figure 11. Reconstructed images (a) and (c) were reconstructed using rosette-scan TOSCA approach with a (b) single-slit reticle and a (d) bisected reticle, respectively. The difference reconstruction algorithm was implemented with the bisected reticle and the standard reconstruction algorithm for the thin-slit reticle. The ISO 12233 Resolution Test Chart[15] was sampled at 256 samples per scan and 128 scan angles. The reticles that are shown, (b) (d), have been altered for demonstration purposes. The reticle width of (b) was increased to be visible in the figure and the greyed areas indicate the opaque areas of the reticle.

3.4 Reticle Width Variation

All the results above utilised a reticle slit with a theoretical width of 1 pixel. Although the reconstructed images produced good results, it is difficult to practically construct such a reticle and to compensate for the lack of light caused by such a thin slit in the reticle. Figure 10 illustrates the effect of the variation of the reticle slit width on the reconstructed image. It can be observed that a large reticle slit results in a soft reconstructed image with image details that are less pronounced and textured compared to the reconstructed image for a 1-pixel wide reticle slit.

Hovland introduced a similar TOSCA reconstruction algorithm that makes use of the derivative signal rather than the intensity signal itself.[12] The application of this approach introduced the use of reticles with transmissive regions much larger than a thin slit in the reticle. Figure 11 attempts to compare the results of a reticle slit and a bisected reticle, which is formed by dividing the circular reticle into one transparent semicircle and one opaque semicircle. The derivative TOSCA reconstruction process is used for the bisected reticle. The reconstructed images in Figure 11 have 256×256 pixels and were obtained from a 4 096×4 096 sized ISO 12233 Resolution Test Chart.[15] Effectively, the required line of averaged scene information is formed by the difference between two images where the reticle has moved to block a different portion of the image. This resulted in a reconstructed image that is perceived as sharp but discretised in Figure 11(a), and a smoother and averaged look in Figure 11(c).

Figure 12 depicts the reconstructed images of rosette-scan TOSCA, utilising a bisected reticle. Although the images seem overexposed, they are comparable to the results for the rosette-scan TOSCA reconstructed images
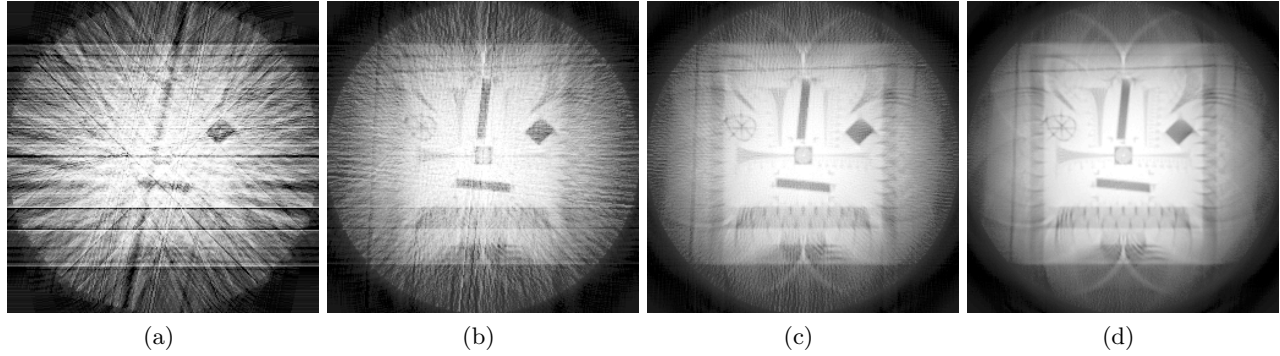
Figure 12. Reconstructed images using an ideal rosette-scan TOSCA approach with an increasing number of distinct scan angles. The difference reconstruction algorithm was implemented with a bisected reticle. The ISO 12233 Resolution Test Chart[15] was sampled at 256 samples per scan. (a) 16 scan angles, (b) 64 scan angles, (c) 256 scan angles and (d) 1 024 scan angles.
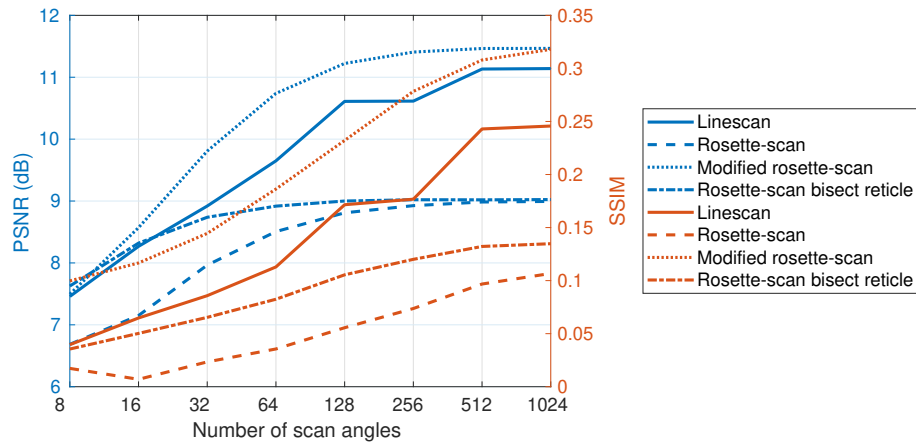
Figure 13. PSNR and SSIM values of the reconstructed images depicted in Figures 3, 6, 9 and 12, indicating an improvement in image quality as the number of scan angles is increased.

using a 1-pixel wide thin-slit reticle. The same artefacts that were highlighted in Section 3.2 for Figure 6 are evident in Figure 12.

3.5 Image Quality Comparison

The SSIM and peak signal-to-noise ratio (PSNR) image quality metrics of the reconstructed images of the four different scanning and reconstruction methodologies are compared in Figure 13. It can be observed that most of the SSIM and PSNR results plateau at 256 scan angles, the same number as the reconstructed images' one-dimensional sizes. Increasing the number of distinct scan angles beyond 256 does not yield a significant improvement in the image quality. The images were compared to the 256×256-pixel apertured ISO 12233 Resolution Test Chart[15] in Figure 3(a). The test image showed significant pixelation due to the conversion of the ISO 12233 Resolution Test Chart to a 256×256-sized bitmap image. Although the data in Figure 13 are not a true reflection of the image quality of the reconstructed image because of the low-resolution reference image, they illustrate a trend of image quality improvement with an increasing number of scan angles for all scan and reconstructed methodologies.

It can be observed that the modified rosette-scan pattern achieved the best results. This is due to the IFOV nutating in a straight line with a constant nutation velocity. The rosette-scan pattern with different reticles performed similarly for a high number of scan angles. The rosette-scan bisected reticle performs slightly better than the thin-slit reticle at a low number of scan angles.

4. CONCLUSION

Rosette-scan TOSCA single-pixel imaging has been successfully implemented in simulation. Furthermore, a modified rosette pattern and the use of a bisect reticle were demonstrated. The bisect reticle implementation made use of the derivative of the intensity function to reconstruct images. The circular line artefacts and distortion observed in all the rosette-scan results were similar for the bisect and thin-slit reticles. The bisect reticle results produced a better image quality for a low number of scan angles. The modified rosette-scan pattern produced better results than the standard rosette-scan pattern due to a constant IFOV nutation and velocity. The rosette-scan pattern can be dynamically updated by changing the angular velocities of the two prisms to obtain more rosette petals for an increased number of line scans. Furthermore, the scanning pattern can be engineered to mitigate and minimise artefacts observed in the reconstructed image.

ACKNOWLEDGMENTS

A. ROSETTE PATTERN EQUATIONS

The rosette pattern, illustrated in Figure 2(a), can be defined in Cartesian space[13] with respect to t as

$$x(t) = \frac{\delta}{2}(\cos{(2\pi f_1 t)} + \cos{(2\pi f_2 t)}) \tag{1}$$

$$y(t) = \frac{\delta}{2}(\sin{(2\pi f_1 t)} - \sin{(2\pi f_2 t)}). \tag{2}$$

Where f_1 and f_2 are the rotational frequencies of the rotating prisms. The deviation angle of the prisms, δ, is also included in the equations. The greatest common denominator, f, can be calculated as the ratio of the prisms' rotational frequencies, $f = f_2/f_1$. The relative rotational rates for both prisms can be calculated to be $N_1 = f_1/f$ and $N_2 = f_2/f$. When both of the relative rotation rates are positive integers, the number of petals can be determined as $N = N_1 + N_2$. The corresponding petal width is represented by $\Delta N = N_1 - N_2$, where $\Delta N = 2$ has no petal overlap and $\Delta N > 2$ produces a petal overlap. Finally, the size of the IFOV can be calculated to be

$$\omega = \frac{2\pi\delta}{N} \cdot \cos{(\frac{\pi}{\Delta N})}. \tag{3}$$

REFERENCES

[1] Kilcullen, P., Ozaki, T., and Liang, J., "Compressed ultrahigh-speed single-pixel imaging by swept aggregate patterns," *Nat Commun* **13**, 7879 (Dec. 2022).

[2] Gibson, G. M., Johnson, S. D., and Padgett, M. J., "Single-pixel imaging 12 years on: A review," *Opt. Express* **28**, 28190–28208 (Sep. 2020).

[3] Hovland, H., *Tomographic scanning imagers*, PhD thesis, University of Oslo, Oslo, Norway (2016).

[4] Hovland, H., "Construction and demonstration of a multispectral tomographic scanning imager (TOSCA)," *Opt. Express* **21**, 4688–4702 (Feb 2013).

[5] Chan, W. L., Charan, K., Takhar, D., Kelly, K. F., Baraniuk, R. G., and Mittleman, D. M., "A single-pixel terahertz imaging system based on compressed sensing," *Appl. Phys. Lett.* **93**, 121105 (Sept. 2008).

[6] Bian, L., Suo, J., Situ, G., Li, Z., Fan, J., Chen, F., and Dai, Q., "Multispectral imaging using a single bucket detector," *Sci Rep* **6**, 24752 (Apr. 2016).

[7] Uzeler, H., Cakir, S., and Aytaç, T., "Image reconstruction for single detector rosette scanning systems based on compressive sensing theory," *Opt. Eng.* **55**(2), 23108 (2016).

[8] Jiang, Y., Tong, Q., Wang, H., Yang, Z., and Ji, Q., "Image Recovery of an Infrared Sub-Imaging System Based on Compressed Sensing," *Symmetry* **9**, 260 (Nov. 2017).

[9] Stoltz, G. and Stolz, M., "Performance estimation of a real-time rosette imager," *Electro-Opt. Infrared Syst.: Technol. Appl.* **11537**(11) (2020).

[10] Hovland, H., "Experimental tomographic scanning (TOSCA) imagers," *Infrared Technol. Appl.* , 150–155 (Jun. 2014).

[11] Hsieh, J., [*Computed tomography: principles, design, artifacts, and recent advances*], Wiley Interscience ; SPIE Press, Hoboken, N.J.: Bellingham, Wash, 2nd ed ed. (2009).

[12] Hovland, H., "Tomographic scanning imager," *Opt. Express* **17**, 11371–11387 (Jul 2009).

[13] Jahng, S.-G., Hong, H.-K., Choi, J.-S., and Han, S.-H., "Reticles: Nutating Systems," in [*Encyclopedia of Optical and Photonic Engineering, Second Edition*], Hoffman, C. and Driggers, R., eds., 1–12, CRC Press (Sept. 2015).

[14] Jahng, S., Hong, H., Han, S., and Choi, J., "Design and analysis of improved instantaneous field of view of rosette scanning infrared seeker," *Electron. Lett.* **33**(23), 1964 (1997).

[15] Westin, S. H., "ISO 12233 test chart," (21 Apr. 2010).

Thermal imagers with boost:
Range prediction with the Johnson Criteria

Uwe Adomeit[*]

Fraunhofer IOSB, Gutleuthausstraße 1, 76275 Ettlingen, Germany

ABSTRACT

Range prediction for thermal imagers applying advanced signal processing is still in its infancy. Boost filters are such an advanced signal processing and here it was assessed if the achieved range when using them is in correspondence with predictions based on the Johnson criteria. Equipment in test was an under-sampled MWIR imager operating with and without five different boost filters, four different Laplace- and one Wiener-filter. Range of this imager using the different boost filter was estimated by perception experiments for identification of numbers. These ranges were compared with limiting frequencies derived from Minimum Temperature Difference Perceived (MTDP) measurements including the boost filters. The comparison showed identification range and limiting frequency derived from the MTDP in good correspondence. Thus, the Johnson Criteria should be able to correctly predict range for thermal imagers including boost-filtering. Further work includes extending the comparison to low contrast and to real targets.

Keywords: MRTD, MTDP, Sensor Performance, Boost Filter, Identification Range, Imager Performance Assessment, Perception Experiment

1. INTRODUCTION

Thermal imagers increasingly employ advanced signal processing, e.g. image fusion, super-resolution, (local) contrast enhancement and boost filtering[1]. Assessment and range prediction have not kept pace with this development. Short et al[2] assessed the influence of boost on range prediction by using either the Johnson criteria or the Targeting Task Performance (TTP). They concluded[2] that "… the TTP metric heavily overestimated the effect of boost, predicting a benefit that was far in excess of what was observed in perception experiments …" whereas "(m)odelling performed using the Johnson criteria agreed with our results slightly better but still significantly undershot the performance increase observed on low-contrast imagery.". Altogether, this indicates that both approaches have problems assessing thermal imagers with boost. As Fraunhofer IOSBs TRM4 range model[3] bases on the Johnson criteria and includes digital filtering, this had to be addressed.

Assessment of the topic used an approach similar to Short et al[2]. Range was measured on IR imagery for identification of numbers. This was compared to the range derived from Minimum Temperature Difference Perceived (MTDP) measurements. Both was conducted for an MWIR undersampled thermal imager with and without applying five different boost filters. Chapter 2 presents the camera and the boost filters. A description of the MTPD measurements and results achieved is found in Chapter 3. The perception experiment and the derived ranges are summarized in Chapter 4. Chapter 5 compares the results of human perception and laboratory prediction and Chapter 6 discussed the results and gives an outlook on further work.

2. THERMAL IMAGER AND BOOST FILTERING

Basis for the range comparison measurements was an AIM Infrarot Module GmbH AIM384m thermal imager. This imager uses an cooled HgCdTe-Focal-Plane-Array with 384 x 288 elements of 24 µm pitch and 66 % fill factor. Spectral range is approximately 1.5 µm to 5.3 µm but as shown in Figure 1, a warm filter mounted in-between optics and detector limited it here to approximately 3.4 µm to 5.3 µm. For the measurements the camera was equipped with 30 mm F/2 optics, resulting in 17.6° x 13.2° field of view. Integration time was 2 ms and frame rate 25 Hz. Image data is available digitally via a RS-422 connection with 14-bit resolution. This data was grabbed using a National Instruments PCI-1422 frame grabber and own designed LabView software. The software enables storing of image sequences, conversion of the

---

[*] uwe.adomeit@iosb.fraunhofer.de

14-bit data to 8-bit for display using a linear conversion function and the boost filtering by the different algorithms with the given frame rate.
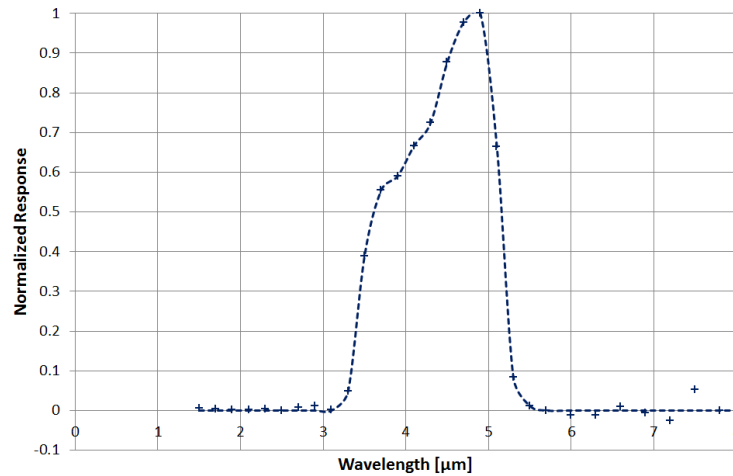

Figure 1. Normalized spectral response of the camera used for the study. Symbols are measured data and the dotted line gives the trend.

Boost filtering here is digital filtering to alter the spatial frequency content of the acquired image with the aim to increase its sharpness. Four different Laplace- and one Wiener-filter were used. The Laplace-Filters increase the local contrast at boundaries, making edges easier to see with the effect of the image look sharper[4]. A selection of the ones available in LabView was used, with Table 1 giving the kernels. For each filter the divider is the sum of the components. The Wiener Filter removes blur from the image by inverse frequency filtering[4]. The one used here was derived from LabView code found in the internet[5]. Testing showed best performance for the camera used here for a sigma of 0.5. As this filter is not truly adapted to the camera, it may not be taken as representative for the performance achievable with Wiener filtering at all.

Laplace 1

0	-1	0
-1	5	-1
0	-1	0

Laplace 2

0	-1	0
-1	6	-1
0	-1	0

Laplace 3

-1	-1	-1
-1	9	-1
-1	-1	-1

Laplace 4

-1	-1	-1
-1	10	-1
-1	-1	-1

Table 1. Filter kernels of the studies Laplace filters.

Noise Equivalent Temperature Difference (NETD) and In-homogeneity Equivalent Temperature Difference (IETD) of the camera was experimentally assessed with and without these filters for a scene temperature of 22 °C. The resulting temporal and spatial noise data are summarized in Table 2. As shown, both are distinct higher when using filtering. The increase is within a magnitude of approximately 3 to 9.

	Camera (None)	Laplace 1	Laplace 2	Laplace 3	Laplace 4	Wiener
NETD	37.5 mK	188.1 mK	111.6 mK	340.9 mK	188.2 mK	105.2 mK
IETD	7.5 mK	37.0 mK	21.8 mK	67.4 mK	37.2 mK	19.5 mK

Table 2. NETD and IETD measured for 22 °C scene temperature.

Using the scanning slit method the Modulation Transfer Function (MTF) was measured for the six configurations. Figure 2 compares the result. The MTF shows the camera to be undersampled and the changing frequency behavior due to the filters is obvious. For the Laplace filter the maximum increase is located at approximately 0.8 times Nyquist-frequency. The Wiener filter results in a more edge like behavior with the maximum increase close to Nyquist-frequency.
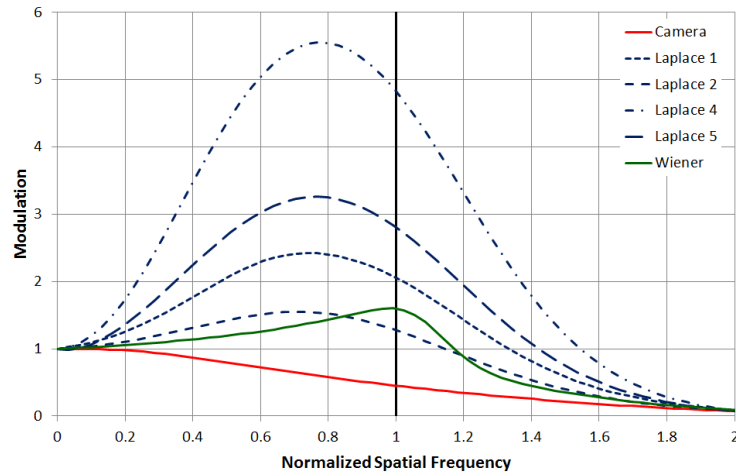
Figure 2. MTF as measured for the six configurations. Spatial frequency is given normalized to the cameras Nyquist-frequency.

3. MTDP-MEASUREMENTS

3.1 Measurement Method

Basing on the Johnson-criteria[6], the Minimum Resolvable Temperature Difference[7,8] (MRTD) allows assessing (nominal) range performance of thermal imagers. The concept found widespread application and NATO STANAG 4347[9], 4349[10] and 4350[11] standardize it. Already STANAG 4349 claimed, "The measurement procedure described applies only to thermal imagers where the MRTD concept is defined, i.e. for properly sampled systems."[10]. With the introduction of focal plane array detectors this became a problem, as they typically are not properly sampled but "undersampled". One of the concepts to overcome this problem is the Minimum Temperature Difference Perceived (MTDP)[12]. The MTDP is a development of the MRTD. From measurement perspective, the only difference between MRTD and MTDP is the allowance that the 4-bar-target used no longer needs to be correctly resolved. MTDP allows measuring when degradation of the test pattern gives three or two bars only, but selection of the maximum number is necessary. This is achieved by shifting the position of the bar target relative to the focal plane array, the so-called phase. The description of the MTDP as applied at Fraunhofer IOSB is as follows:

Validity

The MTDP applies for well and undersampled thermal imagers. The applicable spectral range is 3 μm to 14 μm or parts of this range.

Definition

The MTDP is the minimum temperature difference, which allows an observer to resolve a 4-bar-test pattern in accordance with a given criterion. It depends on the spatial frequency of the test pattern, the environmental temperature, the orientation (horizontal or vertical) and the position of the test pattern relative to the detector (phase).

Measurement Conditions

The 4-bar-test pattern is positioned in front of a blackbody, the temperature of which can be varied, giving positive and negative temperature differences ΔT between the bars and the background. A 2.5 m focal length, 0.3 m diameter off axis collimator projects the bar target on the imager under test. In doing so optical axis of imager under test and projection correspond. The position of the bar pattern relative to the detector is adjustable in small steps relative to the imager's instantaneous field of view. Environmental temperature, which also corresponds to the bar target background temperature is within 22±2 °C unless otherwise specified. Room illumination is set to a level that does not distract the observer from its task. Typically, a low level is used to increase the observer's contrast sensitivity. Alterations according to the observer's demands are allowed. Imager gain is set manually using an as high as possible one for the measurement. It is kept so during the measurements. Imager brightness is adjusted before and during the measurements by the observer to optimize results. Measurements are performed at the display that belongs to the imager. If no display

is belonging to the imager in test, an IOSB standard monitor is used instead. Contrast and brightness of the monitor are adjusted before the measurements to give optimum performance. A suited video generator is used for this purpose. No alterations are allowed during the measurements. The observer is allowed to alter the distance of his eye from the monitor to give optimized results wherever applicable.

Measurement Process

A test pattern is presented to the observer at a temperature difference that allows a good separation from the background. The observer slightly alters the position of the test pattern relative to the detector (phase) and selects the position where the maximum number of bars is visible. This can be four, three or two bars. The number of bars is recorded and the phase is kept for the measurement of the test pattern. Starting from an invisible test pattern, the temperature of the black body is increased until the test pattern can just be seen in positive contrast. The corresponding temperature difference ΔT_p is recorded. The temperature of the blackbody is then reduced causing the test pattern to disappear and then to reappear with negative contrast. When the observer can just resolve the bars, the corresponding temperature difference ΔT_n is recorded again. The two measurements are taken with a time delay as small as possible to minimize drift influence. This measurement is typically repeated at least once and at most four times. The whole procedure is repeated with bar targets of increasing spatial frequencies until the observer is not able to resolve any bars in the first step of the measurement process. Spatial frequency and temperature difference used for the decision are recorded. The criterion for resolving the bar target is to see the bars and not just some modulation on the display, although it is not necessary that each of the bars be visible at the same time. The MTDP is measured with the bar target orientated horizontal and vertical relative to the detector of the imager. Measurements typically are taken at a minimum of six spatial frequencies distributed approximately uniformly over the useful range of the imager.

Measurement Exploitation

From the recorded positive and negative temperature differences the minimum resolvable temperature difference is calculated according to

$$\Delta T = \frac{\Delta T_p - \Delta T_n}{2} \tag{1}$$

The MTDP at a given spatial frequency is then the average of these calculated temperature differences.

Results

The average temperature difference versus spatial frequency is the MTDP-curve.

3.2 Measurement Results

One trained MTDP observer performed the measurements for horizontal direction only. Using only one spatial direction was possible because former measurements showed MTDP for this camera to be independent of spatial direction. Additionally, all boost filters are spatially symmetrically. For the measurements the image data was presented on a NEC MultiSync PA241W display set to 1920 x 1200 resolution and 60 Hz framerate. It was adjusted for optimum brightness and contrast before the measurements. Image presentation used a zoom factor of 2 with pixel replication. A linear 14-bit to 8-bit conversion with 256 digital levels (DL) gain was used for all measurements. Offset was adjusted according to the observers needs. Four repetitions were performed and the values given as results in Figure 3 are average and standard deviation of these four independent measurements.

For analysis a TRM4.v3[3] modelling was adapted to the measurements. It based on the technical data of camera and display and the presented measured data. For the camera modelling eye signal-to-noise ratio and eye integration time were used for the necessary fine-tuning. These values were assumed to be the right ones for the other five configurations also. Surprisingly, when doing so, it was found that using Table 2 noise figures underestimated the performance, i.e. the MTDP-observer resolved the bar target at much lower temperature differences than expected. To overcome this and adapt the modelling to the measurement, the NETD was set to TRM4.v3s lowest possible value of 0.1 mK and the IETD was adjusted. The resulting values are summarized in Table 3. Because temporal noise is practically neglected, this value can be seen as overall noise. For comparison, the equivalent value for the camera without filter is also given in Table 3. The reason for this difference in the perceived compared to measured noise is currently unknown.

	Camera	Laplace 1	Laplace 2	Laplace 3	Laplace 4	Wiener
Noise	22.9 mK	25.0 mK	25.0 mK	30.0 mK	28.0 mK	37.5 mK
Cut-Frequency	None	1.18 mrad^{-1}	None	1.10 mrad^{-1}	1.10 mrad^{-1}	None

Table 3. Overall noise and cut-off frequency of the additional rectangular MTF used for TRM4.v3 modelling of the MTDP. See the text for details.
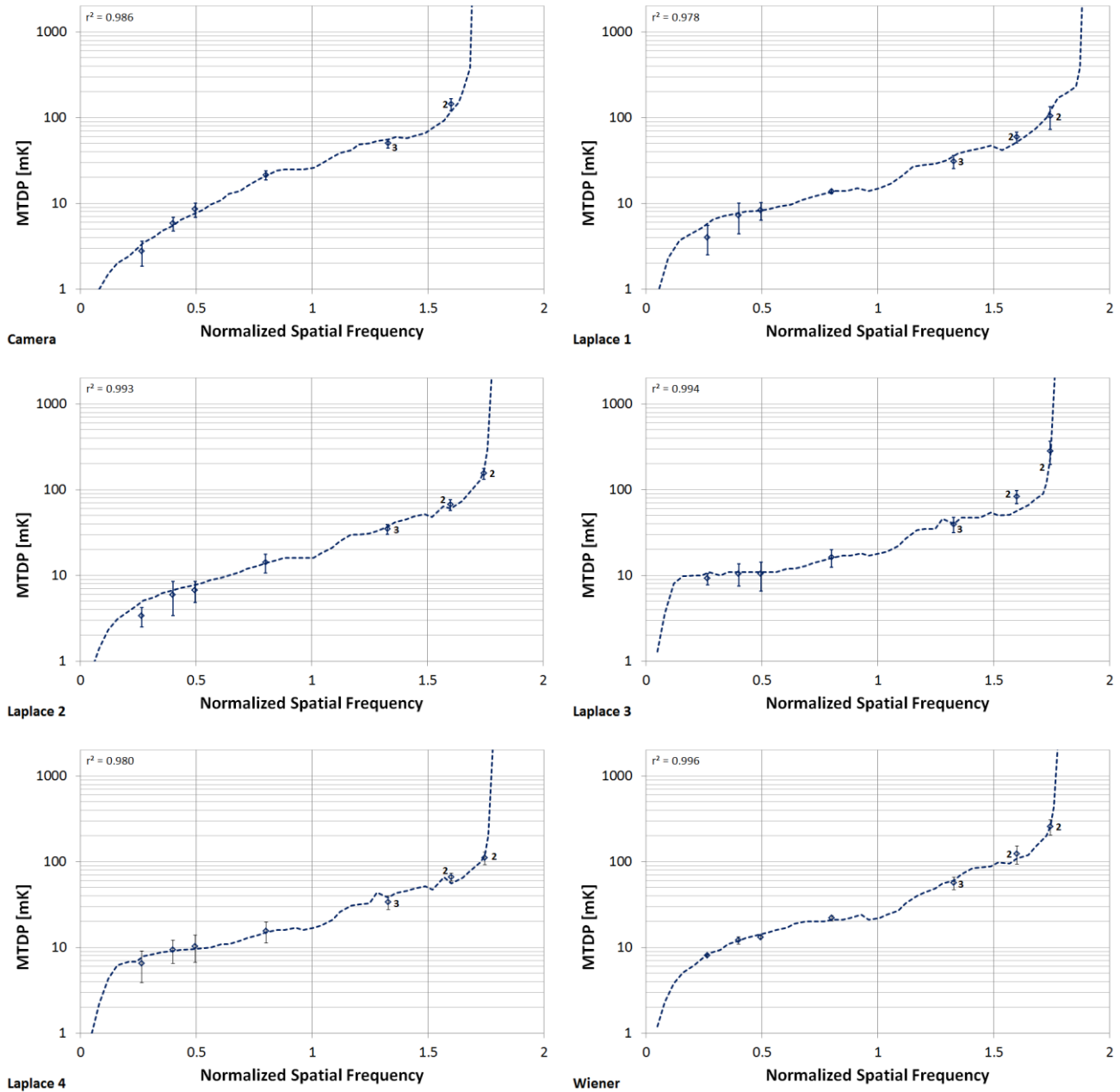
Figure 3. MTDP as measured for the six configurations. Spatial frequency is given normalized to the Nyquist-frequency. Symbols are the measured average values and the corresponding vertical bars are the standard deviation of the four measurements. Numbers close to symbols give the resolved bars in the undersampled region. The dotted line is the adapted TRM4.v3 modelling. r² gives the correlation coefficient between this modelling and the measurement.

Another problem was the MTDP cut-off frequency. For Laplace 1, 4 and 5 the modelling dramatically overestimated the observer's possibility to resolve the spatial frequencies for unknown reasons. Solution here was to include an additional rectangular MTF in the post filter. Selection of the cut-off frequency of this MTF was in-between the frequencies of the last resolved and the unresolved bar target. The used value is also given in Table 3.

As shown by the correlation coefficient given in the graphs, the correlation between measurement and adapted modelling is high, i.e. there is a good correspondence between the two.

4. NUMBER IDENTIFICATION RANGE PERFORMANCE

4.1 Remark

Range data was collected here for identification of numbers. Compared to military targets this has some advantages. As there are only the numbers 0 to 9, the target set is small and complete. It can be expected that numbers are known to all people and so learning effects are largely excluded. This enables larger experiments with a limited number of observers. Target size and thus range can be adapted to enable data acquisition under controlled conditions. Thus, it is easy to repeat the experiment e.g. with other cameras. Environmental effects are largely excluded and deriving size and contrast is simple compared to real targets. Altogether they are something in-between a target for laboratory assessment (e.g. bar target, triangle or Landolt ring) and a real target (e.g. tank or weapon). They are untypical for thermal imager range assessment and closest to the topic is probably the work of Miller and Wiltse[13].

4.2 Data Acquisition and Preparation

Target for data acquisition was a sand blasted aluminum plate of 2 mm thickness containing cut-out numbers 0 to 9 of 10 mm height. It was mounted in front of a CI-Systems SR 80 black-body. The black-body was operated switched off, so the contrast was created only by the different material properties. The camera was mounted on a rail with its optical axis perpendicular to the targets center. Moving the camera on this rail enabled recording the target from set distances. Here these distances were 1.15 m to 6.15 m in 0.5 m steps. At each distance image sequences of 200 images were recorded. At closest range the black-body set to temperatures in-between 15 °C and 40 °C with steps of 5 K was additionally recorded without the target for in-homogeneity correction and calibration purposes.

In preparation for the perception experiment, the following processing was applied:

1. 2-point in-homogeneity correction using the data collected at 15 °C and 40 °C
2. Boost filtering using above filters
3. 14-bit data to 8-bit conversion using a linear scaling of 256 DL gain and a defined offset
4. Separating the numbers using suited quadratic regions of interests

At the end six data sets with of 110 sequences resulted. Figure 4 presents examples for the processed imagery up to step 3.

For comparison with the MTDP measurement the contrast between numbers and background is needed. It was estimated for each distance by averaging the signal in regions of interest on an uncovered black-body and a background area. For calibration of these signals the recorded reference data was used. The resulting temperature difference was 230±20 mK.

4.3 Perception Experiment

The perception experiment was conducted in a darkened room on a NEC MultiSync PA241W display set to 1920 x 1200 resolution and 60 Hz framerate. Distance to the monitor could be selected by the observer as appropriate. The monitor was adjusted for optimum brightness and contrast using a gray-scale of 256 DL.

The form of the experiment was 10 answer forced choice (10AFC) with unlimited observation time. The image sequences were presented in the middle of the display in infinite loop. A zoom factor of 2 with pixel replication was used for all imagery. The software stored the presented image name and the start time of the presentation (stimulus), the observer selection and end time of the presentation (response) and automatically calculated the probability of identification for each range.
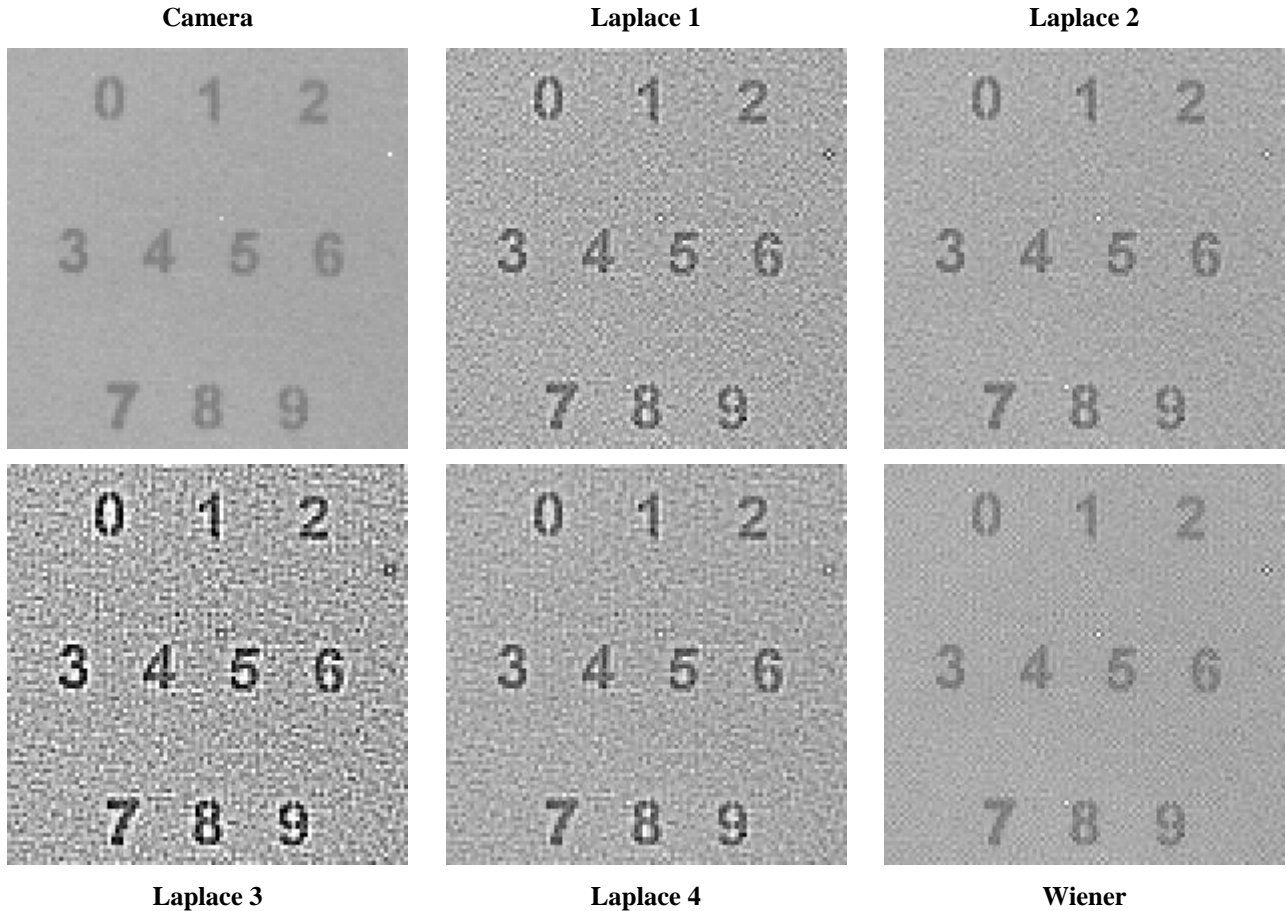
Figure 4. Numbers recorded from a distance of 1.15 m. The differences in sharpness and noise caused by the boost filter are obvious.

Only one observer participated at the experiment, a 54 aged male with corrected eyesight and high familiarity with infrared imagery. This observer also did the MTDP-measurements described before. To minimize strain on the observer each configuration was addressed in a separate session. The duration of such a session was about 7 minutes. Next to the 110 image sequences in test this included 10 introduction images and 0.5 s presentation of a black image in-between two consecutive image sequences. The introduction images were not intended to introduce the observer to the data, but to adapt his mental state to the experiment. All configurations together were seen as one set and ten repetitions of the sets were performed. A random selection of the configurations within the set was used. These 60 experiments were distributed over roughly six weeks, with multiple tests on the same day with breaks in-between possible.

4.4 Analysis and Results

The perception experiment results in probability of correct number identification $P_{Measured}$ versus range R for each configuration. It was corrected for chance probability P_{Guess} to give the actual probability of identification P_{Id} according to[14]

$$P_{Id}(R) = \frac{P_{Measured}(R) - P_{Guess}}{1 - P_{Guess}} \tag{2}$$

With 10 possible answers P_{Guess} equals 0.1 here. For each configuration average and standard deviation of the 10 measurement were calculated. They are given in the following graphs. Fitting the psychometric function is no simple task and a lot of literature is available on this topic, e.g. Klein[14] and Wichmann and Hill[15], both with further references. Here a logarithmic cumulative normal distribution

$$P_{Id}(R) = 0.5 + 0.5 \cdot erfc \left[\frac{-ln\left(\frac{R}{R_{50}}\right)}{\sqrt{2}\cdot 0.24} \right] \qquad (3)$$

was fitted to the data using TableCurve 2D® software[16]. Parameter R_{50} is the range at 50 % probability of identification. The steepness of the transition was variable in a first analysis, but came out quite similar for all configurations. The average was 0.24±0.02 and this value was used for the actual fitting given here. It was also found that for unknown reasons 3.65 m data produced an outlier for all configurations. The value is given in the graphs but was excluded from the curve fitting. Figure 5 gives the resulting data.
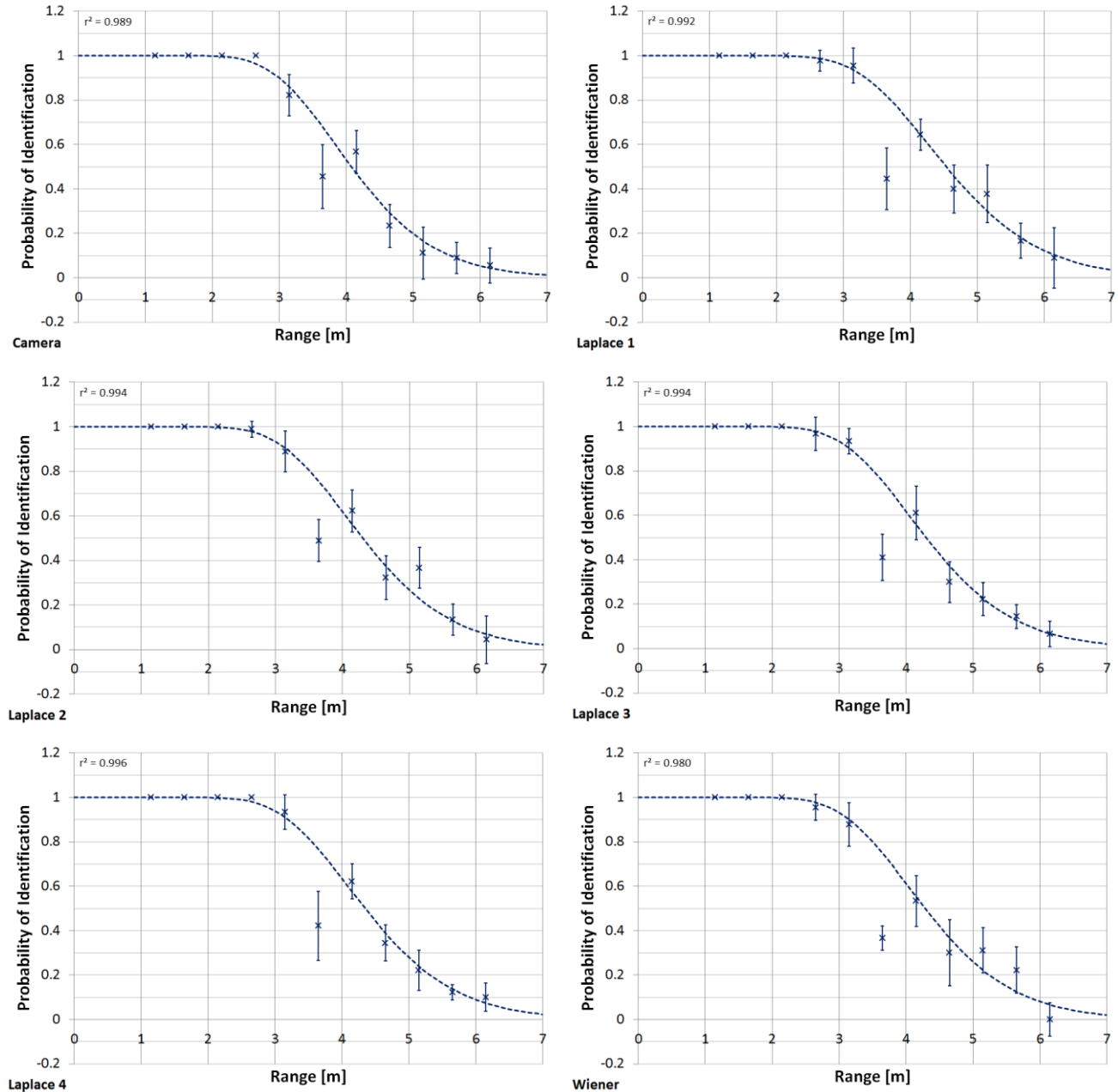

Figure 5. Probability of Identification versus range as measured for the six configurations. Symbols are the average and the corresponding vertical bars the standard deviation of the ten measurements. The dotted lines give the adapted logarithmic cumulative normal distribution and r² gives the correlation coefficient between measurement and curve fit.

5. COMPARING MTDP AND PERCEPTION EXPERIMENT DATA

According to the theory of the Johnson criteria[10,17,18] the connection between range R and limiting frequency $\zeta_{Cut}(\Delta T)$ is given by

$$R = \frac{d_{Target}}{N} \cdot \zeta_{Cut}(\Delta T) \tag{4}$$

using the task difficulty in line pairs on target N, the characteristic size of the target d_{Target} and the target-background contrast ΔT. Rearranging the formula to

$$\frac{R}{\zeta_{Cut}(\Delta T)} = \frac{d_{Target}}{N} = constant \tag{5}$$

shows that for checking if the Johnson criteria apply here, it is sufficient to know range and corresponding limiting frequency for the temperature difference of the target.

Equation (3) shows that the ranges for 50 % probability of identification are directly available from the curve fit. Limiting frequency has to be extracted from the MTDP-measurements for the number temperature difference of 230 mK. Typically, the effect of the atmosphere on temperature difference has to be considered. Because of the small ranges used here this can be neglected. Limiting frequency was calculated from the adapted TRM4.v3 curves by linear interpolation of the closest temperature differences and spatial frequencies. The resulting range and limiting frequency data is summarized in Table 4 and presented as graph in Figure 6. Also presented in Table 3 is the ratio of R_{50} and ζ_{Cut}. With an average of 3.94±0.02 mm it shows only limited variations and this ratio is also plotted in Figure 6. Assuming a letter height of 10 mm this corresponds to 2.54±0.01 line pairs.

	Camera	**Laplace 1**	**Laplace 2**	**Laplace 3**	**Laplace 4**	**Wiener**
R_{50}	4.08 m	4.54 m	4.31 m	4.30 m	4.35 m	4.29 m
ζ_{Cut}	1.033 mrad^{-1}	1.160 mrad^{-1}	1.095 mrad^{-1}	1.089 mrad^{-1}	1.100 mrad^{-1}	1.085 mrad^{-1}
Ratio	3.95 mm	3.91 mm	3.94 mm	3.95 mm	3.95 mm	3.95 mm

Table 4. Range for 50 % probability of identification and corresponding limiting frequency as derived from the measurements. Additionally, the ratio of the two is presented.
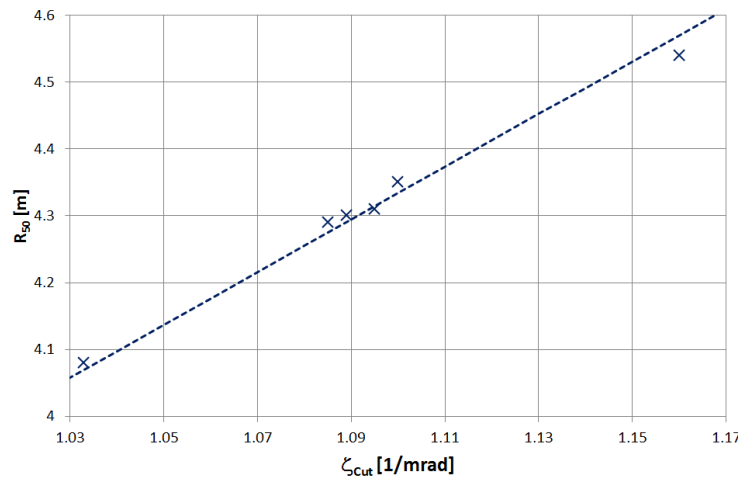
Figure 6. Range versus limiting frequency as derived from MTDP-measurement and perception experiment for the different configurations. The dotted line corresponds to the average ration given in the text.

Another possibility to look on the topic is to calculate the performance gain by applying boost. This can simply be done by calculating relative data based on range and spatial frequency of the camera alone. Table 5 gives the result of this calculation. As shown, predicted and measured gain is practically on spot.

	Camera	**Laplace 1**	**Laplace 2**	**Laplace 3**	**Laplace 4**	**Wiener**
R_{50} relative	1	1.11	1.06	1.05	1.07	1.05
ζ_{Cut} relative	1	1.12	1.06	1.05	1.06	1.05

Table 5. Range and limiting frequency data of Table 3 calculated to relative values with the camera data as reference.

Altogether, perception experiment and MTDP-measurement are in good agreement. This indicates that at least for the camera and boost filter used here the Johnson criterion is able to correctly predict range.

6. CONCLUSION, DISCUSSION AND FURTHER WORK

The presented study aimed on checking if the Johnson criteria applies for thermal imagers with boost filtering. To do so, it compared number identification range to limiting frequency for an undersampled MWIR thermal imager. This imager was operated with and without five different boost filters. Range was measured in perception experiments and limiting frequency was derived from MTDP measurements. As result a very good correspondence between the two figures was found. This indicates the validity of the Johnson criteria for thermal imagers with boost. To value this result some topics have to be considered:

Because of the "high" (230 mK) target-background-contrast used here, limiting frequency often was close to the cut-off frequency of the thermal imager. It was thus not always covered by measured data but had to be derived from model extrapolation. This introduced some kind of subjectivity in the result. In future experiments the contrast should be better adapted to the frequency behavior of the thermal imager.

The study is limited to an undersampled MWIR camera. Repeating it with different imagers is necessary to check if the conclusion is generally valid. Measurements using a LWIR QWIP camera in different configurations already started. They are not finished yet, but first results seem to confirm the findings presented here.

Finally, this study has to be compared to the one by Short et al[2]. Using a similar approach to the one presented here, they found the Johnson criteria to underestimate range when using boost filtering. The major differences between the two studies are well sampled LWIR thermal imager (Short et al) versus undersampled MWIR thermal imager (here), tank targets (Short et al) versus numbers (here) and range modelling (Short et al) versus MTDP-measurement (here). Each of them as well as all together are candidates to explain the conflictive results. However, when analyzing the data (their tables 2, 3 and 4) given by Short et al, it is possible that their result is more indicative for modelling difficulties than the quality of the Johnson criteria for range performance prediction. To assess the differences between the two studies, extensions of the presented work are intended. As discussed above, assessing other cameras is under way. Additionally, a perception experiment for identification of real targets is in preparation.

REFERENCES

1. Weiß, R., Adomeit, U. Chevalier, P., Landeau, S., Bijl, P-, Champagnat, F., Dijk, J., Göhler, B., Landini, S., Reynolds, J. P., Smith, L. N., "A standard data set for performance analysis of advanced IR image processing techniques," Proc. Of SPIE 8355, 835512-2 (2012).
2. Short, R., Littlejohn, D., Bailey, J., Driggers, R., "Infrared sensor performance with boost and restoration filtering," Applied Optics 60(3), 571-579 (2021).
3. Pérez, J., Steiner, D., Keßler, S., "Imager performance assessment with TRM4 version 3: an overview," Proc. Of SPIE 11866, 11866-13 (2021).
4. Russ, J. C., [The Image Processing Handbook 2nd Edition], CRC Press, Boca Raton Ann Arbor London Tokyo (1995).
5. https://forums.ni.com/t5/Example-Code/Color-and-Grayscale-Wiener-Filter-in-LabVIEW/ta-p/3490972, (16.08.2023).
6. Johnson, J., "Analysis of Image Forming Systems," Image Intensifier Symposium 1958, Reprinted in SPIE Milestone Series, SPIE Vol. 513 Part 2, 761-781 (1985).
7. Lloyd, J. M. [Thermal Imaging Systems], Plenum Press, New York and London (1975).

8. Bijl, P., Toet, A., Valeton, J. M., [Electro-Optical Imaging System Performance Measurement], in Driggers, R. G. (ed.), [Encyclopedia of Optical Engineering], Marcel Dekker Inc., New York and Basel (2003).
9. ''STANAG 4347 Definition of Nominal Static Range Performance for Thermal Imaging Systems,'' NATO (1995).
10. ''STANAG 4349 Measurement of the Minimum Resolvable Temperature Difference (MRTD) of Thermal Cameras,'' NATO (1995).
11. ''STANAG 4350 Calculation of Minimum Resolvable Temperature Difference (MRTD) for Thermal Imaging Systems,'' NATO (1995).
12. Wittenstein, W., "Minimum temperature difference perceived - a new approach to assess undersampled thermal imagers," Optical Engineering 38(5), 773-781 (1999).
13. Miller, J. L., Wiltse, J. M., "Resolution requirements for alphanumeric readability," Optical Engineering 42(3), 1-6 (2003).
14. Klein, S. A., "Measuring, estimating, and understanding the psychometric function: A commentary," Perception & Psychophysics 63 (8), 1421-1455 (2001).
15. Wichmann, F. A., Hill, N. J., "The psychometric function: I. Fitting, sampling, and goodness of fit," Perception & Psychophysics 63 (8), 1293-1313 (2001).
16. [TableCurve® 2D Automated Curve Fitting & Equation Discovery Version 5.01], SYSTAT Software Inc., Chicago IL (2002).
17. Dudzic, M. C. (Editor), [The Infrared & Electro-Optical Systems Handbook Volume 4 Electro-Optical Systems Design, Analysis, and Testing], SPIE Optical Engineering Press, Bellingham, Washington (1993).
18. Holst, G. C., [Electro-Optical Imaging System Performance Forth Edition], SPIE Optical Engineering Press, Bellingham, Washington (2006).

Parameterised MRTD

Steinar Børve[a]

[a]Norwegian Defence Research Establishment (FFI), P.O. Box 25, 2027 Kjeller, Norway

ABSTRACT

A large number of factors may influence the performance of thermal surveillance systems used in any given scenario. Highly accurate predictions of acquisition range for a sensor therefore requires the access to specialized numerical tools with a large number of input parameters. At the other hand, simple range estimations with acceptable accuracy can be made for situations of ideal conditions by applying the Johnson Criteria. However, such an approach completely ignores the effect of low signal contrast and atmospheric attenuation and would therefore be unsuited for many real-world scenarios. This work proposes an alternative method, of medium accuracy and complexity, for estimating the acquisition range of thermal sensors. It relies on the well-known concept of Minimum Resolvable Temperature Difference (MRTD), and the method represents the MRTD information for a given sensor by a parametric curve. The form of the parametric curve is chosen so that the observation range can be estimated from a simple second-order equation. The new method has several advantages. First, uncertainties in calculated acquisition range can easily be estimated based on input parameter uncertainties. Secondly, linear approximations can be made for classes of scenarios by making specific assumptions about thermal contrast and atmospheric attenuation. Thirdly, the method can form the basis for a more generalised solver that can handle an even wider range of scenarios. In this work, the new method called Parameterised MRTD (PMRTD) is outlined. Linear approximations to the solutions are derived. In addition, the solution for relevant examples are shown and discussed.

Keywords: Thermal sensors, signature, modelling and analysis of IR systems, defence and security applications of IR sensor technology

1. INTRODUCTION

Target acquisition using thermal sensors is an essential part of many operations, both civilian and military. For this reason, it is of vital importance to have a realistic view of the capabilities of the electro-optical equipment involved in a given scenario. There are many factors that play a role in determining the range capability of a sensor. Some of these factors are determined by the technical specifications of the camera in question, such as the instantaneous field of view (IFOV), but weather and atmospheric conditions can also put severe limitations to the effective observation range. In addition, the characteristics of the observed scene should also be taken into account. What is the target-background contrast? Is the target moving or is it static? What is the orientation of the target and is the target partially covered by other constituents of the scene? These are examples of relevant questions regarding the scene which might affect how easily observable a specific target is in a given scenario. Last but not least, there must be a human or virtual operator which makes the actual observation based on the signal recorded by the sensor. How well a system performs will inevitably depend on the properties of the operator, whether the aim is to *detect*, *recognise*, or *identify* the target, and in the case of a human operator, how the physical and mental state of the operator is at the time of observation.

The Johnson Criteria[1] relates the distance at which an observation task has a certain probability of success to the number of sensor pixels subtending the target in question; it links resolution requirement to the task requirement assuming strong thermal contrast and negligible atmospheric attenuation. One of several possible approaches to extending the Johnson Criteria to include non-ideal effects, is the function called **Minimum Resolvable Temperature Difference (MRTD)**.[2] It is a subjective measure of image quality defined as the minimum temperature difference above 300 K required for an observer viewing through the tested imaging

---

Further author information: (Send correspondence to S. Børve)
S. Børve: E-mail: Steinar.Borve@ffi.no, Telephone: +47-456 23 674

system to resolve a four-bar pattern from the background. When properly established, the MRTD function should reflect the finite temperature difference between target and background and the weather dependent atmospheric attenuation (in a simplified manner). Factors such as detector noise, image processing in the camera and observer performance are all assumed to be incorporated in the experimental data which the method relies on. It is considered quite difficult to obtain good quality, reproducible data for the MRTD function.[2,4] However, theoretical models for the MRTD function also exist.[5]

2. THE PARAMETERISED MRTD (PMRTD) EQUATION

A well-established approach to thermal sensor range prediction combines the Johnson Criteria, the MRTD function and the concept of apparent temperature difference (ATD).[6] If the MRTD function is specified directly from experimental data, it most likely cannot be represented on a compact, analytical form and a solution can therefore only be obtained through a graphical or discreet solver approach. The parameterised MRTD (PMRTD) model is a mathematical approximation to the MRTD model.[7] The idea is to replace an experimentally obtained MRTD curve with a parametric function which fits sufficiently well with the original MRTD curve and which allows the acquisition range to be estimated analytically. Uncertainties in the MRTD curve for the sensor in question should be reflected in the parameters determining the approximate function and result in a corresponding uncertainty in the estimated range.

2.1 MRTD on parametric form

The functional form of the PMRTD function should be chosen both to fit well with typically MRTD functions obtained experimentally as well as provide a basis for deriving a simple analytic expression for the predicted range. Also, it is desirable that the number of parameters needed for specifying the function is kept to a minimum. First, we note that it is common for the MRTD function to be plotted on a logarithmic scale as a function of the normalised spatial frequency of the target, $x = \nu/\nu_N$, where ν_N is the Nyquist frequency of the sensor.[6] The PMRTD function is therefore defined as the logarithm of the target-background temperature difference with 1 K as the normalising unit, hereafter denoted y. Secondly, the PMRTD function should be a non-linear function where the function itself and its first derivative should be monotonically increasing with frequency. The minimum required temperature contrast for detection is given by the zero-frequency value of y, $y(0) = y_{\min} = \log(\Delta T_{\min})$. There should also exist an upper frequency limit, here referred to as the critical frequency $x_c > 1$, above which observations are no longer possible irrespective of temperature contrast. At $x = x_c$ the derivative of the PMRTD function goes to infinity. The last parameter we need is $y_N = \log(\Delta T_N)$, the PMRTD functional value at the Nyquist frequency ($x = 1$).

With these 3 parameters specified, we choose to define the PMRTD function y as

$$y = y_{\min} + D_y \frac{x_c - 1}{x_c - x} x, \tag{1}$$

where we have introduced the quantity $D_y = y_N - y_{\min}$.

Fig. 1 shows two examples where the MRTD data are approximated by 3 PMRTD functions each. The PMRTD parameters used for the PMRTD curves in panel **a**[6] include $\Delta T_{\min} = 0.05$ K and $\Delta T_N = 4.0$ K, In addition, three different options for x_c are shown in the plot; 1.5 (blue curve), 2.0 (green curve), and 3.0 (red curve). We see that $x_c = 2.0$ fits better with the MRTD data for $x = \nu/\nu_N \leq 1$, while $x_c = 3.0$ seems like a better option when looking at $\nu/\nu_N > 1$. In practise, it is probably more important to have a good fit below the Nyquist frequency than above. The MRTD data shown in panel **b** is approximated using the same ΔT_N value as in the previous case, but with $\Delta T_{\min} = 0.025$ K. Here, the three options for x_c are 1.4 (blue curve), 1.8 (green curve), and 2.6 (red curve). A good fit is achieved with $x_c = 1.8$ for $x = \nu/\nu_N \leq 1$, with the exception of the smallest frequencies where the PMRTD curve cannot match the additional drop seen in the original MRTD curve. A strategy for handling this type deviation between the PMRTD function and the original MRTD curve is discussed in the conclusion.
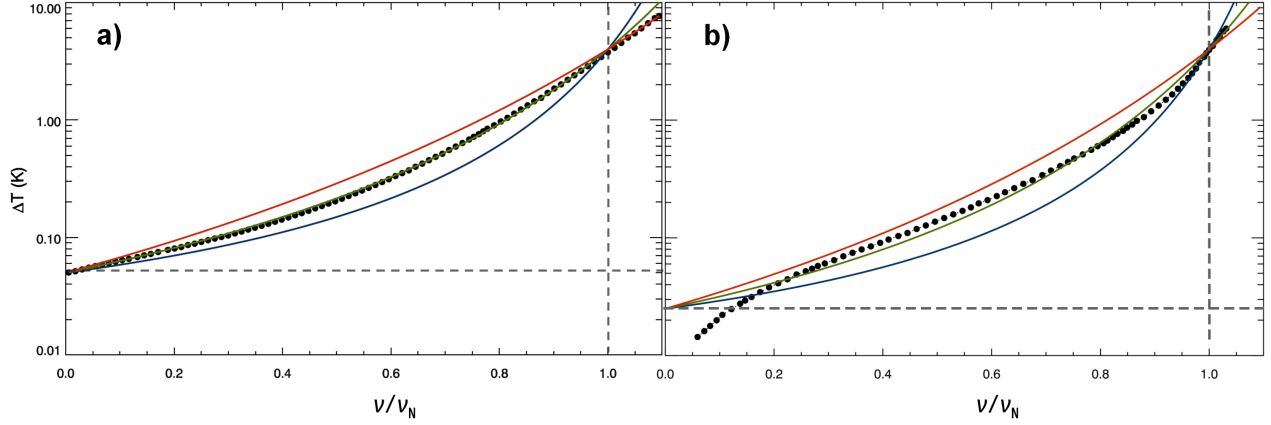
Figure 1. Two MRTD curves plotted as functions of spatial frequency ν normalised by the Nyquist frequency (ν_N). They are taken from Perić et al. (2019)[6] (panel **a**) and Howe (1993)[3] (panel **b**) and are represented by the filled circles. The solid curves represent PMRTD approximations to the MRTD curves. See main text for details.

2.2 Deriving the PMRTD equation

To derive the PMRTD equation,[7] we start with the Johnson Criteria which gives us the Nyquist range, R_N, as:

$$R_N = \frac{L}{2J_{\mathrm{bcc}}\epsilon} \equiv \frac{L}{J_{\mathrm{bcc}}}\nu_N, \tag{2}$$

where we replace ϵ, the sensor IFOV, by the Nyquist frequency, $\nu_N = 1/(2\epsilon)$. In addition, R_N depends on the Johnson bar cycle criteria, J_{bcc}, providing the number of pixels required across the critical target length L in order to achieve the specified probability of success for a given observational task. It is also important to note that since R_N is proportional to ν_N, we also have $R = R_N x$. The apparent temperature difference (ATD) between target and background as measured at distance R is found using Beer's law:

$$\Delta T_a(R) = \Delta T_0 e^{-\gamma R}, \tag{3}$$

where ΔT_0 is the effective target-background temperature difference as measured by the sensor at close range and γ is the total, band-averaged atmospheric attenuation coefficient. By taking the logarithm of Eq. 2, we can express the logarithmic ATD as

$$y_a \equiv \log(\Delta T_a) = y_0 - \gamma \log(e)R = y_0 - \Gamma x, \tag{4}$$

where $y_0 = \log(\Delta T_0)$ and $\Gamma = \gamma \log(e)R_N$.

Since we are interested in the *largest* acquisition range, we set the ATD, given by Eq. 4, equal to the PMRTD function, given by Eq. 1. After some rearrangements, this gives us the easily solvable second-order equation:

$$\Gamma x^2 - \beta x + E_y x_c = 0, \tag{5}$$

where we have defined

$$\beta \equiv \beta_1 + \beta_2 + \beta_3 = \Gamma x_c + E_y + D_y(x_c - 1) \tag{6}$$

and

$$E_y = y_0 - y_{\min} = \log\left(\frac{\Delta T_0}{\Delta T_{\min}}\right). \tag{7}$$

The first term of the sum in Eq. 6, β_1, is proportional to the reduction in logarithmic ATD due to atmospheric attenuation at the critical range $R_N x_c$, while the second therm, β_2, is the effective target-background temperature

contrast relative to the temperature sensitivity of the sensor. The last term, β_3, can be interpreted as the increase in MRTD from the Nyquist frequency to the critical frequency if the MRTD was a linear function.

Solving Eq. 5 gives us the following compact expression for the acquisition range estimate:

$$R = R_N \frac{\beta - \sqrt{\beta^2 - 4\Gamma E_y x_c}}{2\Gamma}. \tag{8}$$

2.3 Linear solutions to the PMRTD equation

Although the solution given by Eq. 8 is already on a highly compact form, it is still a model which relies on a fairly large number of parameters; 4 camera parameters, 2 target-related parameters, and 1 atmospheric parameter (not including parameters related to the underlying atmospheric attenuation model). Therefore, it might be useful to look at three special cases where some of the parameters no longer are important in determining the range. These special cases are identified by which of the three terms in Eq. 6 dominate the expression of β. In all three cases, we will assume that

$$4\Gamma E_y x_c = 4\beta_1 \beta_2 \ll \beta^2. \tag{9}$$

This means that β_1, β_2 or both are negligible compared to β. This assumption enables us to linearise Eq. 8 and thereby obtain a simplified approximation to the normalised range estimate which is given as

$$x \approx \frac{x_c E_y}{\beta}. \tag{10}$$

If $\beta \approx \beta_1$, we have what we could call the **high attenuation approximation**. Combined with the assumption in Eq. 9, this means that the ATD will be low, and that the solution will be in the low frequency regime. The Johnson Criteria range estimate will overestimate the observation range, either because the atmospheric conditions are far from perfect or because the critical target size is large. The predicted observation range can then be expressed as

$$R = R_N \frac{x_c E_y}{\Gamma x_c} = \frac{E_y}{\gamma \log(e)}. \tag{11}$$

In this regime, the predicted observation range depends neither on the critical size of the target nor on any PMRTD parameters other than $y_{\min}$, the global minimum resolvable temperature difference.

When β is dominated by E_y it means that the target-background contrast is very high and the atmospheric attenuation is relatively low. The solution to Eq. 5 will therefore be in the high frequency regime. In this case, Eq. 8 simply becomes

$$R \approx R_N x_c = C_N L, \tag{12}$$

where $C_N = \nu_N x_c / J_{\mathrm{bcc}}$. The observation range in the **high contrast regime** is thus proportional to the critical dimension of the target, L. Increasing the target size by a factor two will therefore lead to a doubled maximum observation range. The expression does not depend explicitly on temperature contrast because it is assumed that the solution is close to the asymptote where the PMRTD curve goes to infinity. If we assume the critical frequency to be equal to the Nyquist frequency, this approximation becomes identical to the Johnson Criteria range estimate. However, if for instance $x_c = 2$ is assumed, the predicted observation range becomes twice the Nyquist range.

Parameter β_3 is only determined by the PMRTD curve of the sensor in question. Since neither thermal contrast nor target size affects this parameter, an especially useful approximation to the observation range can therefore be found in the limit where β_3 dominates β. We refer to this as a **low contrast and low attenuation approximation**. The estimated range in this case becomes:

$$R = R_N \frac{x_c E_y}{\beta} = \frac{C_N}{D_y(x_c - 1)} Z_T. \tag{13}$$

We have here introduced the scenario-specific parameter

$$Z_T = L E_y = L \log\left(\frac{\Delta T_0}{\Delta T_{\min}}\right), \tag{14}$$

which combines the critical dimension of the target L with the ratio of the target-background temperature difference to the minimum detectable temperature difference on a logarithmic scale. Hereafter we refer to Z_T as the **thermal size** of the given combination of target, background and sensor. The predicted range in Eq. 13 serves as an upper limit for low-to-moderate temperature contrast scenarios applicable to near perfect weather conditions.

2.4 Comparsion of linearised and non-linearised solutions

We will illustrate the properties of the PMRTD model by estimating the acquisition range for a fairly large sample of randomly chosen target-background combinations. The assumed MRTD curve is given by Fig. 1a with PMRTD parameters $\Delta T_{\min} = 0.05$ K, $\Delta T_N = 4.0$ K, and $x_c = 2$. For the purpose of simplicity, we will for now assume that the uncertainty in these parameters is negligible. The Nyquist frequency for this sensor is about 38.5 mrad^{-1}.[6]
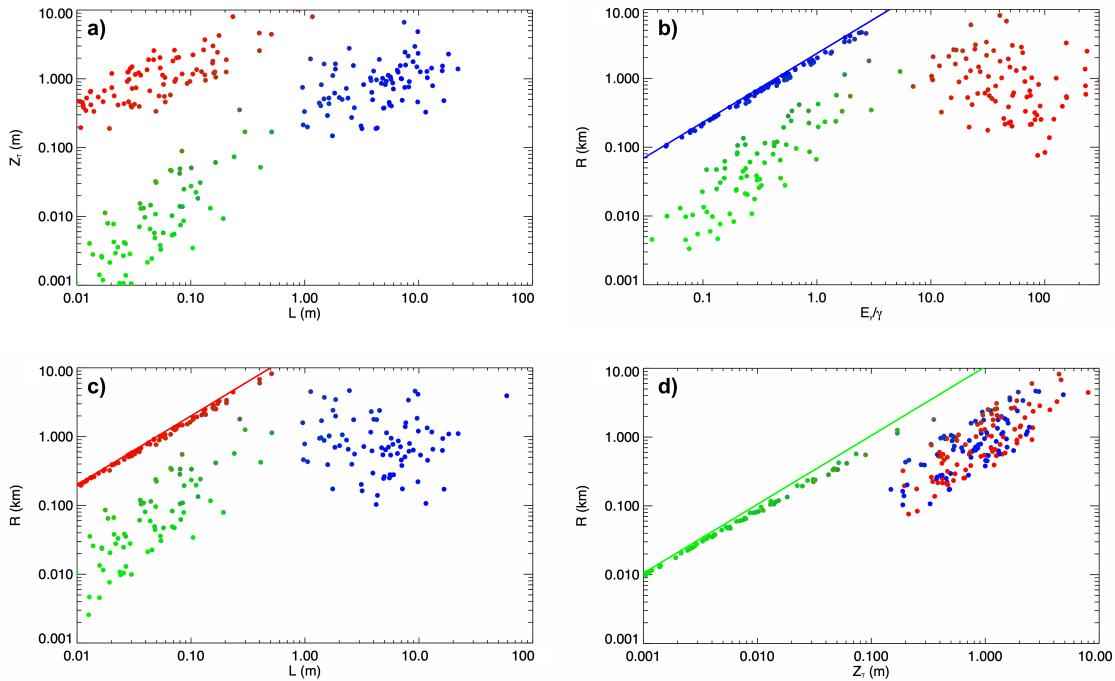

Figure 2. Predicted range according to PMRTD for target recognition with 50% success rate assuming a sensor with MRTD curve. Each dot represents a target-background measurement, and the colours blue, red, and green indicate which of the three terms, β_1, β_2, and β_3 dominate in Eq. 8. Panel **a** plots thermal size (Z_T) as a function of critical size (L), while panels **b-d** plot the predicted range as a function of E_y/γ, L, and Z_T, respectively. The blue, red, and green solid lines indicate the linear approximations given in Eqs. 11-13.

Combinations of target size, target-background temperature difference, and atmospheric attenuation have been randomly selected for 200 test cases. This is done so that either β_1, β_2 or β_3 dominates the expression of β. Fig. 2a plots all test cases in an $L - Z_T$-plot. The colours indicate which of the three β-terms, β_1 (blue dots), β_2 (red dots) or β_3 (green dots), dominate. The recognition distance with 50% probability of success is plotted as a function of E_y/γ, L, and Z_T in panels **b**, **c** and **d**, respectively. The solid line in each of the three panels indicates the relevant approximation given by Eq. 11, 12 or 13. Panel **b** confirms that the predicted observation distance is close to a linear function of E_y/γ when β is dominated by β_1. Similarly, we see in panel **c** that observation distance is almost linearly dependent on the critical size L when β is dominated by β_2. And finally, that R is nearly a linear function of Z_T when β_3 is the main contributor to the parameter β. Notice how the non-linear PMRTD solutions are all below the linear approximations.

3. TEST CASES

In this section we will look at some specific examples of how the PMRTD model can be used. Specifically, we will illustrate how the model can be useful in multi-dimensional parameter studies. Since the only target parameters relevant to the model is the critical length and the effective temperature (surface temperature measured under the assumption of black-body radiation), targets are identified by these two properties only.

3.1 Detection range as a function of atmospheric attenuation

The first example is detection of a target with 50% probability of success where the critical dimension of the target, L, is estimated to be about 1 m. This could for instance be a human. The sensor used in the operation has an assumed IFOV of $\epsilon = 0.12$ mrad. This gives a Nyquist range for the sensor-target pair of $R_N = 4.2$ km. The PMRTD parameters for the sensor are given as follows: $\Delta T_{\min} = 0.015$ K, $\Delta T_N = 0.25 - 4.0$ K, and $x_c = 1.4 - 2.0$. Fig. 3 plots the PMRTD-estimated detection range as a function of the atmospheric attenuation, ranging from 0 km^{-1} to 4 km^{-1}. The uncertainty in the predicted detection range due to the uncertainties in ΔT_N and x_c is indicated by the light blue shaded region, while the solid black curve represents the average prediction. The dashed lines indicate the range estimated with one specific combination of the PMRTD parameters, and the grey solid line marks the Nyquist range that would be obtained from the Johnson Criteria.
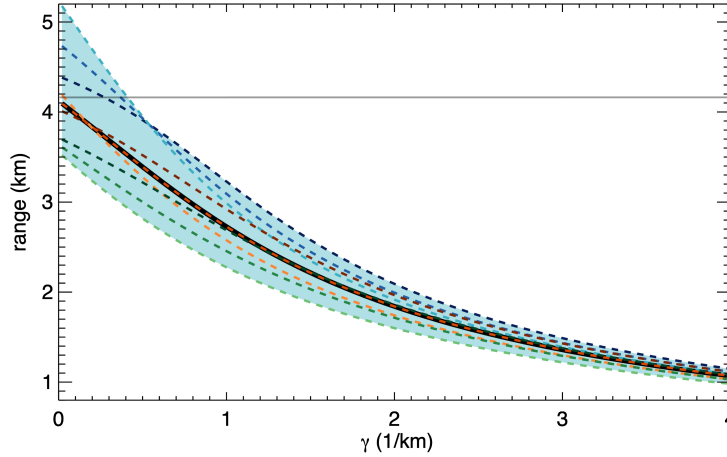

Figure 3. PMRTD-estimated range for 50% successful detection of a 1-m tall target as a function of the extinction coefficient (γ) using an $\epsilon = 0.12$ mrad sensor. The target-background difference, ΔT_0 is 2 K, while the sensor has a minimum temperature sensitivity of $\Delta T_{\min} = 0.015$ K. For the two remaining PMRTD parameters, ΔT_N is 0.25 K (blue curves), 1.0 K (orange/red curves), and 4.0 K (green curves), while x_c is chosen to be 1.4 (darkest shading), 1.8 (medium shading), and 2.6 (brightest shading). The grey line indicates the Nyquist range (R_N).

With no atmospheric attenuation, the PMRTD estimate with average model parameters is very close to R_N. Taking into account the uncertainty of the PMRTD parameters, dominated by the uncertainty in ΔT_N, the zero attenuation range estimate is found to be in the interval 3.5-5.2 km with an estimated uncertainty of about ± 20 %. As the attenuation increases, the estimated range drops as expected. With $\gamma = 1$ km^{-1}, the estimated range of the average model is down to 2.8 km, or about 2/3 of R_N. The relative uncertainty due to the PMRTD parameters also drops as γ increases, and when γ has reached 4 km^{-1}, the range is estimated to be about 1.1 km with an uncertainty of ± 10 %.

3.2 Detection range as a function of temperature difference and IFOV

In the next example, we want to compare the detection range (with a success rate of 50%) predicted by PMRTD to the Nyquist range, the range predicted from the Johnson Criteria. The target size is still kept at 1 m while the atmospheric attenuation is fixed at a relatively high value, 1.2 km^{-1}. This time, the IFOV of the sensor is allowed to vary, in the range 0.05-0.5 mrad, and the target-background temperature difference is varied in the range 0.02-60 K. The PMRTD parameters of the sensor are set as follows: $\Delta T_{\min} = 0.015$ K± 10%, $\Delta T_N = 1.0$ K± 50%, and $x_c = 1.8 \pm 20$%.

Fig. 4 shows the ratio of the PMRTD predicted detection range to the Nyquist range for 17×11 combinations of the IFOV and the temperature difference. For each of these cases, 2^3 combinations of the three PMRTD parameters are tested. The minimum and maximum range estimates found among the PMRTD-combinations are recorded. In Fig. 4, the lower prediction limit is indicated by the colour of the smaller square while the upper prediction limit is indicated by the corresponding larger square. Yellow corresponds to the PMRTD estimate to be roughly equal to the Nyquist range.
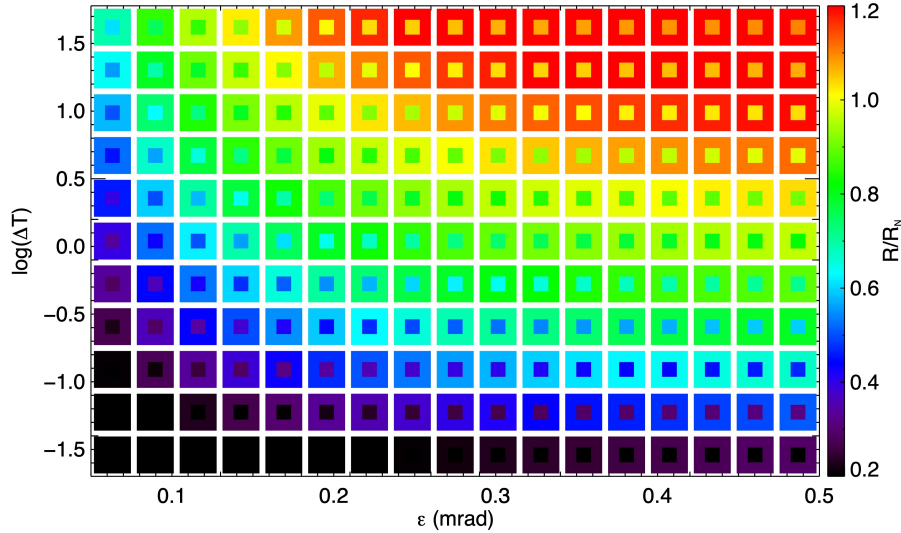

Figure 4. PMRTD-estimated range for 50% successful target detection as a function of the IFOV (ϵ) and the temperature difference on a logarithmic scale. The critical length of the target is set to 1.0 m and the atmospheric attenuation is chosen to be 1.2 km^{-1}. The predicted detection range is normalised by the Nyquist range (R_N). The colours of the smaller squares indicate the lower prediction limit, while the colours of the larger squares mark the upper prediction limit.

Since sensor resolution is inversely proportional to ϵ, the theoretical detection range, as given by R_N, goes to infinity as ϵ approaches 0 mrad. In reality, atmospheric attenuation and finite temperature contrast will be a limiting factor on the range when ϵ is sufficiently small. We see in Fig. 4 that the PMRTD model captures this phenomena. As ϵ is reduced, the temperature difference ΔT must increase in order to keep R/R_N constant. For the chosen target size, atmospheric attenuation and $\epsilon - \Delta$T domain, R/R_N ranges from under 0.05 to just over 1.3. The uncertainty in the predicted detection range due to uncertainties in the PMRTD parameters are on average around 12% with a maximum uncertainty of 33%.

3.3 Importance of view angle on the recognition range for a high aspect ratio target

The last test case, is an example of a more specific analysis performed with the PMRTD model. The targets in question is similar in size to a bus with length equal to 12 m and a height of 3.5 m. The target width is either 2.4 m (case **a**) or 1.2 m (case **b**), corresponding to a length-to-width aspect ratio of 5 and 10, respectively. The high aspect ratio causes the apparent cross section of the target to vary substantially with the observer's viewing angle. The critical length, to which the Nyquist range is proportional, is defined as the square-root of the cross-section area. It is therefore relevant to investigate to what extent the view angle would affect the range at which a given sensor can perform a specified acquisition task. In the current example, we assume the goal is to recognise the targets with a 50% chance of success. The sensor IFOV is chosen to be 0.1 mrad, while the PMRTD parameters for the sensor are $\Delta T_{\min} = 0.02$ K±10%, $\Delta T_N = 1.0$ K±30%, and $x_c = 1.8$±20%. Results are obtained for atmospheric attenuation (γ) in the range 0.2-2.4 km^{-1}. In practical use, there might be considerable uncertainty associated with γ. In this example, we therefore assume that the accuracy of γ is 10%.

Recognition distance for the targets with the specified sensor are calculated for 20 different values of the view angle in the range 0° to 90°, where 0° corresponds to the smallest surface area of the targets being directed

towards the observer. The temperature difference between the targets and their surroundings is set to 2 K for all target surfaces. As before in this paper, we do not consider possible differences in surface emissivity. The results of this simple analysis is shown in Fig. 5 for γ equals 0.2 km^{-1} (black curve), 0.6 km^{-1} (turquoise curve), 1.2 km^{-1} (blue curve), and 2.4 km^{-1} (red curve). The thickness of the curves indicates the uncertainty in the predicted range due to the uncertainty in PMRTD parameters and in γ. Panels **a** and **b** show results for aspect ratio 5 and 10, respectively.
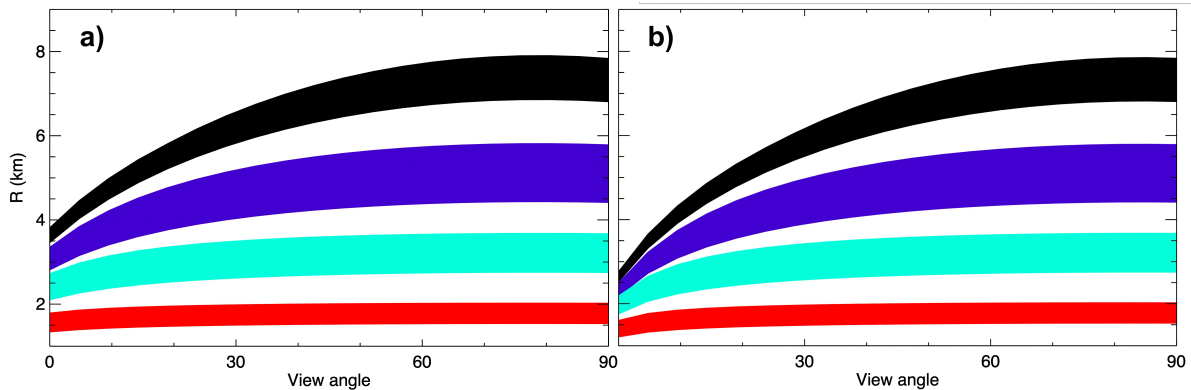

Figure 5. PMRTD-estimated range for 50% successful recognition of targets with length 12 m, height 3.5 m and an aspect ratio of 5 (panel **a**) and 10 (panel **b**) plotted as a function of the view angle assuming a target-background temperature difference of 2 K. The atmospheric attenuation is chosen to be 0.2 km^{-1} (black curve), 0.6 km^{-1} (turquoise curve), 1.2 km^{-1} (blue curve), and 2.4 km^{-1} (red curve) with an assumed accuracy of 10%.

Let us first consider the case with aspect ratio equal to 5. For this combination of sensor and target, we see that the recognition range dependence on view angle is small when compared to the input parameter uncertainties if the atmospheric attenuation is large and the range therefore is small. Take $\gamma = 2.4$ km^{-1} as an example; the model-averaged range increases from around 1.6 km to roughly 1.8 km as the view angle increases from 0° to 90°. In comparison, the prediction uncertainty is around ±0.2 km. This implies that even if the view angle were completely unkown, we still could limit the recognition range for the $\gamma = 2.4$ km^{-1} case to the interval 1.4-2.0 km. For $\gamma = 1.2$ km^{-1}, the relative importance of the view angle is slightly larger. The model averaged recognition range increases from around 2.4 km at 0° to 3.2 km at 90°. The PMRTD model uncertainty in this case is in the range ±0.3 − 0.5 km. As the atmospheric attenuation decreases further and the recognition range increases correspondingly, both the range prediction itself and the model uncertainty increases notably with increasing view angle. For $\gamma = 0.6$ km^{-1}, the average range increases from 3.1 km at 0° view angle to 5.1 km at 90° view angle. The model uncertainty correspondingly increases from ±0.3 km to ±0.7 km. Finally, for very good atmospheric conditions ($\gamma = 0.2$ km^{-1}), the model averaged recognition range increases from around 3.6 km at 0° to a maximum of 7.4 km occurring at a view angle of around 80°. In this case, the model uncertainty is less important, increasing from approximately ±0.1 km to ±0.5 km.

The target aspect ratio does not affect the recognition range when the targets are viewed at 90° since the target length and height is the same in both case **a** and **b**. However, when the targets are view head-on, the doubled aspect ratio in case **b** causes the recognition range to drop signficantly when compared with case **a**, in particular in good weather conditions. The drop in the model averaged estimated recognition range at 0° view angle is from around 18% when $\gamma = 1.2 − 2.4$ km^{-1} to around 25% when $\gamma = 0.2 − 0.6$ km^{-1}.

4. CONCLUSION

In this paper, I have presented the method Parameterised MRTD (PMRTD) for easily estimating the acquisition range of thermal sensors which combines the Johnson Criteria, the MRTD function and the concept of apparent temperature based on Beer's law. This approach to obtaining range estimates is well known from the literature.[6] However, an essential ingredients to the current method is the introduction of a suitable parameterised approximation to the MRTD function. This makes it possible to obtain an acquisition range estimates which takes into

account both non-linear system performance and atmospheric attenuation from a simple second order equation. The simplicity of the solution therefore makes it very well suited for multi-dimensional parameter studies. The author is not aware of previously published work where this particular approach to sensor range estimation has been applied.

A limitation to the PMRTD method is to what extent the MRTD function can be sufficiently accurate represented by the chosen parametric form and how accurately the functional parameters can be specified. In Fig. 1b, we saw an example where the original MRTD function had low-frequency characteristics which were not accurately reproduced by the global parametric approximation. A solution in this case would be to split the frequency domain into several zones, e.g. one zone covering frequencies below roughly $0.2\nu_N$, one zone for the mid frequencies ($0.2 < \nu/\nu_N < 0.8$) and one zone covering all frequencies above $0.8\nu_N$. For each zone, a separate set of PMRTD parameters can be determined, with the additional requirement of continuity at the zone borders. When estimating the acquisition range in a given scenario, Eq. 8 should be solved for each set of PMRTD parameters. Since the PMRTD function is still continuous with a positive derivative, one and only one of the zones will provide a solution which is within the valid frequency range of the zone. In principle, an arbitrary number of zones might be constructed and thus making the PMRTD function in reality able to approximate any continuous, monotonically increasing MRTD function.

In its current (and simplest form), the method is also restricted to using the broadband Beer's law in order to calculate the apparent temperature difference (ADT).[8] This also implies that the target and background is assumed to be more or less co-located. A generalisation of the method to incorporate a more accurate and flexible expression of ADT is possible. However, this would come at the expense of increased complexity as Eq. 8 then will have to be replaced by an iterative solver. This solver will be less suited for quick, analytic estimations, but will have the potential of handling more complex scenarios involving e.g. surface emissivity effects and non-negligible target-background separation, as well as allow more accurate ADT models to be used. Work is in progress to explore some of these possibilities.

REFERENCES

[1] Johnson, J. **Analysis of Image Forming Systems**, Proc. Image Intensifier Symp.,U.S. Army Eng. Res. Development Lab, Ft. Belvoir, AD 220-160, 249-273, 1958

[2] Holst, G.C. **Infrared Imaging System Testing**, in "The Infrared & Electro-Optical Systems Handbook. Volume 4: Electro-Optical Systems Design, Analysis, and Testing", M.C. Dudzik (Ed.), SPIE Optical Engineering Press, Bellingham, USA, 195-243, 1993

[3] Howe, J.D. **Electro-Optical Imaging System Performance Prediction**, in "The Infrared & Electro-Optical Systems Handbook. Volume 4: Electro-Optical Systems Design, Analysis, and Testing", M.C. Dudzik (Ed.), SPIE Optical Engineering Press, Bellingham, USA, 55-120, 1993

[4] Krapels, K. et al. **Minimum resolvable temperature difference (MRT): procedure improvements and dynamic MRT**, Infrared Phys. & Tech., 43, 17-31, 2002

[5] Ratches, J.A. **Static Performance Model for Thermal Imaging Systems**, Optical Engineering, 15, 525-530, 1976

[6] Perić, D. et al. **Thermal Imager Range: Predictions, Expectations, and Reality**, Sensors, 19, 3313-3325, 2019

[7] Børve, S. **(U) Thermal sensor acquisition range estimation**, Norwegian Defence Research Establishment (FFI), 2021

[8] Vinson, J.G., Driggers, R.G. & Deep, R. **Teqhniques and Errors for Apparent Differential Temperature Calculations**, SPIE Proc. Infrared Imaging Systems: Design, Analysis, Modeling, and Testing IX, 3377, 2-10, 1998

Impact of motion blur on recognition rates of CNN-based TOD classifier models

Daniel Wegner[a] and Stefan Keßler[a]

[a]Fraunhofer IOSB, Gutleuthausstr. 1, 76275 Ettlingen, Germany

ABSTRACT

This work investigates the impact of various types of motion blur on the recognition rate of triangle orientation discrimination (TOD) models. Models based on convolutional neural networks (CNNs) have been proposed as an automated and faster alternative to observer experiments for range performance assessment. They may also give insights into the impact of system degradations on the performance of automated target recognition algorithms. However, the effects of many image distortions on the recognition rate of such models are relatively unknown. The recognition rate of CNN-based TOD models is examined in terms of different forms of motion blur, such as jitter, linear and sinusoidal motion. For model training and validation, simulated images are used. Triangles with four directions and different sizes, positions are used as targets, which are superposed on natural images as background taken from the image database "Open Images V7". Motion blur of varying strength is applied to both the triangle and the entire image to simulate movements of the target and imager. Additionally, common degradation effects of imagers are applied, such as white sensor noise and blur due to diffraction and detector footprint. The recognition rates of the models are compared for target motion and global motion as well as for the different motion types. Furthermore, dependencies of the recognition rate on blur strength, triangle size and noise level are shown. The study shows interrelationships and differences between target motion and global motion regarding TOD classifications. The inclusion of motion blur in training can also increase model accuracy in validation. These findings are crucial for range performance assessment of thermal imagers for fast-moving targets.

Keywords: CNN, triangle orientation discrimination, motion blur, target motion, camera stabilization, imager assessment

1. INTRODUCTION

Motion blur is an image degradation occurring and used in a wide field of applications. In photography and film recording, motion blur is applied to increase the perceived motion and let recorded images appear more realistic.[1] In games industry, simulated motion blur is used for temporal anti-aliasing and make fast animations appear smoother.[2] Several tracking methods have been proposed for visual tracking applications under motion blur.[3] Thermal seekers of missiles, e.g. for air defence, are subject to motion blur due to flutter, rotation and moving targets. Also uncooled microbolometer cameras used in applications such as remote sensing can suffer significantly from motion blur of moving objects due to the slow decay of electrical signal, with decay times of tens of milliseconds[4].[5]

For imager performance assessment, a well-established approach is triangle orientation discrimination (TOD).[6] The original idea of this method was that human observers discriminate between four orientations of a triangle shown on a display. These presented images are degraded depending on several camera-specific properties leading to varying blur due to optical diffraction, aberration, detector footprint as well as temporal and fixed pattern noise. These degradations impair the ability for discrimination making the accuracy a measure for imager performance. For automated imager assessment, models for TOD classification based on convolutional neural networks (CNN)[7] trained and validated on artificial images have been presented. While these models have been investigated for some degradations such as Gaussian sensor noise and clutter variance, the impact of target and global motion blur on model accuracy is yet unexplored.

Further author information: (Send correspondence to D. W.)
D. W.: E-mail: daniel.wegner@iosb.fraunhofer.de

In this paper, several TOD models trained on artificial images are validated on images with target and global motion at varying strengths. Different types of motion such as linear motion, sinusoidal motion, jitter, zooming and rotation are investigated.

The paper is structured as follows: In Section 2 a framework for simulating target and global motion is presented. Based on this framework, images with single triangles are superposed with natural backgrounds and impaired by target and global motion blur. These images are used to train and validate TOD models. Section 3 presents the results of validating several models with different contributions of motion blur in the training for varying target and global motion. Finally, Section 4 concludes the paper.

2. METHODS

2.1 Model architecture and training

Models for triangle orientation discrimination can be formed by convolutional neural networks. The model architecture and procedure for model training are adopted from a former work.[7] Such models predict the orientation of single triangles contained in an image with 64x64 pixels, with 4 possible orientations: "left", "right", "up", and "down". In the former work several image degradations were taken into account for data augmentation to increase model robustness such as white Gaussian sensor noise as well as Airy blur and Gaussian blur representing optical diffraction and aberration. In this paper, target and global motion blur are used in addition for model training and validation. Target motion basically impairs the sharpness of the image section covered by the target and leads to blending effects with the background at the target edges due to partial occlusion of the background. The major part of the background is unaffected by target motion. In contrast, global motion caused by camera movements blurs the entire image. For a linear motion shift the image blur is uniform.

2.2 Calculation of target and global motion blur

Target motion blur can be calculated by affine transformation of an image representing a single target $T_0(x, y)$. $T_0(x, y)$ are floating-point numbers within the dynamic range of $[0, 1]$. Then shifted versions of a target $T_k(x, y)$ with N frames, i.e. $k \in [0, N-1]$, can be calculated by affine transformation as

$$T_k(x, y) = T_0(A_{target,k}(\vec{r} - \vec{r}_{center,target}) + \vec{r}_{center,target} + \vec{b}_{target,k}), \tag{1}$$

with a matrix A_i, the pixel coordinates $\vec{r} = \begin{pmatrix} x \\ y \end{pmatrix}$, the rotation and zoom center $\vec{r}_{center,target}$ and shift vectors $\vec{b}_k$. These N frames can be thought of as the projections of the input scene with varying positions of the target captured and integrated over the integration time t_{int}. If rotation, zooming and shearing are neglected, A_k is the unit matrix.

With a mask normalized to $[0, 1]$, i.e.

$$M_k(x, y) = \frac{1}{T_{k,max} - T_{k,min}} \cdot T'_k(x, y) + 1 - \frac{T_{k,max}}{T_{k,max} - T_{k,min}}, \tag{2}$$

$$T_{k,max} = \max_{x,y} T'_k(x, y), \tag{3}$$

$$T_{k,min} = \min_{x,y} T'_k(x, y), \tag{4}$$

and

$$T'_k(x, y) = \begin{cases} T_k(x, y) & \text{positive contrast} \\ 1 - T_k(x, y) & \text{negative contrast} \end{cases}, \tag{5}$$

this moving target can be overlayed with a background image $B(x,y)$:

$$I_k(x,y) = M_k(x,y)T_k(x,y) + (1 - M_k(x,y))B(x,y). \tag{6}$$

If then global motion is applied on the entire image, e.g. due to camera movements, the shifted image for the frame k can be calculated as

$$I'_k(x,y) = I_k(A_{global,k}(\vec{r} - \vec{r}_{center,global}) + \vec{r}_{global} + \vec{b}_{global,k}) \tag{7}$$

with matrix $A_{global,k}$ and shift vector $\vec{b}_{global,k}$. Finally, the blurred image can be calculated as

$$I_{average}(x,y) = \frac{\sum_{k=0}^{N-1} I'_k(x,y)}{N}. \tag{8}$$

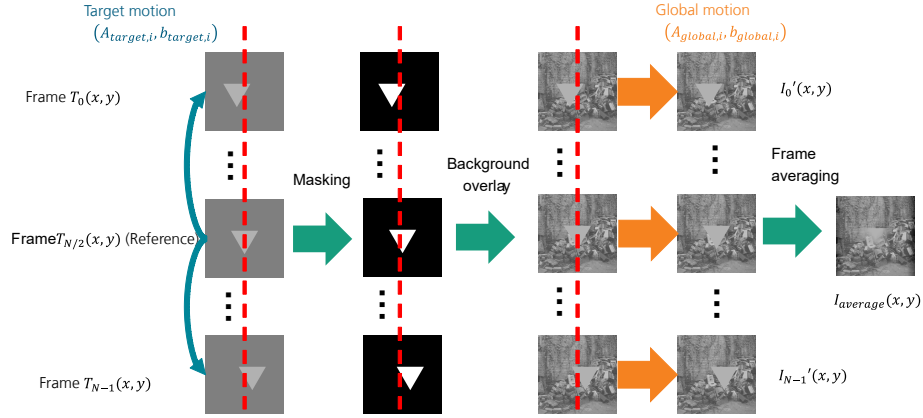The procedure of overlay is depicted in Figure 1.

Figure 1. Target and global motion applied on a single triangle target with background overlay (from Open Images V7).

In order to preserve the mean position of the target, horizontal and vertical shifts for linear and sinusoidal motion are defined as

$$s_{x/y,global/target,linear} = \frac{k - N/2}{N} \cdot s_{x/y,linear,max}, \tag{9}$$

$$s_{x/y,global/target,sinusoidal} = \sin\left(2\pi\frac{k}{N}\right) \cdot s_{x/y,sinusoidal,max}. \tag{10}$$

If a physical target has a velocity $\vec{v}_{target,\parallel}$ parallel to the plane of observation, the maximum motion shift in sensor pixels can be expressed as

$$\vec{s}_{target,max} = \frac{\vec{v}_{target,\parallel} \cdot t_{int}}{2R \cdot IFOV} = \frac{\vec{v}_{target,\parallel} \cdot t_{int} \cdot fl}{2R \cdot p_{x/y}} \tag{11}$$

with integration time t_{int}, target-observer distance R, focal length fl and pixel pitch $p_{x/y}$. The magnitude is given by

$$\left|\vec{v}_{target,\parallel}\right| = \sin\theta \cdot v_{target} \tag{12}$$

with the scalar total target velocity v_{target} and the angle θ between the direction of motion and the line of sight.

Jitter positions $j_{x,k}$ and $j_{y,k}$ can be realized by random sampling from a normal distribution $\mathcal{N}(0, j_{max})$. Then intermediate positions can be calculated by interpolation as

$$j_{x/y,global/target,l} = (1 - r)j_{x/y,l} + rj_{x/y,l+1}, \tag{13}$$

with $r = k/N_{jitter} - \lfloor k/N_{jitter} \rfloor$ and the floor function $\lfloor \circ \rfloor$ rounding to the largest integer smaller than the argument. The number of points per linear trajectory N_{jitter} allows to limit the bandwidth of the jitter and enables a smooth transition to linear motion if N_{jitter} approaches the number of frames N. The total shifts result from summing these three types of motion:

$$s_{x/y,global/target,total} = s_{x/y,global/target,linear} + s_{x/y,global/target,sinusoidal} + j_{x/y,global/target}. \qquad (14)$$

Approximation blur can result when approaching a target, e.g. when a cruise missile or drone is headed for a close target. This blur may be modelled by zoom matrices

$$A_{k,zoom} = \begin{pmatrix} z_k & 0 \\ 0 & z_k \end{pmatrix}, \qquad (15)$$

with zoom factors z_k inversely proportional to the distance between target and imager. If a constant velocity of the imaging platform $v_\perp$ perpendicular to the plane of observation is assumed, z_k can be written as

$$z_k = \frac{R_0}{R_k} = \frac{R_0}{R_0 - k\Delta R} = \frac{R_0}{R_0 - k\frac{v_\perp t_{int}}{N-1}} = \frac{1}{1 - k\Delta z}, \qquad (16)$$

with the ranges $R_k = R_0 - k \cdot \Delta R$ and $z_0 = 1$. With the definition $z_{n-1} := z_{max}$, the zoom factor step Δz can be expressed as

$$\Delta z = \frac{\Delta R}{R_0} = \frac{v_\perp t_{int}}{R_0(N-1)} = \frac{z_{max} - 1}{z_{max}(N-1)}. \qquad (17)$$

Additionally, missiles may spin around the axis of motion, which can be modelled by

$$A_{k,rotation} = \begin{pmatrix} \cos\alpha_k & -\sin\alpha_k \\ \sin\alpha_k & \cos\alpha_k \end{pmatrix}, \qquad (18)$$

with

$$\alpha_k = \left(k - \frac{N}{2}\right) \cdot \alpha_{max}/(N-1). \qquad (19)$$

In this work we use $N = 20$.

2.3 Introducing motion blur in the model training and evaluation

Images containing single triangles with random orientations, contrasts and sizes are generated. Then images for model training and evaluation are created by random sampling $s_{x/y,linear}$, $s_{x/y,sinusoidal}$ and $j_{x/y,global/target}$ from a uniform distribution in $[-20, 20]$ pixels and applying target and global motion. Example images with target and global motion are shown in Figure 2 as well as the corresponding pristine images.
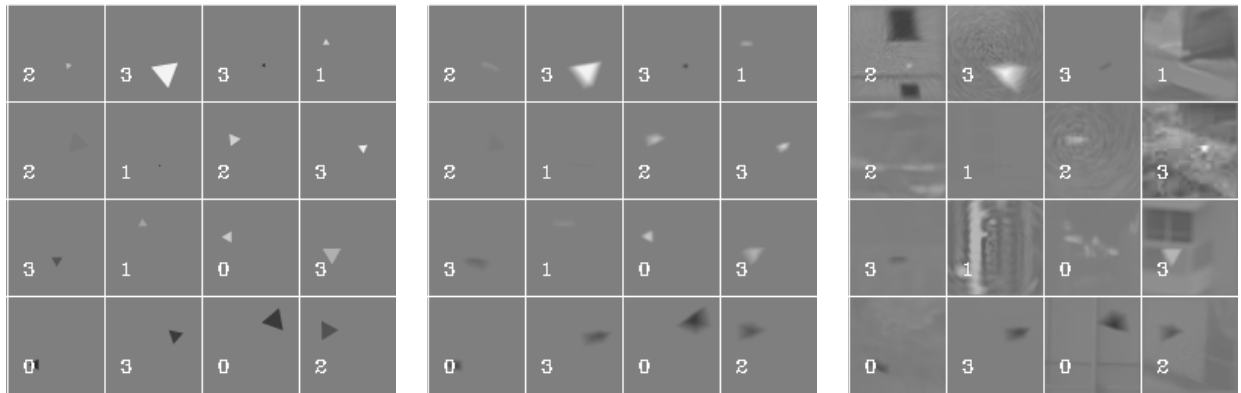

Figure 2. Training examples: Images of single triangles with varying position, contrast, size and orientation with 64×64 pixels: Pristine (left), with additional target motion blur (center), with additional target motion blur and background overlay (right). For each image label is shown, i.e. "left"(0), "up"(1), "right"(2), "down"(3).

In order to increase model robustness against typical camera effects, further degradations are applied such as Gaussian sensor noise and blur due to optical diffraction, detector footprint, and Gaussian blur representing aberration. To investigate the effect on the validation accuracy of whether motion blur is considered in the training phase or not, different models are trained, where different types of motion are toggled. An overview of these models is given in Table 1.

Table 1. Different trained models considered in this paper with different types of motion toggled. Each cell represents the probability of application on single training images.

Model name	Target linear	Global linear	Global sinusoidal	Global jitter	Global zooming	Global rotation	All motion types enabled
M_0	0	0	0	0	0	0	0
M_{single}	1/4	1/4	1/4	1/4	0	0	0
M_{all}	1/5	1/5	1/5	1/5	0	0	1/5
$M_{single.RZ}$	1/6	1/6	1/6	1/6	1/6	1/6	0
$M_{all,RZ}$	1/7	1/7	1/7	1/7	1/7	1/7	1/7

The models $M_{single,RZ}$ and $M_{all,RZ}$ use zoom factors z_{max} randomly sampled from a uniform distribution in $[0.8, 1.2]$ and a rotation with maximum angle α_{max} randomly sampled from a uniform distribution in $[-15°, 15°]$.

2.4 Model architecture

As in the former work[8] a CNN-based architecture is used as shown in Figure 3. A sequence of blocks and fully connected layers is used to calculated four probabilites for directions "left", "right", "up" and "down" from an image with 64×64 pixels.
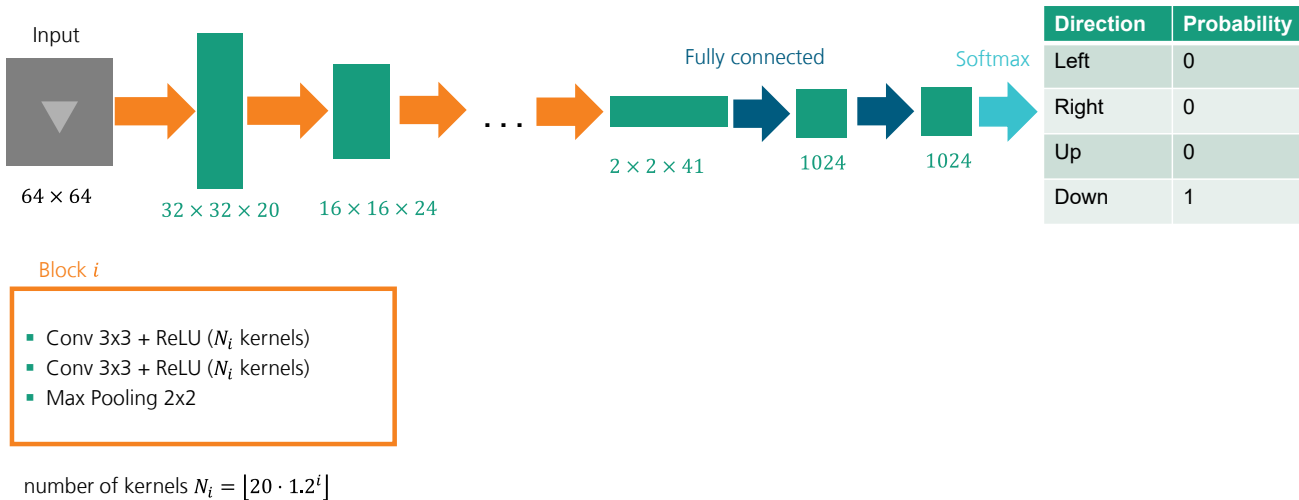

Figure 3. Model architecture for Triangle Orientation Discrimination (TOD): A number of blocks downsamples spatial dimensions and increases number of features N_i. Each block consists of two convolutional layers with 3×3 kernels with ReLU and a subsequent 2×2 max pooling layer. Two fully connected layers with subsequent ReLUs give 4 probabilites for "left", "right", "up" and "down".

2.5 Model training

The model is trained and validated in python/tensorflow. The learning method is ADAM[9] with a learning rate of $\eta = 0.0005$, $N_{epoch} = 1000$ epochs and a batch size of $N_{batch} = 1024$. This gives an amount of $N_{training} = N_{epoch} \cdot N_{batch} = 1024000$ images for training. To counteract overfitting, images for training and validation are generated on-the-fly.

3. RESULTS AND DISCUSSION

3.1 Linear motion

To investigate the impact of whether different types of motion blur are included in the training phase or not, additional models are trained:

- model M_T (only linear target motion),

- model M_G (only linear global motion),

- model M_{TG} (linear target motion and global motion).

These models were evaluated using images with a centered triangle of circumradius $r = 10$ pixel as shown in Figure 4. $N = N_{orientation} \cdot N_{background}$ images are generated by background overlaying of $N_{background} = 1000$ random crops of 64x64 pixels from the database Open Images V7.[10] Either horizontal linear target motion or horizontal linear global motion was applied.
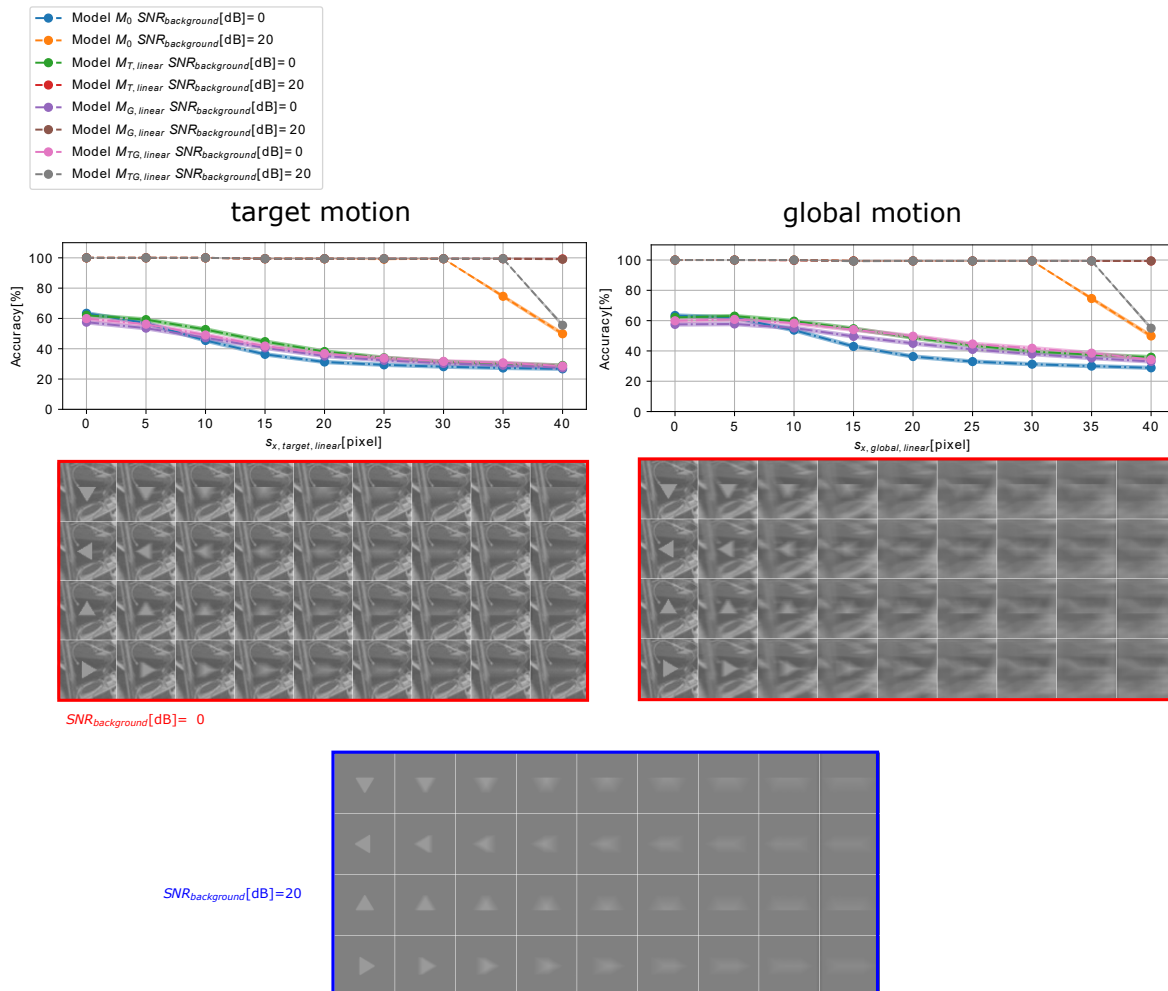

Figure 4. Model accuracies for the models M_0, M_T, M_G and M_{TG} based on $N = N_{orientation} \cdot N_{background} = 4000$ images (top). Example images for model evaluation: A centered triangle with constant contrast to background $SNR_{background} = 0$ dB. From left to right $s_{x,max,linear}$ takes the values 0, 5, 10, 15, 20, 25, 30, 35 and 40 (middle). Example images for an almost negligible background level $SNR_{background} = 20$ dB (bottom).

It can be observed, that for weak backgrounds, i.e. $SNR_{background} = 20$ dB, the model accurarcies are close to 100% with significant drops only from target and global motion shifts $s_{x,target/global,linear} \geq 35$ pixel. For the human observer and the corresponding example images it seems much harder to clearly determine the triangle orientation. For images with strong backgrounds, i.e. $SNR_{background} = 0$ dB, there are smooth decays of the accuracy with increasing motion shifts $s_{x,target/global,linear}$. While the motion-agnostic model M_0 is slightly better for low motion shifts $s_{max,target/global,linear} \leq 5$ pixel, the model M_{TG} including both motion types in the training is better for high motion shifts $s_{max,target/global,linear} > 15$ pixel. For comparable motion shifts, model accuracies are higher for global motion than for target motion. This fact could be understood by the observation, that target motion mostly affects the sharpness of the target, while global motion also blurs the background and hence decreases the fluctuations of gray levels within the background. Therefore, the effective signal-to-background ratio $SNR_{background}$ is increased.

Analogously, evaluations are performed for trained models with varying maximal motion shifts $s_{x/y,linear,max} \in [2, 5, 10, 20]$ conforming the trend of increasing model accuracies on the validation set for large motion shifts $s_{x/y,linear}$ when increasing the maximal motion shifts in the training, but also with decreasing model accuracies on the validation set for small motion shifts $s_{x/y,linear/sinusoidal/jitter} \leq 5$ pixel.

3.2 Dependency of model accuracy on triangle size

In addition to a horizontal shift $s_{x,target}$ and $s_{x,global}$ for target and global motion, also the triangle circumradius r[pixel] is varied for a centered triangle and random crops of Open Images V7 validation images are used as backgrounds. Then the model M_0 is applied on $N = N_{orientation} \cdot N_{background}$ images for different combinations of $s_{s,target/global}$ and r. Strong backgrounds with $SNR_{background} = 0$ dB are used, as target motion and global motion are close to identical with weak or no backgrounds. The corresponding model accuracies are shown in Figure 5.
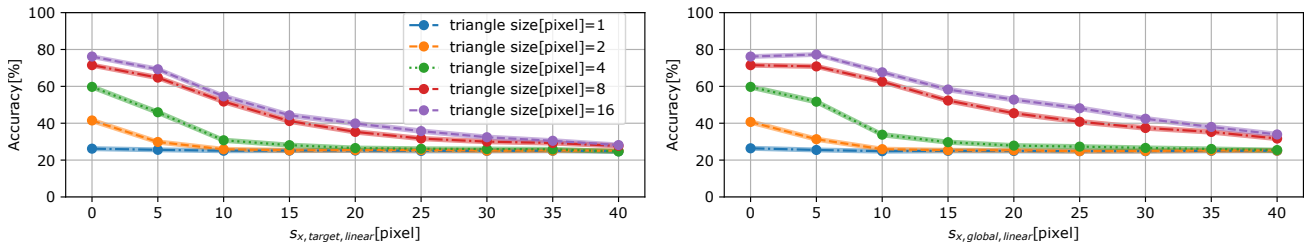

Figure 5. Model accuracy based on $N = N_{orientation} \cdot N_{background} = 4000$ images with different triangle circumradii r[pixel] for target motion (left) and global motion (right).

It can be observed that the accuracies are smoothly dropping for increasing horizontal shifts $s_{x,target/global}$ as well as for decreasing triangle sizes. For sufficiently large triangles, i.e. $r \geq 4$ pixel, accuracies are slightly higher for global motion compared to target motion.

3.3 Dependency of model accuracy on noise level

In an analogous way, horizontal shifts $s_{x,target}$ and $s_{x,global}$ for target and global motion were varied for different sensor noise levels SNR_{noise} and fixed triangle circumradius $r = 10$ pixel. Each combination is validated on $N = N_{orientation} \cdot N_{background} = 4000$ images based on $N_{orientation} = 4$ orientations and $N_{background} = 1000$ backgrounds. The resulting model accuracies for the model M_0 are shown in Figure 6.
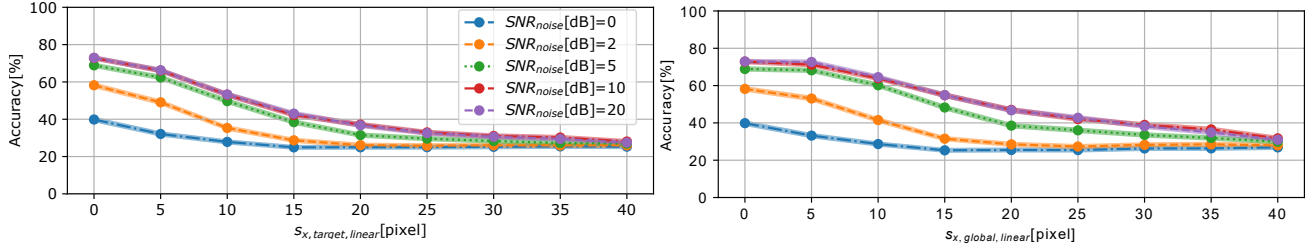
Figure 6. Model accuracy based on $N = N_{orientation} \cdot N_{background} = 4000$ images with different SNR_{noise}[dB] for horizontal target motion (left) and global motion (right).

Again the accuracies are smoothly decaying for increasing horizontal shifts $s_{x,target/global}$ and decreasing signal-to-noise ratio SNR_{noise}. For $SNR_{noise} \geq 5$ dB, accuracies are slightly higher for global motion compared to target motion.

3.4 Joint target and global motion

When the target and the observing camera move simultaneously on the same axis, the directions of these movements are crucial for image blur:

- If the target and global motion shifts are in the same direction, i.e. target and global motion are in opposite directions, the target motion blur and global motion blur are accumulated.

- It the target and global motion shifts are in opposite directions, i.e. target and global motion are in the same direction, the blur effect on the target is reduced by the global motion. The optimal image sharpness of the target is achieved when target motion shift and global motion shift have same absolute values.

In Figure 7 model accuracies of the models M_0, M_{single} and M_{all} are shown for joint target and global motion. Target and global motion is again applied on $N = N_{orientation} \cdot N_{background} = 4000$ images with $N_{background} = 1000$ different random crops from the validation set of Open Images V7.[10] Linear, sinusoidal and jitter global motion is considered as cameras may be subject to swivel movements and vibrations, while for the target only linear motion is taken into account.
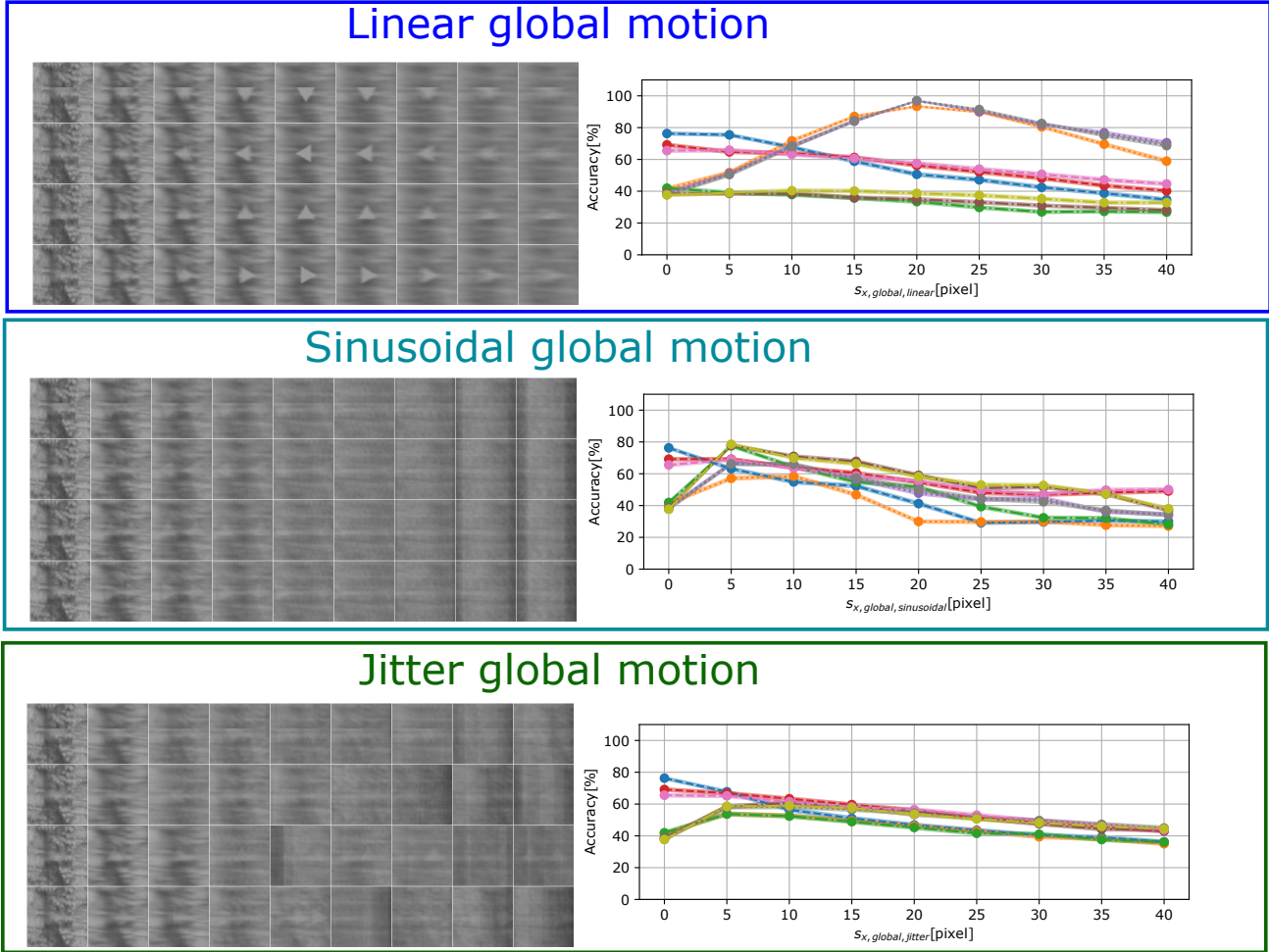
Linear target motion

Linear global motion

Sinusoidal global motion
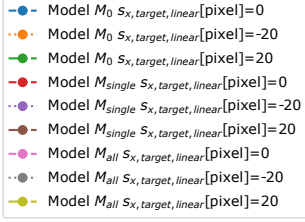
Jitter global motion

Figure 7. Model accuracies for models M_0, M_{single} and M_{all} on $N = N_{orientation} \cdot N_{background} = 4000$ validation images. Horizontal linear, sinusoidal and jitter global motion ($s_{x,global} \in [0, 5, 10, 15, 20, 25, 30, 35, 40]$)
as well as horizontal target motion at 3 different levels $s_{x,target} \in [-20, 0, 20]$ is applied to these images. Collage images show examples for $s_{x,target,linear} = -20$ pixel and the different levels of global motion.

For strong motion shifts, border expansion used in the affine transformations for the calculation of target and global motion can lead to visible artifacts which might degrade the model performance. These border effects can be alleviated by applying the procedure of background extraction with target and global motion on images larger than 64×64 pixels and subsequently forming center crops of 64×64 pixels. Hence, this procedure was applied on images of 128×128 pixels. However, no significant changes in model accuracy could be observed compared to the case of no image enlargement.

As was to be expected, for linear global motion, there is a local maximum when target motion and global motion compensate each other, i.e. $s_{x,global,linear} = -s_{x,target,linear} = 20$ pixel. This maximum of accuracy close to 100% even exceeds the accuracy for a still target of same size without global motion, since the accuracy is about 75% for $s_{x,target,linear} = 0$ and $s_{x,global,linear} = 0$ in Figure 7. Remarkable is also the fact, that for $s_{x,target,linear} = -20$ pixel, the model accuracies are higher for $s_{x,global,linear} - s_{x,target,linear} > 0$ than for $s_{x,global,linear} - s_{x,target,linear} < 0$. While the target sharpness is similar for same $|s_{x,global,linear} - s_{x,target,linear}|$, global motion increasingly blurs the background. Therefore, the effective signal-to-background ratio $SNR_{background}$ is also increased resulting in higher model accuracies. For the other target motion shifts $s_{x,target,linear} = 0$ and $s_{x,target,linear} = 20$ pixel, the accuracies smoothly drop for increasing global motion shift $s_{x,global,linear}$. While the base model M_0 is better for low target motion shifts $s_{x,target,linear} \leq 5$ pixel, the models M_{single} and M_{all} including motion in the data augmentation for training are better for large target motion shifts $s_{x,target,linear} \geq 20$ pixel. For sinusoidal global motion there is a trend of dropping accuracies for increasing global motion shift $s_{x,global,sinusoidal}$ as well as for jitter global motion. For sinusoidal motion and jitter, the models M_{single} and M_{all} also tend to be better than M_0 for motion shifts $s_{x,global,sinusoidal/jitter} \geq 10$ pixel. Curiously, the model accuracies increase for target motion $s_{x,target,linear} = 20$ and low target motions $s_{x,target,sinusoidal} = 5$ pixel. This could be due to the fact, that the bimodal distribution caused by the sinusoidal motion and jitter provides multiple sharper images of triangles in the same direction compared to the blurred single triangle for the case of no global motion.

Also simulations with diagonal and vertical linear target motion combined with horizontal global motion and the three motion types (linear, sinusoidal and jitter) were applied. While the local maximum for $s_{r,target,linear} = -s_{x,global,linear}$ became smaller for the diagonal global motion, there are similar accuracies for both directions of target motion, i.e. $s_{x,target,linear} = \pm 20$ pixel, and vertical global motion.

3.5 Zooming and rotation

A realistic scenario where zooming and rotation of the imager's field of view may occur is a spinning missile heading for a linearly moving target. Therefore, global motion with different maximal zoom factors z_{max} and rotation angles α_{max} were applied, while linear horizontal target motion was applied with $s_{x,target} \in [-20, 0, 20]$. The center of the image is both the triangle center and center for zoom and rotation. The model accuracies of the models M_0, $M_{single,RZ}$ and $M_{all,RZ}$ are shown in Figure 8.
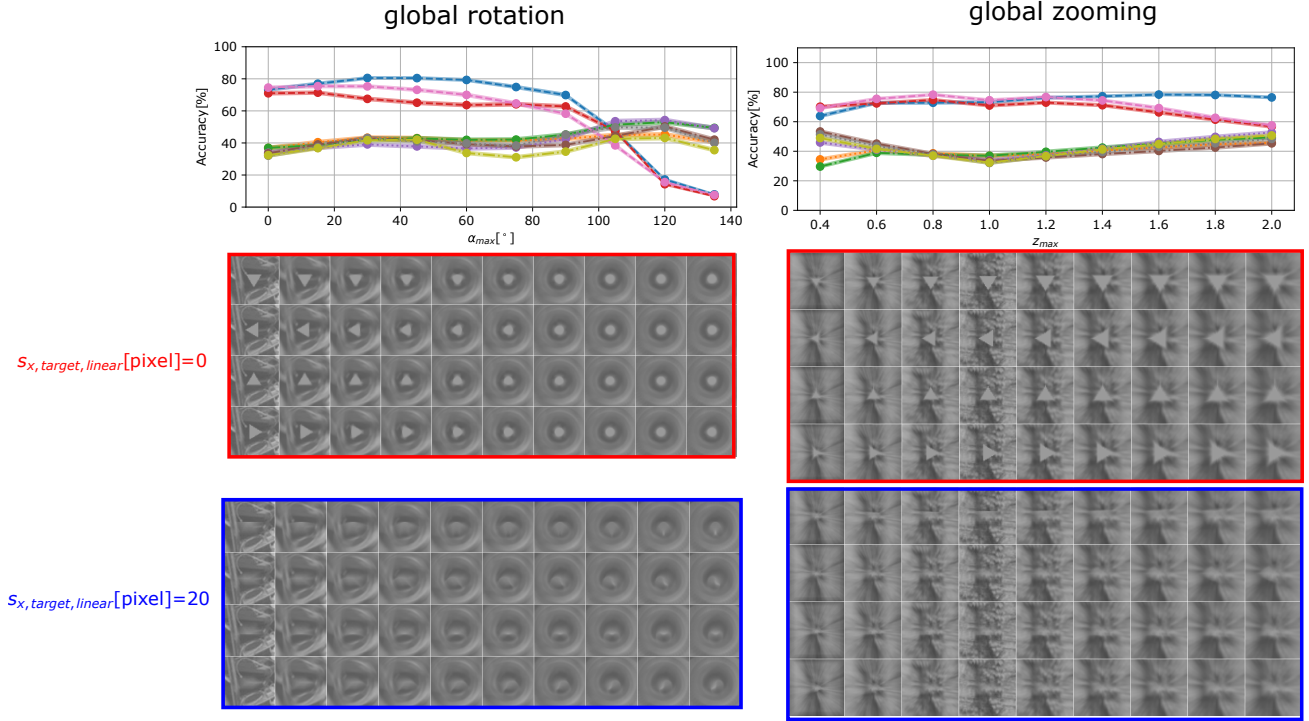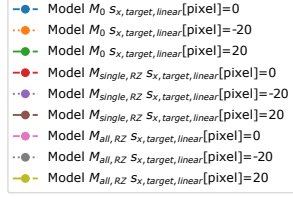
Figure 8. Model accuracy for models M_0, $M_{single,RZ}$ and $M_{all,RZ}$ based on $N = N_{orientation} \cdot N_{background} = 4000$ images with a centered triangle of circumradius $r = 10$ pixel and $N_{backgorund} = 1000$ random crops used as backgrounds. Different global rotation and global zooming around triangle center at different rotation angles α_{max} and zoom factors z_{max} are applied, as well as horizontal linear target motion at levels $s_{x,target,linear}[\text{pixel}] \in [-20, 0, 20]$.

For global rotation, model accuracies for all models are quite stable up to $\alpha_{max} \leq 90°$. For the still target, i.e. $s_{x,target,linear} = 0$, the accuracies decay for angles $\alpha_{max} > 90°$ even below the guess rate of 25%, while they slightly increase for a moving target, i.e. $s_{x,target,linear} = 20$ pixel. Obviously, global rotation leads to a blur of the triangle corners, while the triangle's inner circle is conserved. When inspecting the corresponding example images in Figure 8 triangle orientation discrimination seems difficult for the human observer for even the lowest rotations $\alpha_{max} = 15°$ in contrast to the models. The models $M_{single,RZ}$ and $M_{all,RZ}$ including global rotation do not reveal a benefit over the motion-agnostic model M_0.

Also, slight variations can be found for varying zoom factor z_{max}. For very low zoom factors $z_{max} = 0.4$ the models $M_{single,RZ}$ and $M_{all,RZ}$ are significantly better than M_0, while M_0 is better for large zoom factors z_{max}. Again, when inspecting the example images, for the moving target the models seem to reveal super-human performance. It is also remarkable, that the model accuracies are relatively similar over ranges of α_{max} and z_{max} above the ranges $\alpha_{max} \in [-15°, 15°]$ and $z_{max} \in [0.8, 1.2]$ used in the training data. This fact proves a good generalization ability of the model for these motion types.

4. CONCLUSION

In this paper the impact of motion blur on TOD classifier model performance was investigated. These models were trained and validated on artificial images of single triangles overlayed by natural images as backgrounds. Smoothly dropping model accuracies were found for increasing target and global motion shifts, decreasing triangle sizes and signal-to-noise ratios SNR_{noise}. Comparisons of models trained with motion blur (M_{single} and M_{all}) and without motion blur (M_0) showed, that M_{single} and M_{all} are slightly worse than M_0 if the motion shifts are low ($s_{x,target/global} \leq 15$), while they are better than M_0 for strong motion shifts ($s_{x,target/global} \geq 15$) due to more or less relevant training data in the respective domain. Therefore, the benefit of including motion blur in the model training depends on expectable motion shifts in real case scenarios.

Furthermore, global motion can increase effective signal-to-background ratio $SNR_{background,eff}$ leading to higher model accuracies. Hence, initial acquisition of a moving target and compensation of target and global motion can have great impact on the performance of tracking applications. The model evaluations for joint target and global motion show that image blur on target and background can have opposing effects on target recognition. This fact can render a vast amount of image quality metrics (IQM) unsuitable for task-specific performance assessment, as many of them are not specifically sensitive to predefined targets.

The framework of target and global motion and the procedure for model training and validation presented here can be easily extended to real targets. Therefore, it could be a useful building block for the assessment of unmanned aerial vehicle (UAV) applications as well as for the evaluation of their countermeasures.

REFERENCES

[1] Navarro, F., Serón, F. J., and Gutierrez, D., "Motion Blur Rendering: State of the Art," *Computer Graphics Forum* **30**(1), 3–26 (2011).

[2] Sharan, L., Neo, Z. H., Mitchell, K., and Hodgins, J. K., "Simulated motion blur does not improve player experience in racing game," in [*Proceedings of Motion on Games*], *MIG '13*, 149–154, Association for Computing Machinery, New York, NY, USA (Nov. 2013).

[3] Ma, B., Huang, L., Shen, J., Shao, L., Yang, M.-H., and Porikli, F., "Visual Tracking Under Motion Blur," *IEEE Transactions on Image Processing* **25**, 5867–5876 (Dec. 2016).

[4] Nihei, R., Tanaka, Y., Iizuka, H., and Matsumiya, T., "Simple correction model for blurred images of uncooled bolometer type infrared cameras," in [*Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXX*], Holst, G. C. and Krapels, K. A., eds., **11001**, 1100118, International Society for Optics and Photonics, SPIE (2019).

[5] Oswald-Tranta, B., Sorger, M., and O'Leary, P., "Motion deblurring of infrared images from a microbolometer camera," *Infrared Physics & Technology* **53**, 274–279 (July 2010).

[6] Bijl, P. and Valeton, J. M., "TOD: a new method to characterize electro-optical system performance," *Proc. SPIE* **3377**, 182–193 (1998).

[7] Wegner, D. and Keßler, S., "Comparison of algorithms for contrast enhancement based on TOD assessments by convolutional neural networks," in [*Electro-optical and Infrared Systems: Technology and Applications XIX*], Hickman, D. L. and Bürsing, H., eds., **12271**, 122710H, International Society for Optics and Photonics, SPIE (2022).

[8] Wegner, D. and Keßler, S., "Comparison of algorithms for contrast enhancement based on triangle orientation discrimination assessments by convolutional neural networks," *Optical Engineering* **62**(4), 048103 (2023).

[9] Kingma, D. P. and Ba, J., "Adam: A method for stochastic optimization," *CoRR* **abs/1412.6980** (2014).

[10] Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M., Kolesnikov, A., Duerig, T., and Ferrari, V., "The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale," *IJCV* (2020).

Error Analysis of a Multi-Sensor Maritime Targeting System

Duncan L. Hickman[a], Maria Niebla[a], Daekyu Sang[b], Boomin Kim[b], and Jeonghyun Baek[b]

[a] Tektonex Ltd, The Long View, Argyll Rd, Kilcreggan, Argyll & Bute, UK
[b] Agency for Defense Development, 160 Bugyuseong-dero 488beon-gil, Yuseong-gu, Daejeon, Korea

ABSTRACT

Targeting systems are subject to multiple sources of error when operating in complex environments. To reduce the effect of these errors, modern targeting systems generally include both imaging and RF sensors. Data processing then provides target detection and classification information, and the detection streams are combined using a data fusion scheme to produce an optimal target location estimate with an associated latency. In this paper, the performance of a multi-sensor system in a maritime application is investigated using a mathematical simulator that has been developed to provide the system performance error analysis for different engagement scenarios and test conditions. This simulator is described together with the sources of targeting error such as image motion blur and radar glint. Additionally, the impact of flare and chaff countermeasures on the targeting performance is reviewed in terms of different types of target recognition and tracking algorithms.

Keywords: Multisensor Systems, Error Analysis, Image Blur, Radar Glint, Model Validation, Countermeasure Model, ATR Processing, Targeting Systems.

1. INTRODUCTION

The combination of sensors with different modalities such as infrared and radar, is an effective means of optimising the range and angular resolution performance of a targeting system. In this paper, we consider several critical issues and challenges in the modelling and simulation of such a multisensor system for a maritime scenario.

For this maritime application, the modelled sensor suite consists of an active RF sensor (2cm wavelength) and two cameras which operate in the thermal and visible spectral bands. The detection outputs are then combined through a data fusion process to give an estimated target position which is then used to control the dynamics of the platform on which the sensors are mounted. For the imaging camera sub-system, it is assumed that these are used in a strapdown configuration [1-3] and gimbal related errors can be ignored. It is also assumed that the two camera image streams are fully aligned and spatially registered, and that their outputs are time synchronised. It is further assumed that the target can manoeuvre and is equipped with passive countermeasures (chaff and flares).

With the above assumptions, the performance errors and the sensitivity to sensor design, image blur effects [4], radar glint [5,6], countermeasures [7,8], and the operational environment can be readily investigated within an end-to-end engagement simulation. Such a simulator must have a fast runtime, particularly if the multiple engagements are performed within a Monte-Carlo Simulation (MCS) framework. Consequently, those functions relating to the sensor assembly must be based on mathematical representations of the processing performance and behaviour such as transfer functions or Look-Up Tables (LUTs) [9,10]. The processing time requirement creates a number of challenges, and the following issues are addressed in this paper:

- Developing an efficient functional design for the simulator;
- Incorporation of sensor platform and target manoeuvres;
- The representation of imaging and RF sensor performance as LUTs and transfer functions;
- Determination of the errors associated with radar glint from a complex target;
- The effect of image motion blur on the ATR and guidance performance;
- System latency effects on the targeting performance;
- Incorporation of countermeasure representations and the assessment of their effectiveness;
- The effect of environmental conditions.

In Section 2, we discuss the modelling approach taken and describe the functional design used. The simulator design uses two different types of models. The first is a time-based mathematical model that simulates each engagement using LUT or transfer functions of the sensors. This is referred to as the *Error Model*. The second type processes images or data to determine the performance and behaviour of the sensors for a range of input conditions. These models can use image or data processing methods and are referred to here as *Procedural Models*. The Procedural Models generate the required sensor performance information which is captured in an *Error Results File* (ERF). The Error Model then reads the ERF and interpolates the sensor data for the required conditions. In Section 2, we also review the characteristics of the sensor platform motion (guidance) and the target and describe how these are integrated into a time-based engagement simulation.

The imaging sensor Procedural Model is discussed in Section 3 together with the generation of the associated performance LUT and its interpolation. To illustrate the model an example is presented where the variation of an ATR algorithm is determined against a specific target as a function of range, image blur, and noise. Experiments have shown that further simplifications can be made without introducing any significant errors and these simplifications are discussed. The RF sensor Procedural Model is then reviewed in Section 4 and the approach to radar glint modelling is discussed. A method for the mathematical representation of glint is proposed as a new and effective means of incorporating glint in an engagement simulation. In Section 5, the approach taken for modelling passive countermeasures is presented together with the methodologies used for assessing the impact of flares and chaff on the system performance.

Combining sensor, target, platform, countermeasure, and environmental functions provides a comprehensive performance analysis tool for design and sensitivity analysis. However, care needs to be exercised when assessing the simulator results: do they accurately represent the required engagement or are they a consequence of model errors or limitations. Unfortunately, increasing the complexity of a model increases the risk of:
- Implementation errors (coding and parameters);
- Operation beyond the assumptions or limitations associated with modelled functions;
- Coupling between two or more functions creates conflicts or compounded errors.

Given the research and development nature of many simulators, there is generally no measurement data available to confirm the accuracy of the model and provide validation accreditation. In Section 6, a number of methods are described which were used to verify and validate the simulators behaviour. These techniques include using methods such as sensitivity analysis and cross-correlation. Additionally, some errors only become evident under certain conditions and parameter combinations, and it is shown that these can be revealed using MCS methods.

2. FUNCTIONAL ARCHITECTURE

2.1. Design Approach

Engagement models are typically implemented as a time-based simulation with sensors being represented as simple transfer functions. However, to fully evaluate the performance of modern sensors, greater detail is required, particularly with regards to the sensor data processing.

For imaging sensors, the most obvious approach is to use image-based processing to determine the required performance measures. This would require either real or synthetically generated imagery that reflect the viewing conditions with the engagement model. In addition to the obvious complexity, such an approach would be extremely time-consuming for MCS implementations. The approach taken here uses algorithmic processing on a small sample of images, where the imagery covers the conditions expected in the simulation (such as range, blur, and noise). This data is then saved in a LUT where it is subsequently interpolated for the conditions in the engagement model [9,10].

Radar glint can be determined using principles of physical optics (PO) and the physical theory of diffraction (PTD) in conjunction with ray tracing techniques [11]. However, such simulations are time consuming and require the availability of expensive software tools. An alternative, and widely used approach, is to represent the target as a number of point scatterers and then calculate the scattered return at the sensor's location [12,13]. Again, this can result in a complex and computationally time-consuming process not suited for use within an engagement model or large numbers of engagement simulations. An alternative approach is used here whereby a target simulation is used to generate a statistical distribution for the target glint, to which a function is then fitted.

As noted earlier, the simulator is separated into two components: the Error Model and the Procedural Models (Figure 1). The Error Model is a time-based representation of the engagement where the behaviour and performance of the sensors are represented as transfer functions. This ensures that the engagement simulation run-time is sufficiently fast to support large MCS. Two Procedural Models are used. The first is for the imaging sensors where image processing is applied to a small sample of images which are manipulated to produce an error map as a function of range and degradations. The second is for the RF sensor where a scattering model is used to determine the glint characteristics for a given target. The data generated by the Procedural Models is pre-calculated and saved in an Engagement Run File (ERF) in the form of Look-Up Tables (LUTs) or equation parameters. The Error Model subsequently interpolates the LUTs to get specific sensor errors for the conditions in the engagement loop.
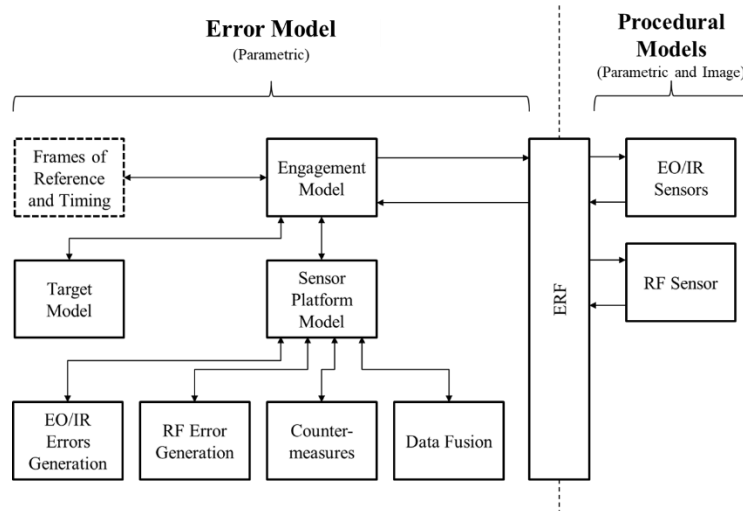

Figure 1: Functional architecture of the sensor system to be evaluated.

The overall processing time is greatly reduced using this functional design. For example, when operated on a standard desktop PC, an engagement comprising 2000 steps which is repeated 1000 times in a MCS, requires approximately 435 seconds of run time.

2.2. Engagement Framework

The engagement framework of the Error Model uses a conventional reference frame approach with the x, y axes being in the plain of the sea surface and the z being positive in the upward direction from the surface. Local frames are used for the sensor platform and the target, and these are connected to the reference frame using translation and rotation vectors. Given the fixed and coincident viewing conditions of the sensors, they are assumed to be aligned with the body-axis of the platform and no further frames of reference (such as for gimballed systems) are required. The target is assumed to remain in the x, y plane.

2.3. Target Model Representation

Maritime targets are slow moving when compared to the sensor platform. As such, ships are generally represented as an object moving along a linear path at constant speed although some authors have modelled a circular trajectory path [14]. It is well known that naval ships use countermeasures in conjunction with a manoeuvre. In recognition of this, the target model used within the simulator combines an initial linear path followed by a rotational path to form an evasive trajectory. The manoeuvre can be either a turn to port or starboard and the time at which the turn is executed can be varied. Additionally, the rate of turn, can be set or varied as required. Given the relatively short engagement times for maritime engagements, the linear speed of the ship can be assumed constant.

The basic geometry of the target trajectory model is shown in Figure 2 and the position vector of the target (T) in the reference frame at a given time t is:

$$\boldsymbol{T}(t) = \begin{bmatrix} S_0 \cos \alpha_0 + v_T\, t \cos \beta + r \cos \beta \sin \varphi(t - t_A) + rL \sin \beta \{1 - \cos \varphi\,(t - t_A)\} \\ S_0 \sin \alpha_0 + v_T\, t \sin \beta + r \sin \beta \sin \varphi(t - t_A) - rL \cos \beta \{1 - \cos \varphi\,(t - t_A)\} \\ 0 \end{bmatrix} \qquad (1)$$

where S_0 is the initial surface range from the reference frame origin, α_0 is the azimuth angle (relative to the x-axis), and t_A denotes the time at which the turn manoeuvre begins. Note that if $t < t_A$ then we set $t - t_A = 0$. The rotation angle $\varphi(t)$ is given by $\varphi(t) = v_T(t - t_A)/r$. L is a sign function with a value of +1 for a starboard turn and -1 for a turn to port. The direction (unit) vector of the target is given by:

$$\widehat{\boldsymbol{D}}_T(\boldsymbol{t}) = [\cos\{\beta - L\varphi(t - t_A)\}, \sin\{\beta - L\varphi(t - t_A)\}, 0] \qquad (2)$$

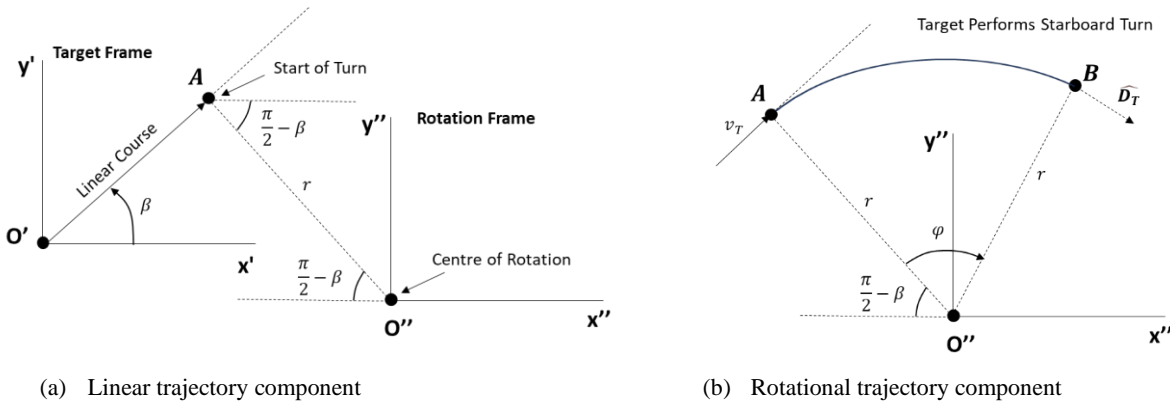(a) Linear trajectory component   (b) Rotational trajectory component

Figure 2: Two component target trajectory where A denotes the start of the target manoeuvre and B is the target's position during the turn. The angle β defines the target's initial course and the radius of the turn, r, is calculated from the target's speed, v_T and angular rate $\dot{\varphi}$.

2.4. Sensor Platform Model

The sensor platform model uses a proportional navigation (PN) guidance method [15-17]. There are two general types of PN: *pure* and *true* and these are illustrated in Figure 3 [18]. The difference between them is the direction of the acceleration. For pure PN, the acceleration is normal to the velocity vector and therefore only its angle of travel changes. For true PN, there is a component of acceleration in the direction of travel. The platform speed in the direction of travel will therefore change under true PN, which implies motor thrust or breaking is required. For simplicity, we assume that the sensor platform is constant and therefore will adopt pure PN for guidance purposes.
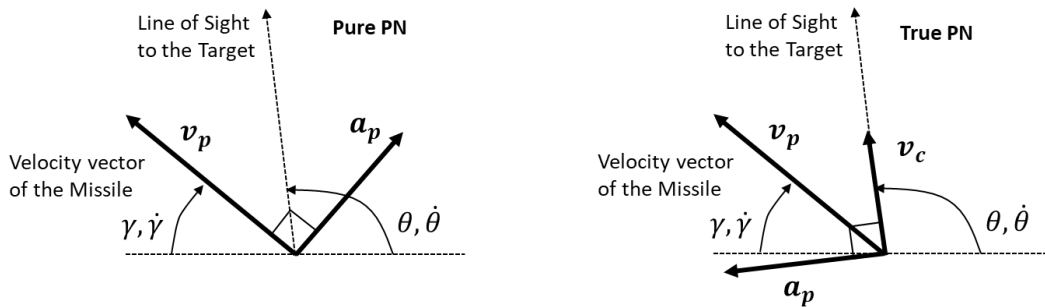

Figure 3: Pure and True Proportional Navigation

Using pure PN for platform guidance, we have the lateral acceleration as [19,20]:

$$a_p = N v_p \dot{\lambda}$$

where $\dot{\lambda}$ is the rate of change of the line-of-sight angle between the platform and the target.

Assuming that the platform is roll-stabilised, we can treat the pitch and yaw angles which define the sensors' lines-of-sight as being independent of each other. Denoting the pitch angle as θ and the yaw angle as ψ, we can write the lateral accelerations as [19,20]:

$$
\begin{aligned}
a_\theta &= N v_p \dot{\theta} \\
a_\psi &= N v_p \dot{\psi}
\end{aligned}
\tag{3}
$$

where N denotes the navigation constant and v_p is the speed of the platform.

Note that the gravitational force in the pitch direction has been ignored in Equation 3 as it does not impact the error analysis [19,20]. The angular rates can be determined from the time-based simulation using:

$$
\begin{aligned}
\dot{\theta}_E(t) &= \frac{\theta_E(t) - \theta_E(t - \Delta t)}{\Delta t} \\
\dot{\psi}_E(t) &= \frac{\psi_E(t) - \psi_E(t - \Delta t)}{\Delta t}
\end{aligned}
\tag{4}
$$

where Δt denotes the time interval between measurements and the subscript E denotes the estimated value of an angle. The sensor measurement errors are incorporated in these estimated values.

We use these estimated angular rates to determine the commanded angles for the guidance system. Assuming small angles and time steps, we have:

$$
\delta\theta(t) = -N\Delta t \cdot \dot{\theta}_E(t)
$$

$$
\delta\psi(t) = N\Delta t \cdot \dot{\psi}_E(t)
\tag{5}
$$

and the direction vector of the platform (and sensors' line of sight) is given by:

$$
\widehat{\boldsymbol{D}}_{\boldsymbol{P}}(t) = \begin{bmatrix} \cos\{\theta_E(t - \Delta t) + \delta\theta(t)\} \cos\{\psi_E(t - \Delta t) + \delta\psi(t)\} \\ \cos\{\theta_E(t - \Delta t) + \delta\theta(t)\} \sin\{\psi_E(t - \Delta t) + \delta\psi(t)\} \\ \sin\{\theta_E(t - \Delta t) + \delta\theta(t)\} \end{bmatrix}
\tag{6}
$$

The platform's position vector in the reference frame at time t is given by:

$$
\boldsymbol{P}(t) = \boldsymbol{P}(t - \Delta t) + v_P \Delta t \cdot \widehat{\boldsymbol{D}}_{\boldsymbol{P}}(t - \Delta t)
\tag{7}
$$

where v_P is the speed of the sensor platform.

The *Projected Miss-Distance* (PMD) was found to be a useful performance measure. It provides a prediction of the separation of the target and the platform at the end of the engagement by projecting the current target and platform directional unit vectors for the remaining time of the engagement. It is given by:

$$
PMD(t) = \left\| \{\boldsymbol{T}(t) - \boldsymbol{P}(t)\} + t_F(t) \cdot \{v_T \cdot \widehat{\boldsymbol{D}}_{\boldsymbol{T}}(t) - v_P \cdot \widehat{\boldsymbol{D}}_{\boldsymbol{P}}(t)\} \right\|
\tag{8}
$$

where t_F is the remaining time of fight at time t and is given by:

$$t_F(t) = \frac{\boldsymbol{P}_z(t)}{v_P \widehat{\boldsymbol{D}}_{\boldsymbol{P},\boldsymbol{z}}(t)} \tag{9}$$

and the z components of the position and direction vectors are denoted by the subscript. An example engagement and PMD plot are shown in Figure 4.
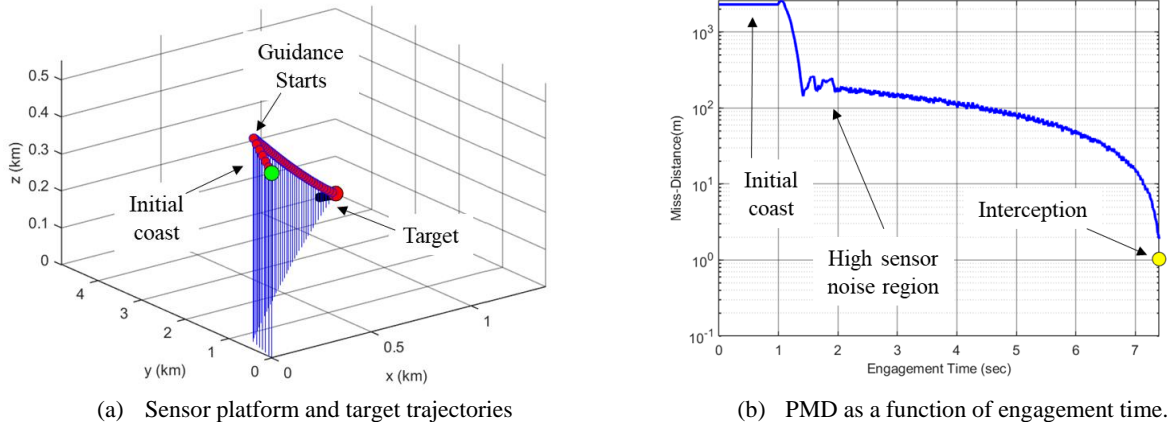

(a)   Sensor platform and target trajectories

(b)   PMD as a function of engagement time.

Figure 4: Example engagement trajectory plot and PMD. The platform initially coasts before the guidance command is invoked. In this case, the initial guidance was provided by the RF sensor which produced a larger angular error.

3. IMAGING SENSOR ERRORS

3.1. Methodology

The imaging sensor Procedural Model applies different ATR processing schemes to imagery in order to determine positional errors due to blur, noise, structured backgrounds, and atmospheric losses. A small number of single image frames of the target are used where imagery matches the geometry for the engagement. The images are of the target only and any background is removed. From this small sample, a subset of images is generated through a range scaling function (Figure 5). The ranges used cover that needed for the engagement.
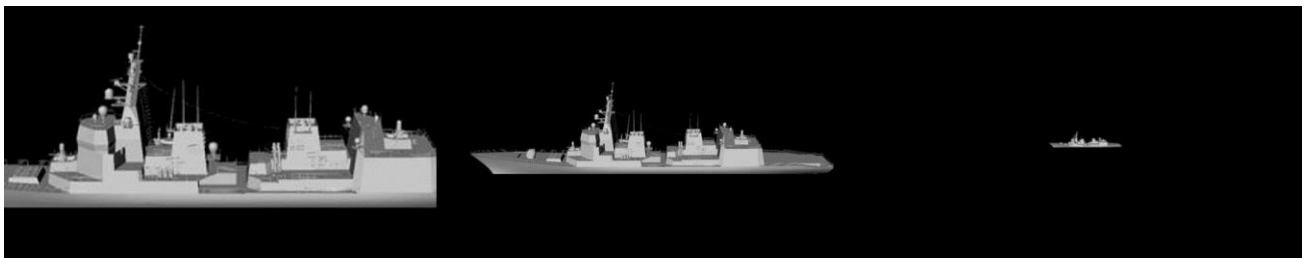

Figure 5: Scaled target image with no background

A statistically derived background is then added to the target image using a suitable power spectral density function scaled for range [21]. At each range step, different types of image blur are applied to the images where the blur has varying amplitudes. Consequently, each image provides an array of M ranges by N blur levels.

The applied blur consists of two functions. The first is a Gaussian blur with a fixed variance which represents the image degradations of the imaging system as well as any high frequency random motion blur (jitter) associated with the platform motion. This is a reasonable assumption since a series of compounded and monotonically decreasing MTF curves tend towards a Gaussian distribution and, consequently, the Point Spread Function (PSF) also tends to a Gaussian profile. The

second type of blur is that associated with linear motion which occurs when the platform orientation changes rapidly in pitch or yaw. Linear motion blur is a result of the image moving across the image-plane during the detector integration time. Such motion is characteristic of sensors mounted on an agile platform and it has a variable amplitude which depends on the time-dependent lateral accelerations.

Once the image blur has been added, noise is introduced. Two types of noise are used: random Gaussian and Salt and Pepper [22]. The results of combining the scaled target with the statistical background, followed by blurring and noise are illustrated in Figure 6 where the base image shown in Figure 5 was used.
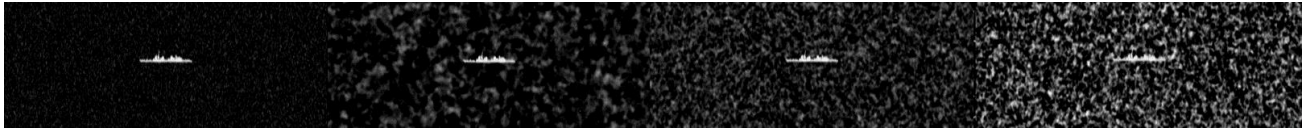
Figure 6: Examples of the longer-range target image from Figure 5 after combination with different background structures, and after the application of image blur and noise. The sequence corresponds to a progressive degradation in image quality (left to right) which will increase the errors associated with ATR processing and aimpoint selection.
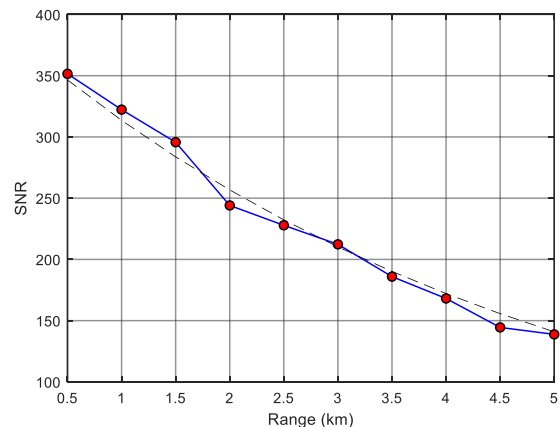
3.2. Imaging Sensor LUT and Interpolation

Starting with the matrix of images at different ranges and with different blur settings, ATR processing is applied to the imagery to determine the target aimpoint. For those ATR algorithms that employ time-based processing, the matrix of single images can be replaced with a matrix of short image sequences.

The aimpoint error is calculated with reference to the position estimates obtained from target-only (non-degraded) images. In Figure 7a, the aimpoint error is plotted as a function of range and image blur and this forms the imaging sensor LUT. We also show the Signal to Noise Ratio (SNR) in Figure 7b. Note that the measurement of the SNR includes the effects of the clutter structure.
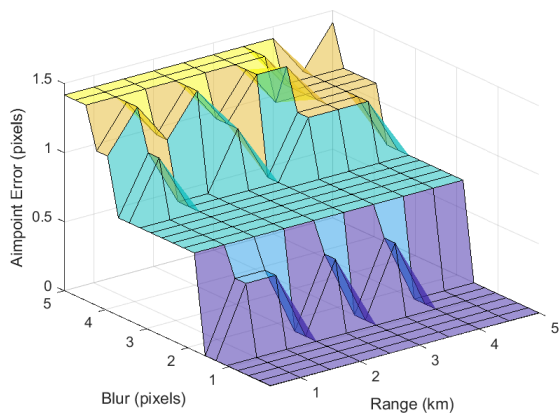
| (a) Sample of an aimpoint error surface | (b) Measured SNR compared to predicted values |

Figure 7: Performance data from the imaging sensors procedural model

It was found in all engagement cases, that the SNR was greater than 10, and greater than 100 in more than 95% of the cases. This was a consequence the high thermal contrast of the target and the relatively short engagement ranges. The imaging sensor performance was therefore primarily limited by blur effects.

3.3. Simplifications to the LUT

The original concept for the imaging sensor LUT was based on the generation of a $M \times N$ matrix of aimpoint errors that would be subsequently interpolated using the range and blur values at each engagement step. However, it was found that

for the implemented ATR algorithms that exploited shape information, the aimpoint error could be represented as a function of the linear blur at each range. This can be seen from the surface profile shown in Figure 7a. Consequently, the LUT could be readily replaced by a table of offset values and gradients (Figure 8).
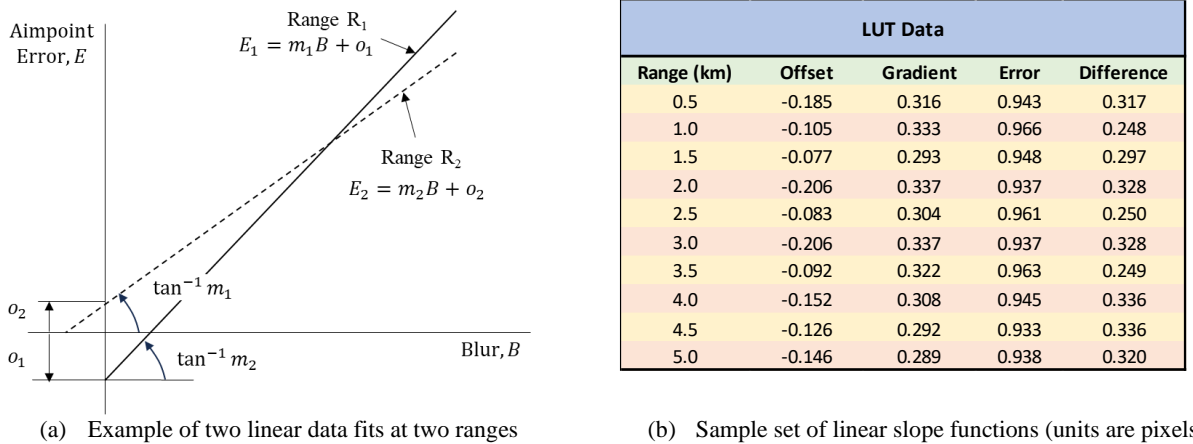

LUT Data				
Range (km)	Offset	Gradient	Error	Difference
0.5	-0.185	0.316	0.943	0.317
1.0	-0.105	0.333	0.966	0.248
1.5	-0.077	0.293	0.948	0.297
2.0	-0.206	0.337	0.937	0.328
2.5	-0.083	0.304	0.961	0.250
3.0	-0.206	0.337	0.937	0.328
3.5	-0.092	0.322	0.963	0.249
4.0	-0.152	0.308	0.945	0.336
4.5	-0.126	0.292	0.933	0.336
5.0	-0.146	0.289	0.938	0.320

(a) Example of two linear data fits at two ranges   (b) Sample set of linear slope functions (units are pixels)

Figure 8: Representation of the LUT surface using a series of linear slope function.

Therefore, the LUT size would be reduced from an $M \times N$ matrix to a $2M$ matrix. However, and more importantly, the interpolation process is simplified and requires fewer calculations than would be needed for a four-point interpolation of the original $M \times N$ matrix of aimpoint errors.

The data table in Figure 8b was found to be typical for linear motion blur and ATR algorithms that exploit shape information and contrast. It can be seen that the gradients were fairly constant with range, for the case shown in Figure 8b, the mean gradient is 0.31 ± 0.02. Therefore, as a further simplification, the table of slope functions can be replaced with a single gradient value with minimal error. Additionally, the small bias on the offset can be ignored and set to zero so that the ATR aimpoint errors is given by the simple relationship:

$$E(\mu_m, B) = \mu_m \cdot B \tag{10}$$

where μ_m is the mean gradient determined from the imaging sensor Procedural Model and B is the blur value set by the engagement model at each step. Equation 10 offers a fast and effective means of incorporating imaging sensor aimpoint errors in an engagement model with minimal error.

4. RADAR GLINT

4.1. Glint and Targeting Issues

A major limitation for maritime detection and tracking systems, particularly at short ranges, is that of target glint. The basis of this is the coherent interference of the returned radar radiation from a complex scattering target [5]. The target can be viewed as a complex set of scattering surfaces and points which reflect the incident radar signature, and these reflected waves combine coherently, resulting in wavefront variations and orientations. The orientation of the wavefront at the radar is used to determine the location of the target, therefore any variations in this wavefront will result in a target location error [12,13]. The resultant radar glint angle error increases as the radar range decreases and is therefore critical to the targeting performance in the terminal phase of an engagement.

Maritime targets are large and complex objects, and they are generally modelled using either statistical or deterministic means [13]. The approach taken here is that of the more widely used method whereby the target is considered as a small number of scattering points [12]. These give a representative indication of glint errors and variations through an engagement although the detailed structure and variation is not necessarily accurate for a given target type.

4.2. Glint Model

The approach taken is to assume that the target comprises a set of scattering centres, s, each with a reflectivity component and a phase retardation that is applied to the reflected RF wave. For the case of N scattering points, we have the returned wave at the radar receiver as:

$$E_T \cdot e^{-j\Phi_T} = \sum_{k=1}^{N} \sigma_k \cdot e^{j\varphi_k} \tag{11}$$

where σ_k is the scattering or reflectivity coefficient of the kth scatterer and φ_k is the associated phase of the received wave from the scatterer. Note that φ_k includes the intrinsic phase retardation from the scatterer as well as the phase associated with the wave propagation. E_T is the combined amplitude of the coherent return waves and Φ_T is the associated phase. When the composite wavefront is generated through multiple scatterers, it can become tilted relative to the normal line of sight as illustrated in Figure 9.
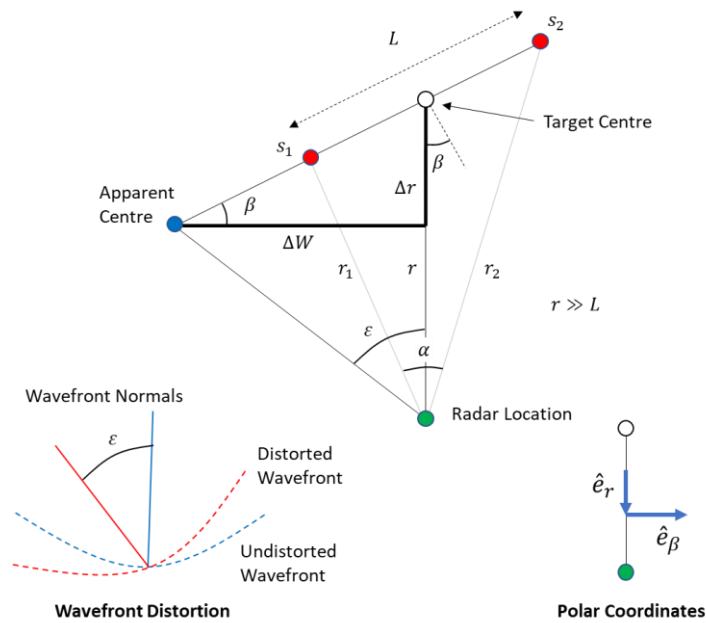
Figure 9: Target glint effects for a simple two scatterer target.

An RF receiver derives the target's direction from the wavefront phase [23] and any asymmetry will give rise to an angular error such that the target's apparent centre becomes offset from the true target centre. This is referred to as *radar glint*.

To evaluate the wavefront at the radar, we use polar coordinates to determine the gradient of the wavefront which, referring to Figure 9, is given by:

$$\nabla\Phi_T(r,\beta) = \frac{\partial\Phi_T}{\partial r}\hat{e}_r + \frac{1}{r}\frac{\partial\Phi_T}{\partial\beta}\hat{e}_\beta \tag{12}$$

where r is the range to target and β is the angle of the target to the sensor system's line of sight. Equation 12 can be solved to give the glint angle ε and the glint displacement term ΔW (refer to Figure 9). For the case of a simple two scatterer model, we have [13]:

$$\varepsilon = \frac{L \cos \beta}{2r} \cdot \frac{(\sigma_1^2 - \sigma_2^2)}{\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2 \cos\left(2\Delta\Phi + \frac{4\pi}{\lambda}L \sin\beta\right)} \tag{13}$$

$$\Delta W = \frac{L \cos \beta}{2} \cdot \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2 \cos\left(2\Delta\Phi + \frac{4\pi}{\lambda}L \sin\beta\right)} \tag{14}$$

Most reported models of radar glint are based on the calculation of the glint angle ε. However, the disadvantage of this is the inverse range dependency (as indicated in Equation 13). In contrast, the glint displacement term ΔW does not have the same range dependency. Using ΔW has the potential for simplifying the implementation of glint calculations and was therefore preferred for the RF Procedural Model. In Figure 10, the glint angle and displacement terms are shown as a function of range and target orientation where the target has been represented as ten scattering centres with independent reflectivity and intrinsic phase.

(a) Glint angle ε (b) Glint displacement ΔW
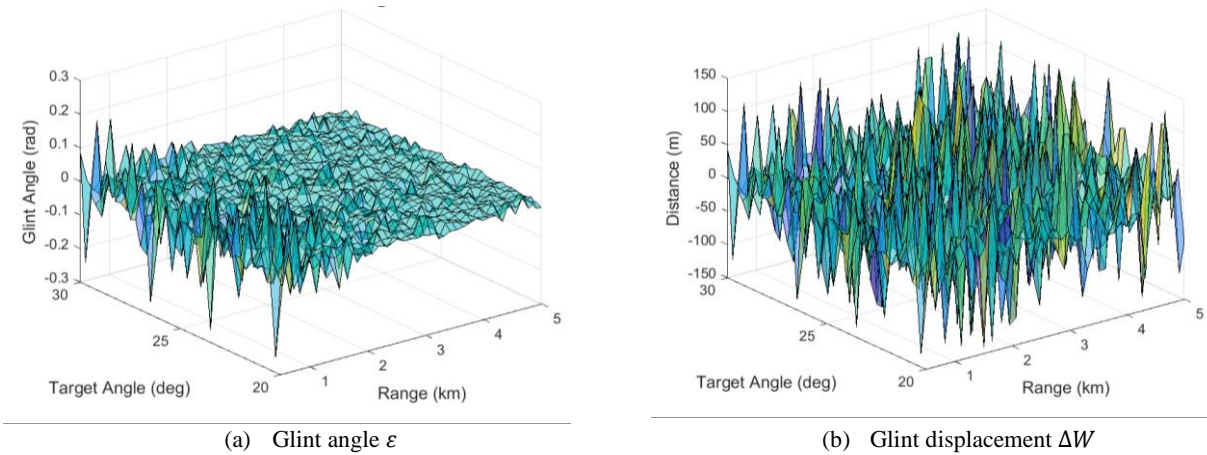
Figure 10: Glint angle and glint displacement as a function of range and target orientation for a 10-point scattering target model. These surfaces could be used as a LUT for the RF glint Procedural Model.

The glint angle data could be calculated and captured in a LUT. However, this would again require the interpolation of a data matrix where the adjacent matrix elements have a nominally random variation. An alternative approach was therefore used which is described below.

A number of points can be noted from these distributions shown in Figure 10. Firstly, the glint model produces a randomized distribution of glint errors for relatively small variations in the range and target orientation. This effect has been observed through measurement [24]. Secondly, the variation of glint angle with range can be seen in Figure 10a but a range dependency is not evident in the glint displacement of Figure 10b. These relationships appear to hold for simple target model (Equations 13 and 14) as well as the more complex target model used to generate the data in Figure 10. This is further illustrated in Figure 11 where the Procedural Model was used to generate 20 independent sets of glint data and the RMS values calculated.

A third point to be noted is that the envelope of the variations is relatively constant over a small angular range (10° in Figure 10) and this has two implications. Firstly, for maritime engagements, the target angular variation is generally small and consequently the glint calculations can be treated as independent of target angle. Secondly, the glint displacement statistics generated over all angles and ranges can be combined as they have the same underlying statistical distribution and associated parameters. This feature is used in the calculation of the glint function.

(a) Glint angle (the dashed line is a fitted $1/r$ function)
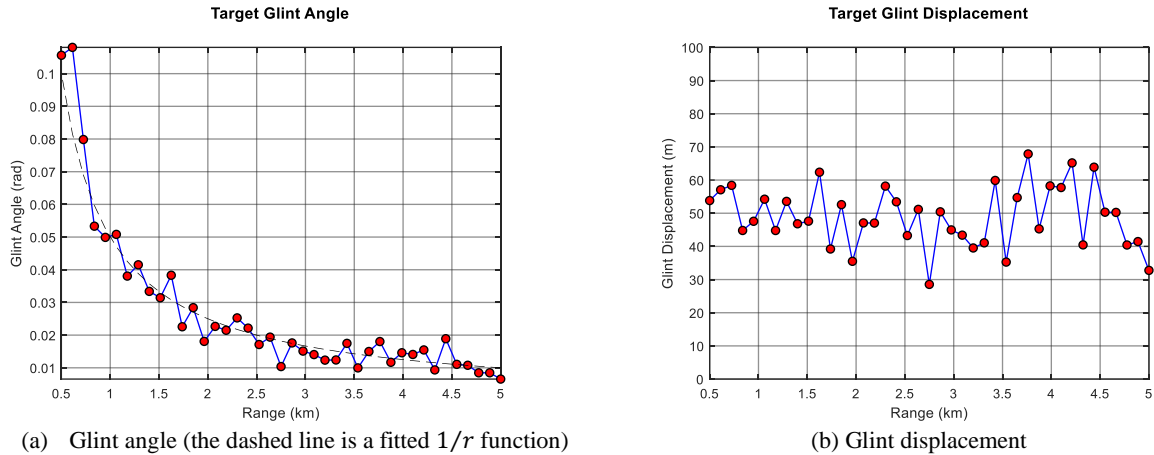
(b) Glint displacement

Figure 11: RMS target glint data to illustrate the underlying trends on the data surfaces shown in Figure 10. These plots were calculated from 20 independent sets of glint angle and displacement data.

4.3. New Approach and the Glint Gain

Using the ΔW function, we use all range and angle values from the glint data to generate a more accurate histogram of glint displacement probability. This is shown in Figure 12a, where the following exponential function has been fitted to the data:

$$H = H_0 \cdot e^{-\lambda_g |\Delta W|} \tag{15}$$

where λ_g is a gain term which we refer to at the *glint gain*. An alternative Gaussian form has also been used for the fitting function although the errors were found to be slightly larger for the targets modelled with the current simulator set-up parameters. Using the calculated glint gain value, we can generate the glint displacement using:

$$\Delta W = \frac{S}{\lambda_g} \cdot \ln(u + \delta) \tag{16}$$

where u is a random number between 0 and 1, δ is a small number to avoid a $\ln(0)$ condition, and S provides a random sign to give left and right displacements from the target centre. Equation 16 is plotted in Figure 12b.

Using the above representation, we therefore use the RF Sensor Procedural Model to calculate the glint gain term λ_g for a specific target and orientation. This single term replaces the need for a glint LUT of the form described earlier. The Error Model then calculates the glint displacement using this gain term together with a random number generator.

In Figure 13a, an example of the glint displacement is shown which has been generated using Equation 15. In addition, Figure 13b shows the variation of the calculated glint gain when calculated over multiple iterations of the Procedural Model. The modelled glint displacement and glint angle calculated during an engagement with a maneuvering target are shown in Figure 13c and 13d.
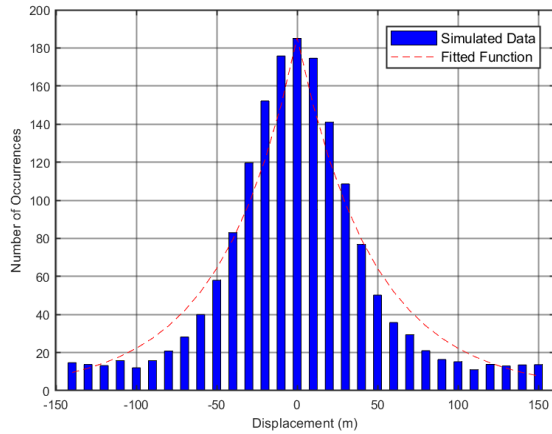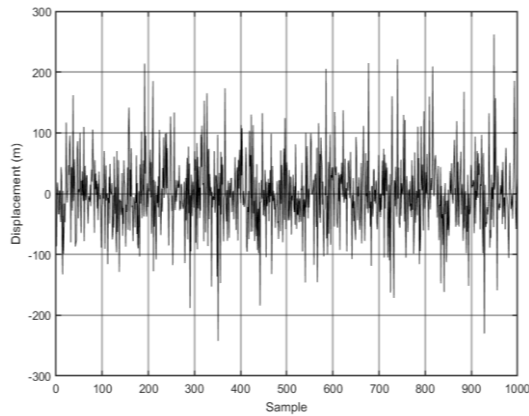
(a) Glint displacement generated using the data in Figure 10b. The function given by Equation 15 is also shown.
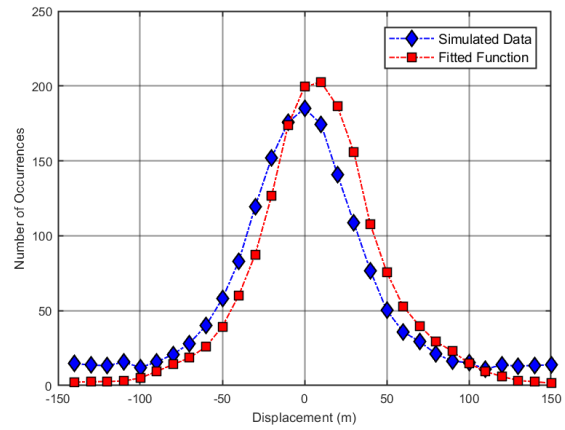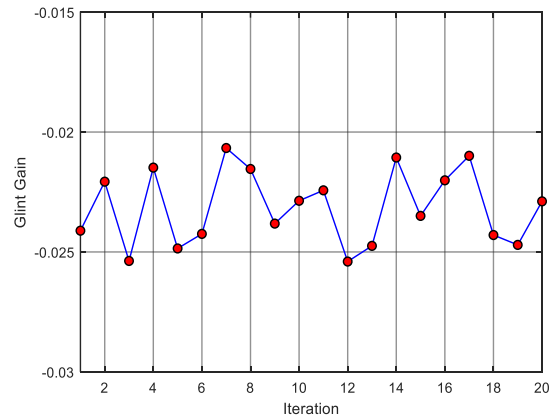
(b) Equation 15 (simulated data) is plotted against displacement data generated using Equation 16
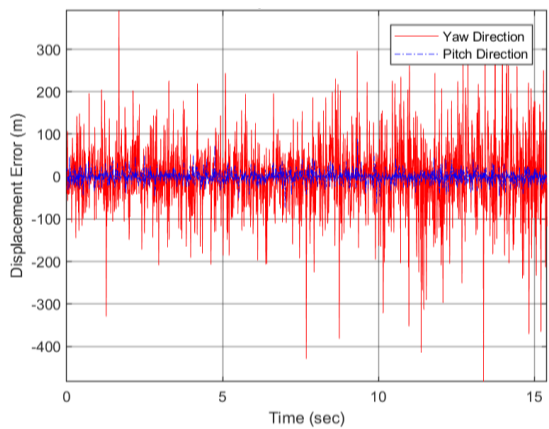
Figure 12: Glint displacement data generated by the Procedural Model and the statistical representation used by the Error Model
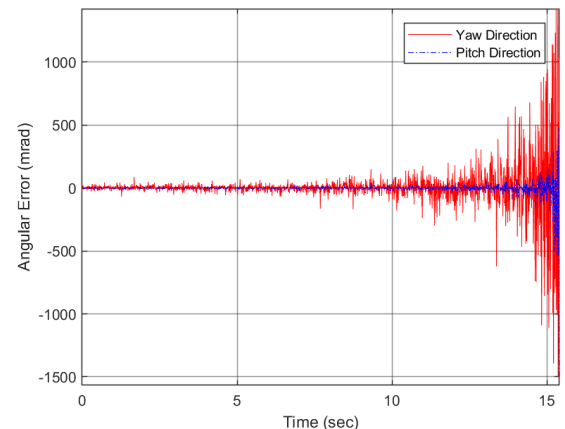

(a) Sample Glint Displacement as a range sequence

(b) Variation of the Glint Gain using the Procedural Model

(c) Glint displacement for an engagement (pitch and yaw directions). Note that the target is maneuvering which results in a small increase in the displacement envelope.

(d) Glint angle derived from (c). Note that this rapidly increases as the engagement time increases (range decreases).

Figure 13: Glint Displacement results and stability of the Glint Gain

5. COUNTERMEASURE MODELLING

5.1. Chaff and Flares

Chaff is a conventional method of defending a naval vessel whereby a false radar signal is generated which creates targeting errors. Chaff traditionally comprises small strips of aluminum which act as dipoles that provide a strong radar reflection signature, although more recent methods include the use of short glass rods which are coated with suitable conductive layers for maximizing the reflected signature. The length of the dipoles is set to maximize the radar return at a specific wavelength [8]. The chaff is launched so that it initially appears in the same resolution cell of the radar as the target. The chaff's signature then aims to shift the target aimpoint away from the target and onto the chaff. The target then tries to manoeuvre out of the radar resolution cell to complete the break lock.

Flares are used to generate a large in-band radiometric signature that seduces a tracking system off the true target position. Typically, these operate at a much higher temperature than the target and early flare-types can be readily countered by dual or multiband imaging systems. Flares have, however, become increasingly sophisticated both in terms of their spectral profile and the use of distributed (multi-leaf) heat sources and smoke [25,26].

5.2. Chaff and Flare Models

The chaff is modelled as an expanding sphere where the position of the nominal centre is determined by the chaff launcher parameters, wind speed and direction, and the buoyancy of the chaff. The Radar Cross Section (RCS) of the chaff at time t is then given by:

$$\sigma_c = \sigma_{c,max} \cdot \left\{ 1 - \exp\left(\frac{t - \tau_L}{\tau_b} \right) \right\} \tag{17}$$

where $\sigma_{c,max}$ is the maximum RCS of the fully matured chaff cloud, τ_L is the launch time, and τ_b is the bloom time. The chaff RCS is also subject to a dispersion function which is represented as a Gaussian with a variable decay time.

The estimated target position is then taken as the centroid of the target and the chaff [27]:

$$E_T = \frac{\sigma_T(\beta) \cdot G(\theta_T) \cdot \begin{pmatrix} x_T' \\ y_T' \\ z_T' \end{pmatrix} + \sigma_c \cdot G(\theta_c) \cdot \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix}}{\sigma_T(\beta) \cdot G(\theta_t) + \sigma_c \cdot G(\theta_c)} \tag{18}$$

where $G(\theta_T)$ is the RF antenna pattern sensitivity at the target position, $G(\theta_c)$ is the sensitivity at the chaff location, the position of the chaff and the target are denoted by the subscripts c and T respectively. The prime on the target position vector indicates that the positions include errors due to glint. The term $\sigma_T(\beta)$ denotes the target RCS which varies with target orientation relative to the sensor system line of sight [28]. A simple RCS representation is used based on the following function:

$$\sigma_t = \sigma_{t,min} + \left| (\sigma_{t,max} - \sigma_{t,min}) \cdot \cos(2\Delta\alpha) \right| + (\sigma_{t,max} - \sigma_{t,min}) \cdot \sigma_N \cdot N \tag{19}$$

where $\sigma_{t,min}$ is the minimum RCS of the target, $\sigma_{t,max}$ is the maximum RCS of the target, and $\Delta\alpha$ is the angle between the target's direction vector and the missile's direction vector. Additionally, N denotes a normal distribution random number and σ_N is a weighting function (equivalent to the standard deviation of the fluctuation). Note that $\sigma_t \geq 0$. An example of a target RCS is shown in Figure 14a and the corresponding target positional estimate using Equation 18 is shown in Figure 14b.

(a) Target RCS generated using Equation 18

(b) Estimated target position with and without the RF cell gain function $G(\theta)$ in Equation 17
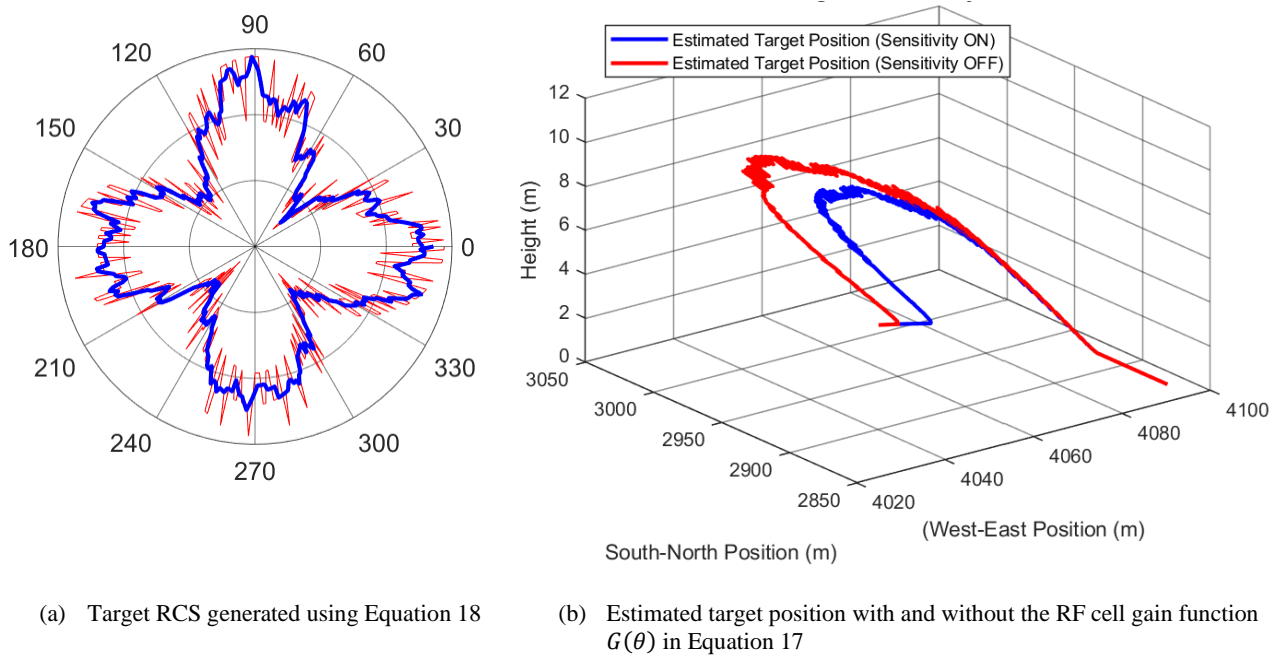
Figure 14: Modelling the effects of target RCS and estimated positions.

Flares are modelled in a similar way to chaff. The major difference here is that two spectral bands are available which allows for flare detection and rejection. It is therefore necessary to consider the physical separation of the target and the flare along the line of sight of the sensor. A proximity measure is determined to create a weighting function w_p. The corresponding target estimated position is then given by:

$$E_T = \frac{A_t \cdot w_p \cdot \begin{pmatrix} x_t \\ y_t \\ 0 \end{pmatrix} + A_f \cdot (1 - w_p) \cdot \begin{pmatrix} x_f \\ y_f \\ z_f \end{pmatrix}}{A_t \cdot w_p + A_f \cdot (1 - w_p)} \tag{20}$$

where A_t is the projected cross-sectional area of the target, A_f is the cross-sectional area of the flare, and the weighting function w_p is given by:

$$w_p = 1 - \exp\left(-\frac{\Delta p}{P}\right) \tag{21}$$

where Δp being the measured separation of the target and flare centres, and P is referred to as the *characteristic distance*.

6. RESULTS AND ANALYSIS

6.1. Methods of Evaluation

Perhaps the most important task in the development of a simulator is the validation of the underlying functions, parameters, assumptions, and range of operation. The simplest form of model validation is the direct comparison with trials data or else comparison with another model that has been subject to validation. In practice, simulators are often developed ahead of the production and trials of the modelled sensor system, and their complexity and scope are such that no comparable model is likely to be available.

In such circumstances, validation must generally rely on spot-check calculations and the performance variation as one or more model parameters are changed. This behavioural assessment is referred to here as *sensitivity analysis*.

Three types of sensitivity analysis are discussed in this Section. The first involves the variation of a single parameter associated with either the engagement scenario or the sensor system. This is an effective means of establishing the correct behaviour of the model and consequently a useful method for verifying a model's operation and the validation of the underlying mathematical representation. The second method is where a small number of parameters are varied to determine if a correlation exists. Both methods lend themselves to a relatively simple assessment of the model. For example, if the detection range sensitivity of a sensor were decreased, this should result in an increase in the predicted aimpoint error at longer ranges.

In some situations, the modelled behaviour of two or more functions can interact to give spurious results which could not be predicted. The third type of sensitivity analysis attempts to address this through a MCS whereby multiple engagement parameters are randomly varied over pre-defined ranges. Potential outliers are identified, and the parameter values are used as a starting point for a detailed investigation of the simulator. This is a more realistic system test although the associated analysis becomes increasingly difficult as the number of variables increases.

6.2. Single Parameter Sensitivity Analysis

Figure 15 provides some examples of the effects of varying a set of model parameters where one parameter is varied whilst the others are held constant. There are many ways of presenting such data and Figure 15a shows a polar plot representation of miss-distance measurement resulting from changing thirteen different target, platform, and sensor parameters (P1 to P13). The shaded region corresponds to the miss-distance values obtained when the model parameters were varied over a pre-defined range. The polar plot is a useful format for assessing the behaviour and sensitivity of a model, particularly when viewed as a time-based sequence. Figure 15a also highlights one parameter (P10) which relates to processing latency. The variation is further illustrated in Figure 15b where the parameter range for P10 covered by the polar plot is highlighted. As expected, the overall performance of the system is strongly dependent on latency, and this gives confidence that the simulator behaviour is correct.

Figure 15c shows a performance sensitivity polar plot for countermeasure effectiveness (CME). CME is defined in terms of the ability of the countermeasure (flare or chaff) to cause the estimated target position to fall outside of a detection gate with the ATR process. Note that the variable parameters are different from those used in Figure 15a. One variable is highlighted (CM3) which corresponds to the impact of wind speed on the effectiveness of flare countermeasures. An extended set of results are shown in Figure 15d, where the effect on chaff effectiveness is also shown for completeness.
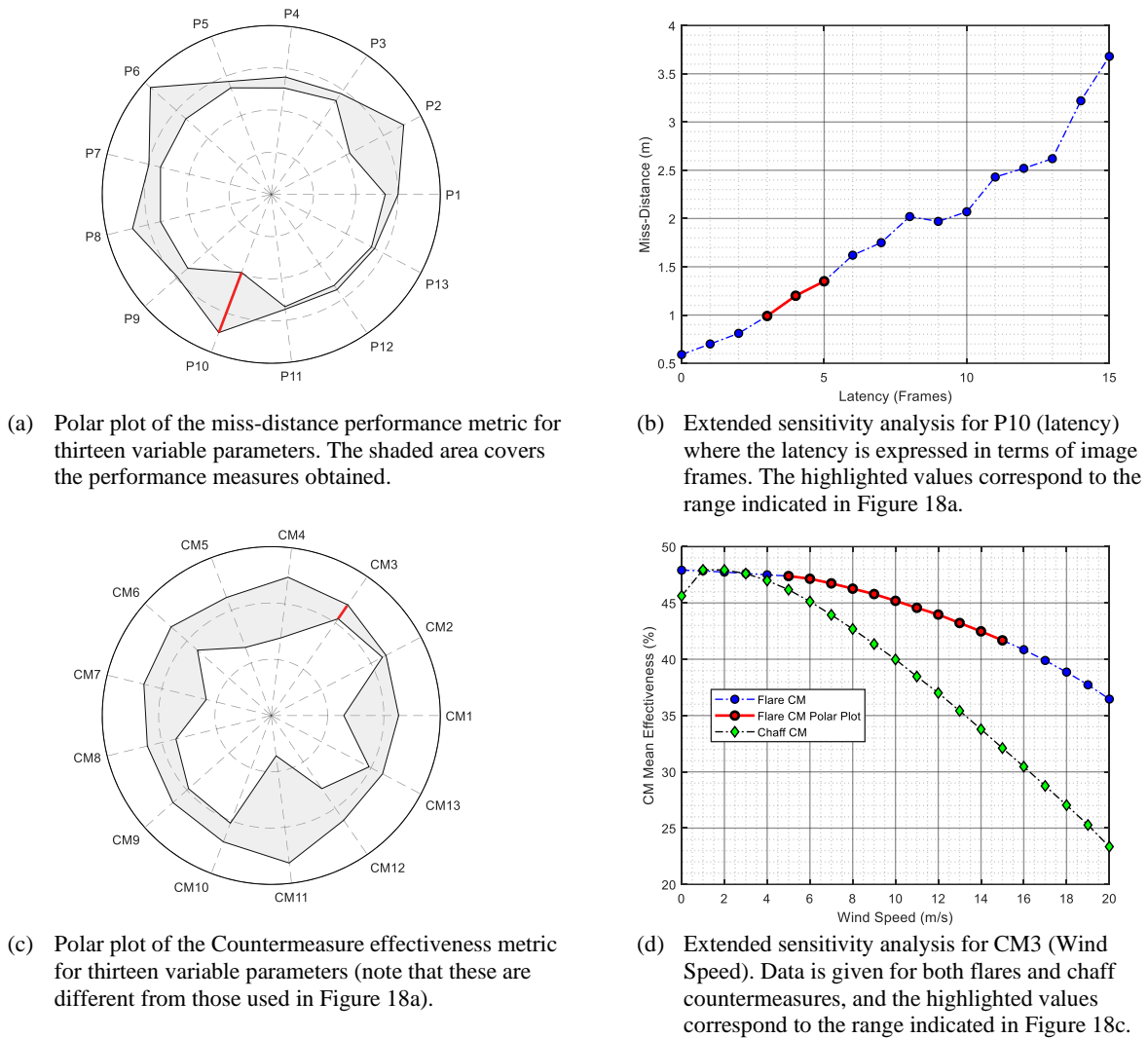
(a) Polar plot of the miss-distance performance metric for thirteen variable parameters. The shaded area covers the performance measures obtained.

(b) Extended sensitivity analysis for P10 (latency) where the latency is expressed in terms of image frames. The highlighted values correspond to the range indicated in Figure 18a.

(c) Polar plot of the Countermeasure effectiveness metric for thirteen variable parameters (note that these are different from those used in Figure 18a).

(d) Extended sensitivity analysis for CM3 (Wind Speed). Data is given for both flares and chaff countermeasures, and the highlighted values correspond to the range indicated in Figure 18c.

Figure 15: Metric presentation for the single variable verification and sensitivity analysis.

6.3. Coupled Parameter Variation

The second type of analysis is where typically two parameters are varied to determine if a correlation exists. The use of two parameters can be readily visualised using a surface plot, and an example is shown in Figure 16. Here, the effectiveness of flares and chaff are examined in terms of two environmental parameters, wind speed and direction. A coupling (correlation) between these two would be expected (the wind direction and wind speed together determine the position of the countermeasure in the fields of view, FOV, of the sensors). As can be seen from Figure 16, the simulator predicts such a relationship. At low wind speed, the wind direction has little effect (the countermeasure position does not change significantly during the engagement). However, at high wind speeds in certain wind directions, the countermeasure positions rapidly move out of the sensor FOV, and their effectiveness consequently falls to a small value.

Although both the flare and chaff variations show a minimum for the same wind direction, the chaff data shows a reduced effectiveness measure over a broader range of angles. This is primarily due to the RF gain sensitivity, $G(\theta_c)$ which reduces the countermeasure effectiveness as the target and chaff separate (Equation 18).
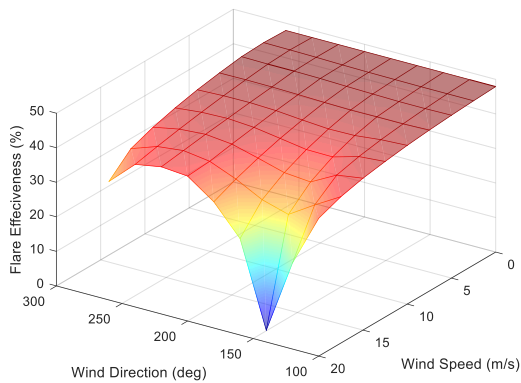
(a) Effect on flare countermeasures
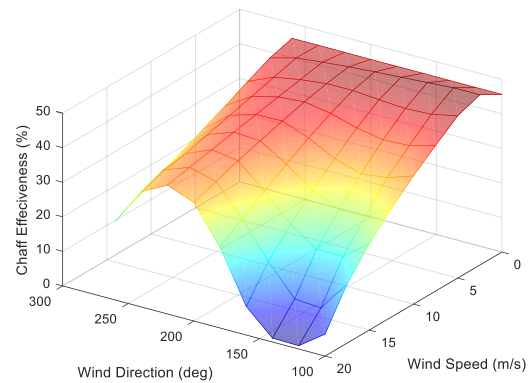
(b) Effect on chaff countermeasures

Figure 16: Connected parameter analysis to determine the relationship between wind speed and direction on the effectiveness of the chaff and flare countermeasures.

6.4. Multiple Parameter Random Variation (MPRV)

The final set of tests presented here relate to the random variation of multiple parameters between otherwise identical engagements. Typically, ten to twenty parameters are varied where each parameter varies over a defined range. This method of testing is particularly useful at identifying performance-related couplings between two or more parameters where this could be caused by an error or limitation of the model, or else be indicative of a system design limitation.

An example of MPRV is illustrated in Figure 17, where approximately fifteen different parameters were varied, and 1000 engagement simulations were used. The initial results shown in Figure 17a indicate some potentially anomalous estimates (circled). An investigation into these results identified an error in the model, and the corrected results are shown in Figure 17b. The corrected results indicate that most engagements produce a miss-distance measure of between 1.8 and 3.5m. However, there is also a dispersion of results above this which is due to variations in countermeasure-related parameters.
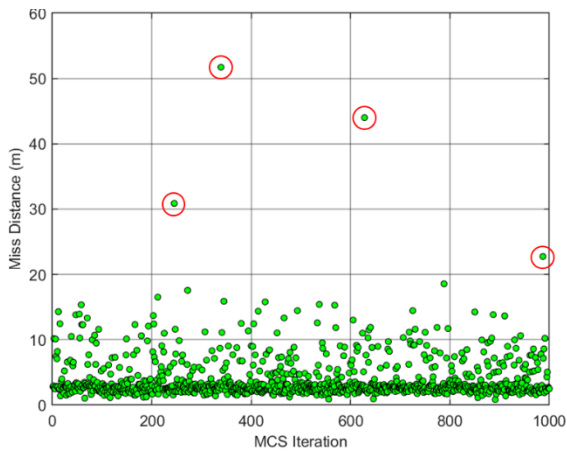
(a) Initial simulation using fifteen parameter variations over 1000 engagement simulations. Potentially anomalous results are circled.
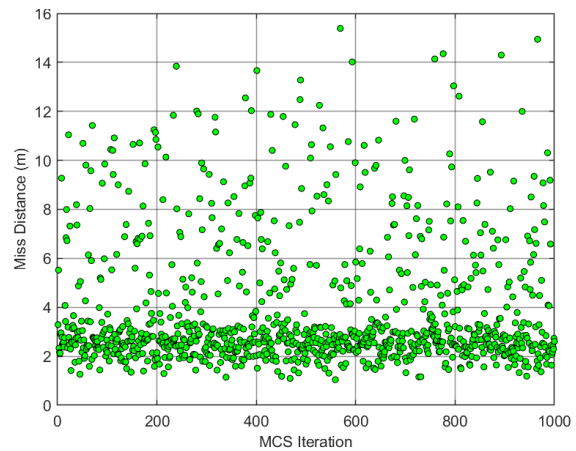
(b) Results following the correction of a model error. Note the change in the y-axis scale.

Figure 17: Illustration of random parameter variations

7. SUMMARY

The design and development of a multi-sensor automatic detection, recognition, and guidance system has been presented. Novel methods and formulations have been presented for both image-based and RF sensors using both LUT representations and mathematically derived forms. The latter was found to provide a fast and effective means of representing sensor performance in Monte-Carlo simulations where the minimisation of processing speeds in critical.

The model has been used to investigate system designs under different engagement and operational conditions. An important consideration in the development and use of a simulator is how to determine whether predicted simulator performance correctly reflects the system behaviour or is a consequence of a model limitation or error. Three methods for analysing the data have been proposed, and all three have been found to be helpful in the model verification and validation process. It is noted here that for complex models and simulations with multiple variable parameters, the evaluation can be time-consuming. It was found that the visual representation of data in polar form was helpful in reducing the analysis time, particularly when the variations are presented in a time-based sequence.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Ehrich R, and Vergez P., 'Strapdown seeker technology for the terminal guidance of tactical weapons', AGARD Guidance and Control Aspects of Tactical Air Launched Missiles, Vol. 81 No 10 (1980)

[2]   Kim D et al, 'Guidance and control for missile with a strapdown seeker', IEEE Proceedings of the 11th International Conference on Control, Automation, and Systems, Vol. 86 (2011)

[3]   Zhou G and Xia Q, 'A guidance strategy for strapdown seeker considering minimum field of view angle constraint', Hindawi International Journal of Aerospace Engineering, Vol. 2022 (2022)

[4]   Kopeika N.S., "A System Engineering Approach to Imaging", SPIE Press (1998)

[5]   Skolnik M.I., "Introduction to Radar Systems", McGraw Hill Third Edition (2001)

[6]   Patterson D., and Ashley R., "A Nonlinear Time Series Workshop: A Toolkit for Detecting and Identifying Nonlinear Serial Dependence", Kluwer Academic Publishers (2000)

[7]   Gray G.J., et al, 'Countermeasure effectiveness against an intelligent imaging infrared anti-ship missile', Optical Engineering, Vol. 52 No. 2 (February 2013)

[8]   Golden A., "Radar Eectronic Warfare", AIAA Education Series (1987)

[9]   Hickman D.L., 'Methodology for the empirical representation of image processing functions in EO/IR sensor system models and simulators', Proc. SPIE Vol. 12271 (November 2022)

[10] Hickman D.L., et al, 'Modeling and simulation framework for ATR design evaluation', Proc. SPIE Vol. 12271 (November 2022)

[11] A F. Weinmann, "Ray tracing with PO/PTD for RCS modeling of large complex objects," in IEEE Transactions on Antennas and Propagation, vol. 54, no. 6, pp. 1797-1806, June 2006, doi: 10.1109/TAP.2006.875910.

[12] Borden B.H., and Mumford M.L., 'A statistical glint/radar cross section target model', IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-19 No. 5 (September 1983)

[13] Bahtiyar S., "Target Glint Phenomenon Analysis and Evaluation of Glint Reduction Techniques", MSc Thesis Middle East Technical University (September 2012)

[14] B Kim K., 'Engagement-scenario-based decoy-effect simulation against an anti-ship missile considering radar cross section and evasive maneuvers of naval ships', Journal of Ocean Engineering and Technology, Vol. 35 No. 3 (June 2021)

[15] Garnell P., "Guided Weapon Control Systems", Pergamon Press 2nd Edition (1980)

[16] Palumbo N.F., and Blauwkamp, 'Basic principles of homing guidance', John Hopkins APL Technical Digest Vol. 29 No. 1 (2010)

[17] Sulochana S., et al, 'Precision targeting in guided munition using infrared sensor and millimeter wave radar', Journal of Applied Remote Sensing, Vol. 10 No. 3 (July – September 2016)

[18] C Shukla U., et al, 'The proportional navigation dilemma – pure or true?', IEEE Trans. Aerospace and Electronic Systems, Vol. 26 No. 2 (March 1990)

[19] D Hong J.H., et al, 'A study on parasite-effect with strapdown-seeker in consideration of time delay', Journal of Guidance, Control, and Dynamics, Vol. 42 No. 6 (June 2019)

[20] E Jang, S.A, 'Guidance algorithms for tactical missiles with strapdown seeker', SICE Annual Conference (August 2008)

[21] Campana S.B. (ed), "Passive Electro-Optical Systems", Volume 5 of "The Infrared and Electro-Optical Systems Handbook", Published by SPIE Press (1996)

[22] Weeks A.R., "Fundamentals of Electronic Image Processing", SPIE Press (1996)

[23] H Mahafza B.R., "Radar Systems Analysis and Design using MATLAB", Chapman and Hall, 2005

[24] N Kim E., 'Measurement and analysis for radar glint effect of maritime target', Journal of the KIMST, Vol. 17 No. 4 (2014)

[25] Berggren J, and Kihlen R., 'Model for simulation of IR countermeasure effect on IR-seeker/missile', Proc. SPIE Vol. 5615 (December 2004)

[26] Butters B., et al, 'Infrared decoy and obscurant modelling and simulation for ship protection', Proc. SPIE Vol. 8187 (October 2011)

[27] Siddiq M., 'Deployment of chaff in centroid mode against anti-ship missiles using a variable azimuth and elevation launcher', Calhoun Archive for the Naval Postgraduate School (1989)

[28] Charris V.D., and Torres J.M.G., 'Analysis of radar cross section assessment methods and parameters affecting it for surface ships', Ship Science and Technology, Vol. 6 No. 11 (July 2012)

Data Gathering Trials for the Development of Military Imaging Systems

Maria Niebla[a], Duncan L. Hickman[a], Eunjin Koh[b], ChanyongLee[b], Hoseong Kim[b], Chaehyeon Lim[b], and Sunghwan Kim[b]

[a] Tektonex Ltd, The Long View, Argyll Rd, Kilcreggan, Argyll & Bute, UK
[b] Agency for Defense Development, 160 Bugyuseong-dero 488beon-gil, Yuseong-gu, Daejeon, Korea

ABSTRACT

Data gathering trials remain an important part of the development and testing of imaging sensor systems. However, the role of trials has evolved to reflect emergent technology, engineering methodologies, operational requirements, and project constraints such as schedule and cost. The changing nature of data gathering trials is reviewed. Although trial programmes are still used for product acceptance, there has been an increased demand for trial data to support the design process. In this paper, the emphasis is on air-to-ground image-based military systems where the timely availability of relevant image data is critical to the development of advanced image processing software which, in turn, underpins the performance of the latest imaging systems. Other factors which affect the nature of trial programmes are also considered. These include the widespread availability of synthetic image generators and the use of low-cost drones as either targets or sensor platforms. Furthermore, the increasing use of AI data processing techniques demands a larger and more diverse image data set for training and evaluation purposes. Against this background of changing requirements, trial planning has become increasingly important. Although the great flexibility of low-cost commercial drones has resulted in them becoming a preferred solution for camera platforms, they present unique challenges, ranging from logistics through to image truthing of target locations. These issues are discussed, and recommendations made based on experience gained through multiple trial programmes.

Keywords: Trials Programmes, Imaging System Evaluation, Field Testing, Drones, UAV, Ground Truth, Image Truth Data, Air to Ground Targeting.

1. INTRODUCTION

The results of trials programmes are widely published, but the planning and procedures used for the trials have received far less attention [1,2]. The reasons for this are three-fold. Firstly, the focus of the scientific and engineering communities tends to be on the analysis of data rather than how it was gathered. Secondly, every trial event is unique regarding arrangements of sensors, platforms, targets, locations, requirements etc, so the applicability of existing trial plans to other trials is therefore limited. The third reason is the lack of the perceived benefits of planning, often coupled with time constraints for undertaking the trail and gathering the data. Although the first two points are understandable, experience has shown that there is considerable value in fully defining the trial process and clearly establishing its objectives and requirements well ahead of time. The third point is a consequence of the priorities and limited trials experience of engineers and scientists who have had little or no exposure to the practical issues and logistic challenges of a trial programme. Therefore, this paper will aim to set out some lessons that have been learnt over numerous trial programmes and common failings will be identified.

The term '*trials*' covers a very wide range of activities which depend on the nature of the equipment being tested as well as its applications. In this paper, we will define it as the *field testing* of imaging sensors and the associated gathering of data for military applications. Here, field testing implies that the trial programmes are conducted outdoors in representative environments. We will also limit the scope to air-to-ground viewing of targets and scene features of interest using remotely controlled airborne platforms or *drones*. Although these various definitions limit the scope of this paper, many of the observations and recommendations will be applicable to other trial activities.

The nature of trial activities has evolved to support the evolution of sensor system design and their testing and acceptance. Key technology events that have influenced the scope and role of trial events include the increased availability and use of synthetic imagery, the emergence of complex sensor processing, new low cost and low weight sensors, and the widespread availability of commercial drone platforms. The evolving nature of trials is discussed in Section 2.

In Section 3, trial requirements and planning are reviewed. As was noted above, trial programmes tend to be different from one another, so there is no one solution that gives an optimal trial plan. It is useful though to discuss various lessons learnt from multiple trial programmes and adapt them to reflect specific trial objectives and activities. Some of the observations offered are based on common sense, as experience has shown that common sense can be by-passed during a trial programme operation.

Section 4 extends the discussion of trial planning and conduct to include observations on the use of drones as either targets or platforms for cameras.

Once the trial has been completed, sense must be made of the gathered data. Unfortunately, this is often a large collection of image sequences and notebooks of hastily scribbled notes containing information on *ground-truth* data. Not only can the image review be a tedious task, but poorly documented ground-truth data coupled with the failings of human memory often leads to errors in the association of ground-truth data associated with the image sequences. It should also be noted that the act of measuring and documenting truth data can be a time-consuming task which inevitably restricts the amount of image data that can be gathered. Automation of the recording of image data and ground-truth data must therefore be considered in trial planning.

For detailed modelling and processing design purposes, an exact knowledge of the target's position in each image frame may also be required together with both local and global scene statistics for contrast and clutter analysis. This additional information is referred to here as *image-truth* data. Typically, image-truth data is generated post-trial. However, there are benefits in determining such truth-data during the trial as it provides confirmation that the target is in the correct position within the image frames and that the background content has the required statistical parameters. Additionally, automated image assessment can be used to confirm the image quality (noise, blur etc). For small volumes of data, it is possible to manually tag the target location and perform image quality spot-checks, but such an approach rapidly becomes non-viable as the volume of data increases. Robust automated processing software then becomes essential. One such potential method is described in Section 5 which is based on an evolving target template scheme [3]. This has been found to work effectively for many different sets of trial data (including scenes with high clutter and low contrast).

Finally, Section 6 provides a discussion and summary of the paper. This section underscores the importance of understanding the trial process, adapting lessons learnt from previous trials, and the need for robust automated processing software for data analysis. The discussion extends to the use of drones as targets or platforms for cameras, and a potential method based on an evolving target template scheme is proposed for handling diverse trial data. It advocates a more holistic approach to trials, emphasizing the importance of learning from experience and adapting to the evolving technological landscape.

2. THE EVOLVING ROLE OF TRIALS

Trials were originally aimed at evaluating camera hardware and proving that supplied equipment met the requirements for a customer's application. Typically, specific scenarios were agreed and then reproduced through the trial programme. Although, the number of trial scenarios were limited, testing of equipment in a 'real-world' environment was highly effective at detecting failures and shortfalls in the design and build. Such proving of equipment continues today as it provides the ultimate assurance that equipment is fit for purpose.

A major constraint in the planning of trials is the ability to fully replicate a military scenario such as targeting or performance evaluation because of cost, safety, and equipment availability. Additionally, it is not always possible to locate suitable available trial sites with terrains that can be used to demonstrate and evaluate the equipment's operational capabilities. Consequently, system modelling has grown alongside trials to provide an evaluation of system performance in scenarios that cannot readily be tested through trials. The scope of modelling has continued to expand, with an increasing emphasis on design development and evaluation through a product lifecycle including 'stress-testing' of the design. Trials can be used to provide data to validation sensor models at specific test points [4] and this has increased the demand on and value of recording ground truth data during trial events.

As both the complexity and scope of applications have grown, trial requirements have evolved from the simple proving of equipment to supporting the lifecycle development of a sensor-based product. The performance and behaviour of sensor systems has become increasingly determined by image and data processing, and to effectively develop and evaluate increasingly complex software requires more data in representative environments. Trials have thus grown beyond the proving of equipment and validating models, to gathering data to facilitate a sensor's design and development.

The role of trials continues to evolve, and three significant changes to trial requirements are mentioned briefly here. The first relates to the growing use of synthetic imagery as a substitute for real image data [5]. The radiometric and visual quality of synthetic scene generators continues to improve and offers many benefits in terms of image data content, data quantity and variability, and the ready availability of the scenario truth data. Setting aside the processing time, initial cost outlay, maintenance charges, and training effort, synthetic scene generators provide a viable solution for sensor system development, so the future relevance of trials will need to be re-evaluated. However, the real world has an infinite level of complexity which synthetic scene generators struggle to reproduce in terms of accuracy and variability. Military imaging system failures and false alarms are often caused by such fine scene detail. An effective use of trials data could be, therefore, to validate synthetic imagery by comparing scene metrics and system performance through a spot-check process. For those organisations with limited budgets, trial programmes are still likely to be the main source of data and the acceptance of critical equipment will probably remain based on physical demonstrations rather than analysis and computer-based demonstrations.

The second factor is that of drones which have become ubiquitous as affordable airborne platforms in both commercial and military spheres [6]. In terms of trial programmes, drones can either be used as a flexible platform for an imaging system or as the target of interest. When used as platforms for sensor systems, the reduced size, weight, and power of current imaging sensors [7] reduces the demands on drone payload, the surplus of which can then be traded for either greater endurance or the use of a lower-cost drone. The benefits and challenges of using drones is discussed later in the paper in the context of trials and trial planning. It is also noted that some military applications are developing the use of multiple drones operating cooperatively, and this has resulted in the emergence of swarm and counter-swarm technology developments [6,7].

The final factor mentioned here is the growing emergence of AI processing for analysis and decision making [8-10]. A key part in the design of an AI processing development is the training of the software against a representative and diverse dataset. Gathering such data from a trial programme can be challenging, particularly if it is required under different lighting and atmospheric viewing conditions. Again, this may perhaps be an area best suited for synthetic image generators with trial data being used to support the data validation process.

3. TRIAL REQUIREMENTS AND PLANNING

3.1. The Importance of Planning

The importance of meticulous planning in the orchestration of a successful trial cannot be overstated. It extends far beyond the mere gathering of necessary materials and arriving at the trial site. A lack of comprehensive planning can precipitate a cascade of undesirable outcomes and wasted time.

First and foremost, planning is the foundation of cost-effectiveness. Trial costs can quickly escalate in the absence of a well-planned strategy due to unforeseen complications, misallocation of resources, and inefficiencies that necessitate additional resources or the correction of avoidable errors. These complications can cause delays, disrupting the trial timeline and potentially impeding the collection of critical data, obstructing the path to conclusive results.

Secondly, planning is critical for optimising the volume, quality, and relevance of data collected during the trial. A well-structured trial, supported by clear objectives and robust methodologies, ensures that the data collected is not only abundant but also of high quality and relevant to the study. In contrast, ill-defined objectives and methodologies may produce data of poor quality or relevance, leading to incorrect results and conclusions and so undermining the trial's very purpose.

Detailed planning should be flexible in order to mitigate any potential issues and is useful in identifying any potential equipment redundancy. Contingency plans and the elimination of redundant equipment can greatly reduce the impact of unforeseen issues during the trial. Failure to plan for such events may compromise safety and increase financial and operational risks. The accumulation of these factors can result in the most serious risk associated with poor planning: failure to meet the trial's objectives. Without a clear plan outlining the trial's goals and the path to achieve them, the trial may fall short of the desired results.

Planning for a trial, particularly one involving sophisticated equipment such as drones and cameras, necessitates careful deliberation and consideration. Here are some key steps and considerations:

- Define the Goals: Determine the trial's goals, such as testing the functionality of a new camera, assessing the quality of images, or assessing the efficacy of a new data collection method.

- Determine Required Resources: Determine the resources needed for the trial, such as drones and cameras, personnel, funding, and time. Obtain all necessary permissions and licences, especially for drone operations.

- Create a Detailed Plan: Make a comprehensive plan that outlines every step of the trial process, from setup and operation to data collection and analysis. Ensure all personnel are fully aware of their roles and responsibilities. Anticipate potential problems and develop strategies to deal with them. Include contingency time to account for unexpected changes, such as weather. Try to arrange access to the site on alternative days ahead of time in case weather on the trial day is bad.

- Test Equipment: Prior to the trial, thoroughly test all equipment to ensure optimal functionality. This includes drones, cameras, and any other technology, as well as ancillary equipment such as batteries, memory cards, and extension cords.

- Equipment Inventory: Generate a list of all required equipment and spares (including batteries for drones). Ensure that power supplies are fully charged, and both general and required specialist tools are available.

- Site Inspection: Conduct a pre-trial site inspection to become familiar with the environment and identify potential issues that may impact the trial. For instance, if the site is large, how will personnel communicate with one another? Consider also airspace restrictions, and potential obstacles (such as electrical pylons and overhead cables) when conducting drone trials.

- General Logistics: The natural tendency in trial planning is to focus on the camera systems and drones. However, many trials fail to achieve their objectives because of issues arising out of general logistic failures. This includes the ability to access the trial location, the availability of shelter from rain and sun, and access to power. Remember that cold weather can shorten battery life.

- Data Annotation, and Storage Plan: This is especially important for trials involving large amounts of data, such as high-resolution images or video. Maintain detailed notes throughout the trials to aid in data identification and to record any significant details that may influence the overall conclusion.

- Ground-Truth Equipment and Data Recording: Determine before the trial what ground-truth measurements will be required, who will record it, and establish the methodology of associating it with the recorded imagery.

- Image-Truth Software: If required, prepare and fully test the image-truthing software, and determine how the information will be used to inform the live trial activities. This includes both the image composition and the image quality.

- Drone Flight Control Software: Commercial tools are available to allow the flightpath of the drone to be pre-defined. However, these may require a network signal to connect to online services which may be affected by poor reception areas on a remote trials site. Additionally, commercial drones tend to perform software checks and

software updates at the most inconvenient times, so always install the latest software before the trial and/or disable the update facility for the duration of the trial.

- Trajectory Accuracy: Flight control software is not always as accurate as required. This can be due to errors in the location of the drone, or effects such as high wind speeds or sudden gusts. It is therefore always useful to establish and measure reference points in the scene that can be used to confirm the location of the drone.

- Remote Camera Control: The means of controlling the camera system (recording cycle, line of sight etc) can be controlled from the ground or by using automated software that is linked to the flight path. The latter has been found to be generally more reliable, particularly when a large volume of data is to be gathered. It is important that the image frames are accurately time-stamped.

- Risk Assessment and Mitigation: Identify potential risks and devise mitigation strategies. This could include equipment failure, data loss, or safety concerns. Having backup equipment and a flexible plan can help you deal with unexpected problems.

- Dry Runs: Perform dry runs of the trial to test the plan and make any necessary adjustments. This can help identify potential issues before the trial and increase the likelihood of success.

- Documentation: Ensure the personnel responsible keep detailed records of the trial process and results. This will be critical for analysing the results and planning future trials.

- Review and Adjust: During the trial, review the images captured in-situ to make sure the systems are correctly calibrated, and that the gathered data is consistent with the quality expected and required. After the trial, go over the results and the process. Use this information to improve future trials.

Finally, planning is more than just a preliminary step in the trial process; it is the foundation upon which successful trials are built.

3.2. Setting Realistic Objectives

The process of setting trial requirements and planning are of paramount importance. Planning tasks are generally badly performed with the result that a trial can fail to meet its objectives. Such failures arise through setting trial requirements in isolation from the conduct of the trials and the over-estimation of what can be achieved during the trial itself. In general, there tends to be a mismatch between the expectations from a trial and what can be achieved in practice. Most trials produce far less data than was planned for. There are many specific reasons for this, but some of the principal ones include:
- Logistics and set-up times
- Equipment failures and lack of spares
- Weather conditions
- Safety and legal constraints
- Lack of basic and specialist tools
- Software failures
- Communication between trial staff

The term *logistics* is used as a general term to describe the movement, placement, and protection of equipment, the facilities for the trial's personnel, and the provision of power. Very often it is logistical problems that are the limiting factor in the effectiveness of a trial. During the planning stage, the logistical challenges should be considered in detail and mitigated through detailed trial planning.

Many of the logistical problems encountered during a trial are a consequence of the trial location which very often involve trial engineers standing in a remote field with no access to water, fuel, or electricity (Figure 1). For example, drone batteries will provide a limited operational window of between 30 and 45 minutes' flight and recording time. However, to fully recharge them can require over three hours. Therefore, when the supply of charged batteries has run out, the trial will become centred on battery charging. This will result in delays and the use of partially charged batteries that provide a much-reduced operational window. To charge the batteries, a generator or portable power generators will be needed if a mains supply or car-powered inverter is not available. In terms of trial planning, the available power should be reflected in the schedule with shorter duration recording tasks being performed towards the end of the trial.

Figure 1: A typical remote trial site with the control post and equipment shown in the background.

Equipment problems always seem to occur with either (or both) hardware and software. All equipment should therefore be tested before the trial and potential problems identified. Where possible, spares should be made available together with tools and drivers required to replace equipment or update software. One issue that occurs on a regular basis is that of automated software updates and remote datalinks associated with commercial products such as drones. This can this cause delays and failures, not to mention frustration.

Weather conditions are self-explanatory, but it is noted that trials are often undertaken without adequate provision for the protection and cleaning of equipment during rainy conditions. However, critical to the operation of drones is the wind speed (both mean and gust levels). It is straightforward to measure wind conditions at ground level. However, the values tend to increase with altitude, and they also depend on the terrain of the trial site. Consequently, a drone can experience unpredictable wind conditions which will impact its ability to follow a pre-defined trajectory. This is particularly so for smaller drones which can be severely affected by gusts.

In terms of the final points in the above list, equipment must be operated in accordance with the relevant safety and legal requirements. This includes the maximum operating altitude, as well as operation near buildings, pylons, and people. When using a drone, it is generally unwise to fly it towards a target because a failure in the drone command link could result in severe equipment damage. In many cases, the same effect can be realised by recording flights away from the target and then reversing the sequence after the data has been gathered (Figure 2).
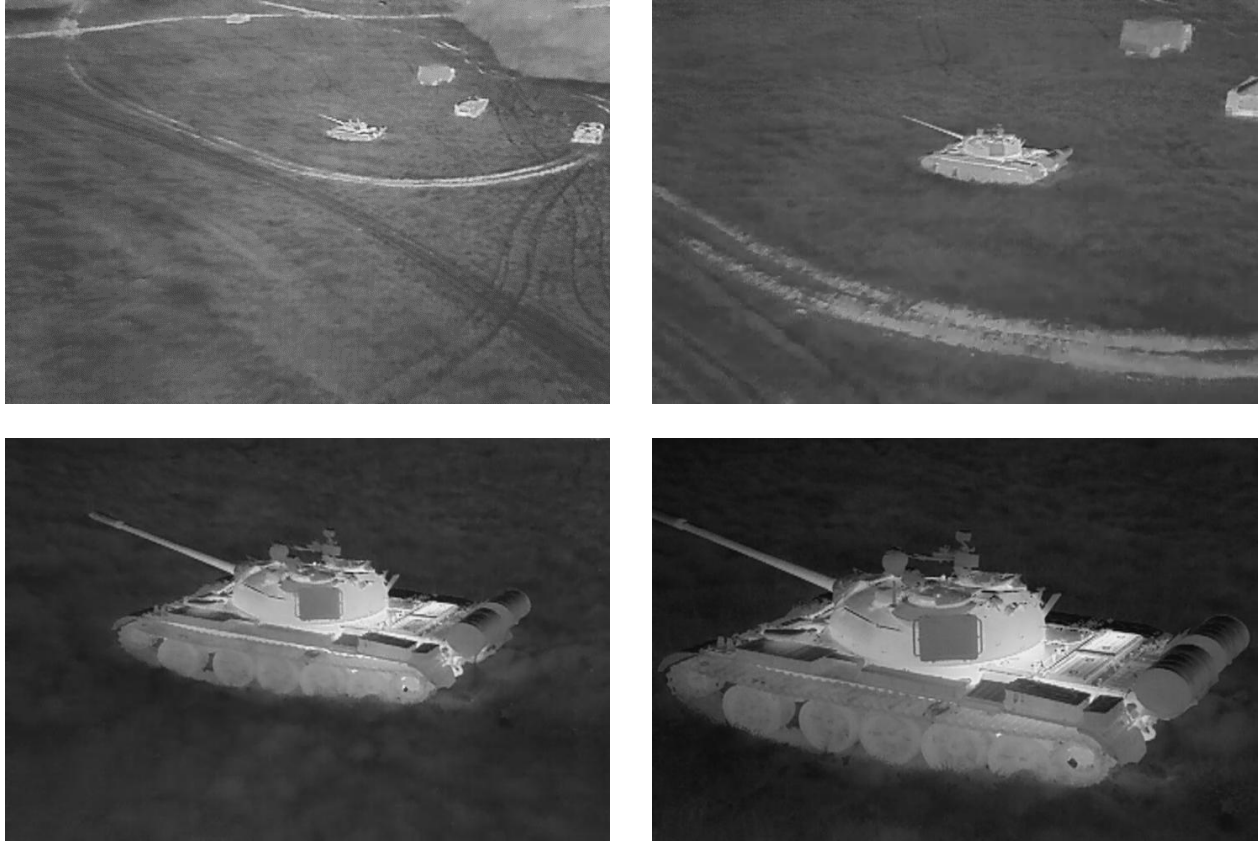
Figure 2: An infrared sequence of images recorded using a drone flying away from the target. The sequence of images was reversed after the data was recorded. This approach works well for static targets in fixed backgrounds.

3.3. Sequencing and Ordering of Events

From an engineering design perspective, trials provide an opportunity to test a sensor system in terms of both its sensitivity and its performance in different scenarios or under different viewing conditions. For example, it is often useful to test a camera in the same scenario but using small variations in the scene content to determine performance sensitivity. Unfortunately, there is a tendency to test such variations before moving onto a different scenario because it is assumed that this will be more efficient with respect to time. Consequently, small variations are effectively prioritised at the expense of the wider data diversity. The reality of most trial programmes is that the time required for the trial is always much greater than that available. As a result, trial events can produce multiple sequences of very similar data.

A logical order of trial tasks from a design perspective is therefore not necessarily the most logical order from a trial perspective. As was noted above, logistical issues such as battery condition should be considered in the planning stage, and this may require the relegation of some important shorter duration tasks to the end of the trial event. Care, of course, needs to be taken to ensure that sufficient time is allowed to do these in a constantly shifting trial schedule. Additionally, evolving weather conditions can influence the ordering of trial events, and often any changes or cancellations can only be made just prior to (or even during) the trial.

It has often been found useful to begin a trial with a relatively simple scenario as a means of familiarisation with the trials equipment and the operation of the sensor systems and platform. Despite pre-trial preparation work, it has been found that hardware, software, and personnel tend to behave differently when they are in a remote trial location, and unforseen problems invariably arise. There is always a pressure to skip such preparatory tests and start with the more important tasks. The problem with this is that mistakes will inevitably be made, and these could comprise the quality of important data sets.

Finally, and perhaps most importantly, it is important to recognise that even the most successful trials will produce less data than is hoped for. This is particularly so for new systems at new trial site locations where the amount of data gathered can be significantly less than 50% of that planned for. Therefore, it is important to know which scenarios are essential and which are of secondary importance.

3.4. Testing, Recording, and Truth Data

A natural aim of a trial programme is to gather as much data as possible. Although the sensor systems may appear to be working, this does not guarantee the quality of the captured data. At the most basic level, such failures can occur if the camera does not record all the details in a defined scenario because of an alignment problem, or if errors occur with the camera or recording equipment. The quality of the imagery may also be degraded through image blurring effects as can occur if a camera stabiliser fails (Figure 3). Although image blur can be reduced using image restoration processing [11], the image quality is still compromised.

It is therefore important to check the image quality during the trial as this allows replacement data to be gathered and any equipment problems fixed on the spot. It is possible to automate the image quality assessment using pre-prepared image metrics. However, in most cases, a regular visual inspection of some of the image data is generally sufficient.

It is also important to assign someone on the team to keep an accurate log of the trial in terms of captured image data. Unfortunately, in many cases this is not done, leading to confusion and assignment errors after the trial has been completed. Each image sequence should be catalogued and linked to specific scenario trajectories as well as any supporting data such as camera parameters, scene temperature measurements, and weather conditions. A common problem that occurs on trials is the recording and saving of data from events that have failed or not been completed correctly. This can cause considerable confusion when data is being analysed weeks or months after the trial.
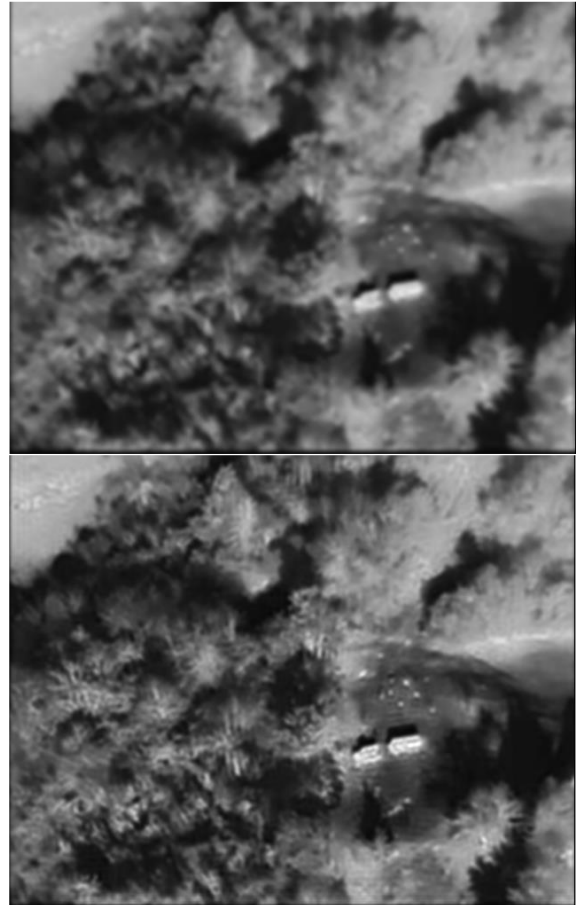

Figure 3: Image blur resulting from a camera stabiliser failure (top) and the restored version (bottom)

An important aspect of trials is the need to measure important parameters associated with a scenario such as temperatures, positions, heights, orientations, and configuration parameters. This is referred to here as *ground-truth* data. The nature of the required measurements depends on the requirements of the trial. For example, if the trial is to provide a demonstration of system-level performance measure such as aim point accuracy for a targeting sensor, then the ground-truth measurements should support the assessment of this metric. If, however, the trial data is to be used in a development programme where the system and scenario are to be modelled, it is useful to measure other parameters such as scene and target temperatures (Figure 4).

As noted above, the ground-truth data should be associated with the recorded trials data such as image sequences.

Figure 4: Instrumenting a target to measure its temperature at several locations for post-trial modelling and analysis

4. OBSERVATIONS ON THE USE OF DRONES

Drones are increasingly being used as either targets or platforms for carrying camera systems or other payloads in military applications. Although drones have been available for many years, they have typically been high performance and expensive systems operating over long ranges at medium to high altitudes [12]. These more specialised platforms generally have greater capability in terms of endurance, payload capacity, navigational control, and stability. However, events in Ukraine have highlighted the potential military role of re-purposed commercial drones for short-range surveillance or weapon delivery [13]. It is anticipated that such lower-cost drone systems will continue to evolve, with increased control, accuracy, and the ability to act cooperatively in groups or swarms that can swamp defensive systems [6,7]. A range of drones used in recent trial programmes is shown in Figure 5.

(a)   Specialist, high-performance fixed wing drone

(b)   Modified commercial drone (6 rotors)

(c)   Commercial drone (4 rotors) with multiple cameras

(d)   Lightweight commercial drone

Figure 5: Examples of drones with varying levels of performance and cost. Many trial programmes now rely on the use of low-cost commercial devices.

The detection of small drones using imaging systems is extremely difficult because of their small size and limited thermal signature. The ability to visually detect such drones is further constrained if they are viewed against a cluttered background and, in most cases, detection can only be achieved if its motion presents a significant relative angular movement. This is illustrated in Figure 6 for the case of a small four-rotor drone and where a larger, fixed-wing drone is shown for comparison. Only images from shorter ranges are shown so that the drones can be seen.

(a)   Fixed-wing drone at a range of 120m

(b)   Fixed-wing drone at a range of 100m
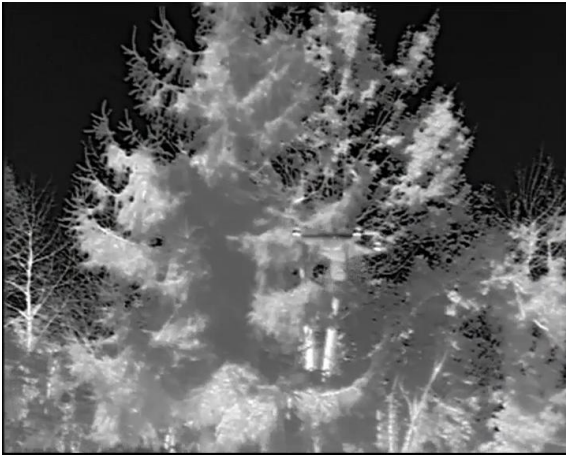
(c)   Small four-rotor drone at a range of 30m

(d)   Small four-rotor drone at a range of 50m

Figure 6: Comparison of the detection of large and small drones. The range values are approximate.

Drones provide extremely versatile platforms and can allow the rapid capture of image data from a variety of angles and positions (Figure 7). Their flexibility and low cost are such that they are replacing more traditional methods such as camera poles and raised platforms. Drones are also capable of executing realistic trajectories to provide dynamic sensor data which is invaluable for many military applications.

By using commercially available software, it is possible to pre-plan the flight trajectories of drones prior to the trial. Additionally, the triggering of data recording and camera lines of sight can be linked directly to the trajectory. This offers enormous potential for optimising a trial. Care should be taken, however, to ensure that the recording time of the image frames is accurate and matches that of other trials equipment.

Figure 7: Illustrating the versatility of drone mounted camera systems

Unfortunately, the performance of current navigation and control software is variable, and the ability of smaller drones to follow pre-programmed courses can be adversely affected by wind speed and gusts. With commercial on-board hardware and software, it is also a challenge to record accurate positional data of a drone that is directly linked to the recorded camera imagery.

Finally, it is noted again that the payload capacity and endurance are critical factors in selecting a drone for trial purposes. They are the most limiting factors when using small commercial drones. Platforms are now available that include mechanically stabilised visible band and infrared (uncooled) cameras which offer a flight time of typically 30 to 45 minutes [13]. However, such commercial systems have limitations such as the ability to record 16-bit imagery at frame rates of 30Hz or above. For the testing of new camera systems, greater payload capacities are generally required, and this demands the use of larger and more powerful drones (Figure 8).

| (a) Manual adjustment of the camera line of sight | (b) Use of a 6-rotor drone for lifting a bespoke imaging and data recording system |

Figure 8: Setting up an using a higher-performance drone for the testing of a new imaging system.

5. POST-TRIAL PROCESSING AND IMAGE TRUTH

Once the trial has been completed, it is necessary to inspect the gathered image data and identify potential issues or degradations. These are most easily seen by viewing the data as movies rather than individual frames. A typical problem which occurs when moving camera platforms are used, is that of variable image blur. For slow moving drone platforms, the blur is due to vibration from the rotors, and platform tilts caused by wind shear or gusts. In these situations, image quality can be improved through a restoration process [14].

The development of image processing and system models requires differing amounts of information from the trial as discussed previously. From the perspective of image processing, one of the most important measures is the location of a target (or targets) in each image frame. For most trial measurement systems, the available truth is not sufficiently accurate, and post-trial processing is required.

The automated determination of a ground target's position is a difficult problem for air-to-ground applications where the target is often seen against a heavily cluttered background. Generally, global image thresholding is unreliable for target segmentation, and more advanced processing techniques are required.

The approach recommended here is based on an evolving template matching scheme where the initial target position is manually set, and a target template is extracted using metrics such as contrast or spatial structure [14]. For the case where there is relative motion, a proximity region can be defined to determine the likely position of the target in subsequent frames or, alternatively, a simple tracking algorithm could be used. A search box is then defined based on that of the previous image, and a template correlation performed to accurately locate the target. The template is then regenerated from the current frame in readiness for the next frame. This is illustrated in Figure 9 for a target in a heavily cluttered background. Furthermore, the technique can be used for multiple targets as shown in Figure 10.
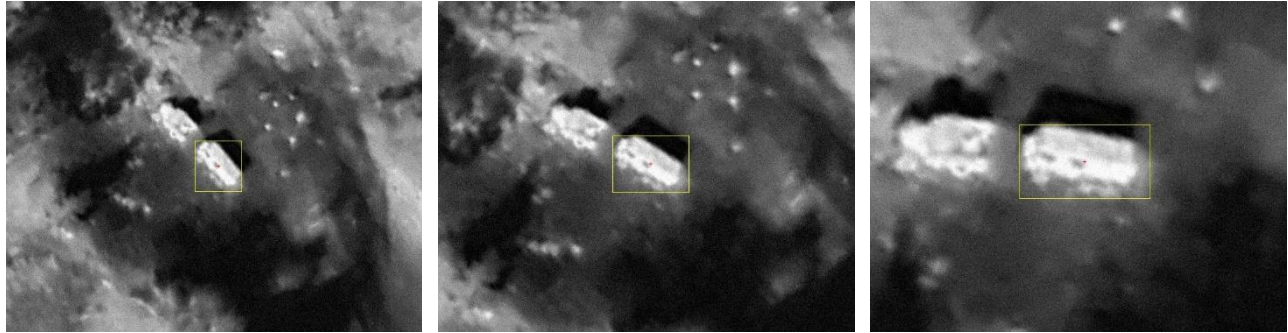
Figure 9: Sequence of image frames with the target location annotated with a boundary box.

Figure 10: Annotation of multiple targets based on the use of multiple evolving template correlation.

The evolving template technique has been found to be relatively robust, although errors can occur if the target passes through a region of high clutter, or if its contrast is reduced. Two effects have been noted. The first is that the outline of the target can change, particularly if the target features are fine details or low contrast (as is the case of the military vehicle shown in Figure 11). This results in a reduced level of accuracy for both the target location and the template used for subsequent image frames. The second effect is that if the target contrast is reduced significantly or is in a highly cluttered environment, the correlation may fail. This effect can be reduced by limiting the size of the search area although a manual reset of the template may be required.
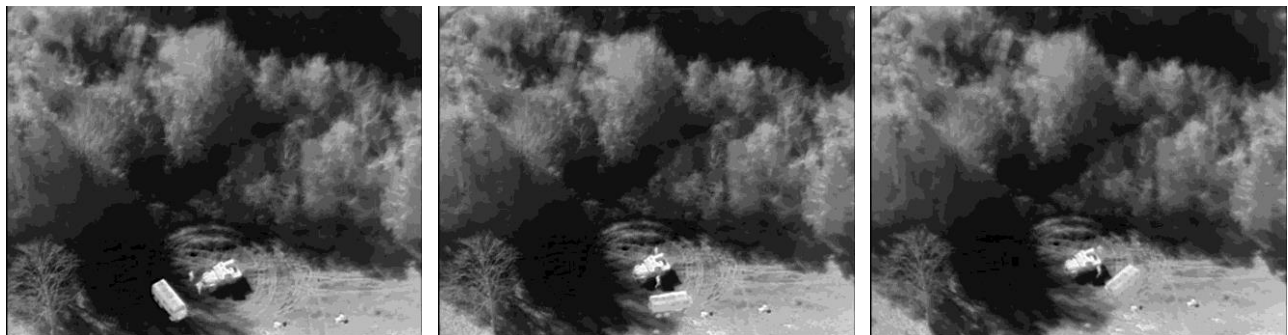
Figure 11: Illustrating the variation of target contrast.

6. DISCUSSION AND SUMMARY

Technological advancements, operational requirements, and project constraints have all had a significant impact on the evolution of data collection trials, particularly in the field of imaging sensor systems. Trial programmes are now used to support the design process rather than just product acceptance, particularly in the context of air-to-ground image-based military systems. The availability of relevant image data is critical for the development of advanced image processing software, which is the foundation of modern imaging system performance.

The introduction of synthetic image generators, as well as the use of low-cost drones as targets or sensor platforms, has also changed the nature of trial programmes. Because AI data processing techniques are becoming more popular, a large and diverse image data set is required for training and evaluation. As a result, trial planning has become increasingly important, with the use of low-cost commercial drones posing unique challenges.

The role of trials is evolving, with synthetic imagery, drones, and AI processing all playing a role in shaping their future. In terms of image data content, data quantity and variability, and the availability of scenario truth data, synthetic imagery provides advantages. The complexity and variability of the real world, on the other hand, frequently cause system failures and false alarms, which synthetic scene generators struggle to replicate accurately. As a result, trial data could be used to validate compatible synthetic imagery.

Drones have become commonplace as low-cost, airborne platforms in both commercial and military applications. They provide flexibility and accessibility, making them a popular choice for camera platforms. They do, however, present unique challenges ranging from logistics to image truthing of target locations.

Artificial intelligence processing is increasingly being used for analysis and decision making. A representative and diverse dataset is required for training in the design of an AI processing development. Obtaining such data from a trial programme can be difficult, especially in different lighting and atmospheric viewing conditions. Synthetic image generators with trial data to aid in data validation may be best suited for this task.

The success of the trial is related to the nuances of trial requirements and planning, which acknowledges the uniqueness of each trial programme and the lack of a one-size-fits-all solution. The significance of learning from experience and adapting these lessons to the specific context of the trial is emphasized, as is the frequent disregard for common sense in trial activities, highlighting the importance of consistent application.

In conclusion, data collection trials have evolved significantly over time, influenced by technological advancements, operational needs, and project constraints. The use of synthetic image generators, low-cost drones, and artificial intelligence processing techniques has influenced the nature and future of these trials. Despite the difficulties, these trials continue to be critical for the development and testing of imaging sensor systems, particularly in military applications. The authors would like to advocate for a more holistic approach to trials, emphasizing the importance of learning from experience and adapting to the evolving technological landscape.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Hickman D.L., 'Modelling and simulation framework for ATR design evaluation', *Proc. SPIE* Vol. 12271 (2nd November 2022)

[2]   Hickman D.L., 'Modelling and simulation framework for ATR design evaluation', *Proc. SPIE,* Vol. 12271 (2nd November 2022)

[3]   Borghgraef A., 'Obtaining ground truth in C-UAS trials', *Proc. SPIE, V*ol. 12270 (2nd November 2022)

[4]   Foks-Ryznar A., 'Trial as a pragmatic and systematic approach for assessing new solutions in crisis management and rescue operations', *Proc. SPIE,* Vol. 11176 (6th November 2019)

[5]   Haddleton G.P., 'Increasing use of high-speed digital imagery as a measurement tool on test and evaluation ranges', *Proc. SPIE,* Vol. 4308 (11th April 2001)

[6]   Edwards C., 'Use of synthetic imagery in target detection model improvement', *Proc. SPIE,* Vol. 4372 (10th September 2001)

[7]   Jackman A., 'Consumer drone evolutions: trends, spaces, temporalities, threats', *Defense & Security Analysis,* vol. 35, no. 4 (2019)

[8]   Mahadevan P., 'The Military Utility of Drones', *CSS Analyses in Security Policy,* Vol. 78 (July 2010)

[9]   Rasch R., Kott A. and Forbus K., 'Incorporating AI into military decision making: an experiment', *IEEE Intelligent Systems,* Vol. 18, No. 4 (July-August 2003)

[10]   Szabadföldi I., 'Artificial Intelligence in Military Application – Opportunities and Challenges', *Land Forces Academy Review,* Vol. 26, No. 2 (2021)

[11]   Harvey C., 'Simulation of a new 3D imaging sensor for identifying difficult military targets', *Proc. SPIE,* Vol. 6950 (13th May 2008)

[12]   Mahdi A. J., 'Performance and Analysis of the Altitude for Optical Channel Drones', *Proceedings of 2nd International Multi-Disciplinary Conference Theme: Integrated Sciences and Technologies*, Sakarya, Turkey (2021)

[13]   Zlepko S., 'Tactical-exploring unmanned aviation systems: Current state and prospects in Ukraine', *13th International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science*, Lviv, Ukraine (2016)

[14]   Gerken M., 'Military reconnaissance platform for the spectral range from the visible to the MWIR', *Proc. SPIE* Vol. 10177 (3rd May 2017)

[15]   D. L., 'Methodology for the empirical representation of image processing functions in EO/IR sensor system models and simulations', *Proc. SPIE V*ol. 12271 (November 2022)

Modular turn-key system for testing the spectral response of detectors and cameras in the VIS/NIR and MID/FAR IR

Dario Cabib, Moshe Elkabets

CI Systems (Israel) Ltd., Ramat Gavriel Industrial Park, Migdal Haemek 2310101, Israel

ABSTRACT

CI has developed a modular optical system, which provides the sensor stimulus to obtain the required quantitative spectral response of single detectors, detector array engines and camera systems with small to large aperture optics. The system includes interchangeable sources, Circular Variable Filter scanning monochromator (CVF) and collimating and focusing optics to project monochromatic radiation (in focused or collimated configuration, according to need) towards the sensor. When spatial patterns are also used on the focal plane, the Unit Under Test (UUT) is fully characterized in both spectral and spatial domains.

1. INTRODUCTION

As a company developing and manufacturing electro-optics systems' testers of many kinds for different applications, CI Systems (CI) has been exposed to a large amount of varied testing situations. In this paper we describe systems used to measure the relative (and in some cases quantitative) response of detectors and cameras as function of wavelength, both alone, and combined with spatial resolution, in case of imaging systems.

We can classify the tested devices into the following categories: i) single detectors, ii) array detector engines, and iii) cameras. We define each of these devices as including, not only the sensitive chip, but also the electronics needed to extract the chip signals, and make them available to a recording system, such as a digital processor for signal analysis. The most common applications for testing such devices are in the R&D stage for optimization of detector performance, and in the manufacturing processes, usually for quality assurance and quality control purposes.

Historically, CI based its core technology on spectral radiometry for remote sensing applications, and over time it developed its own monochromator optical element, a so called CVF (Circular Variable Filter)[1], which is a wheel on whose circumference linearly variable thickness layers are deposited, in order to transmit radiation of continuously variable wavelengths as it is rotated around its axis. This type of scanning monochromator has been originally described in the old Infrared Handbook[2]: it is used in the so called SR 5000 spectroradiometer, a single point Field of View radiometer, measuring spectral radiance and/or radiant intensity emitted by remote sources. In this radiometric configuration the CVF successively decomposes the radiation coming in from a far field into monochromatic components, and a single element detector gives a signal proportional to the intensity of these components. The detector signal can be calibrated by comparison to the spectral signal obtained by measuring a standard source such as a blackbody, and recorded in quantitative physical units of spectral radiance, radiant intensity, or irradiance, as needed according to the application.

In contrast, when the CVF is used in front of a light source with appropriate optics, it can provide monochromatic radiation in successive single wavelengths to test and measure a sensor or a camera spectral response and wavelength dependent spatial resolution, as the case may be. Different configurations of a controlled and calibrated monochromatic radiation projection system are described in this paper.

2. TYPICAL MULTI-SPECTRAL CAMERA TESTER

The simplest system built to measure a camera spectral response, spatial resolution, etc., is according to the following configuration.
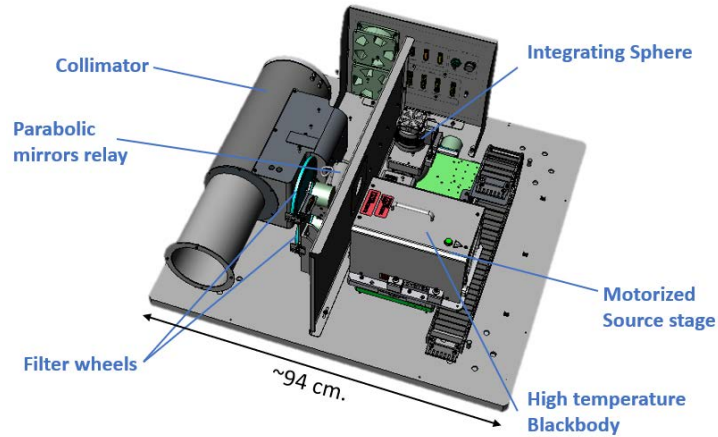
Figure 1. The system shown here is used to measure different camera parameters such as MTF, response, etc., at different discrete wavelengths. The camera to be tested may be visible or infrared: the integrating sphere is used as a uniform visible radiation source, and the blackbody is the infrared source. The filter wheels are one for infrared and one for visible wavelenghts, and are interchanged accordingly with the sources. The target wheel shown in the diagram of figure 2 is included in the container between the collimator and the filter wheels.

The basic feature of the system of figure 1 is that the radiation projected onto the Unit Under Test (UUT) is collimated, so that it can simulate a remote scene from infinity, as this is a camera's usual mode of operation. The source is a blackbody (for infrared), which is interchangeable with an integrating sphere (for visible or near infrared light); the projecting optics is a reflective collection of mirrors (5 or 6" diameter aperture and F#6), for operation in all wavelengths without chromatic aberrations. The test wavelengths are limited to a number of discrete values that can be selected by computer control. Two filter wheels are present, one for infrared filters and blackbody operation, and one for visible wavelengths and integrating sphere source operation. The target wheel is in a container attached to the collimator.

In figure 2 the optical layout of the system of figure 1 is shown. The optical relay focuses radiation from the source onto the filter and target wheels. The target wheel is on the focal plane of the Intermediate Level Electro-optical tester (ILET) collimator, because the patterns on it are imaged by the UUT, while the filter wheel can be slightly off the focal plane, since the filters' task is only to provide wavelength separation. The collimator projects the rays towards the (UUT), being usually an infrared camera. This type of test system allows measuring MTF and other parameters at the different wavelengths of the bandpass filters present in the filter wheel, or in integrated mode, in absence of filters.
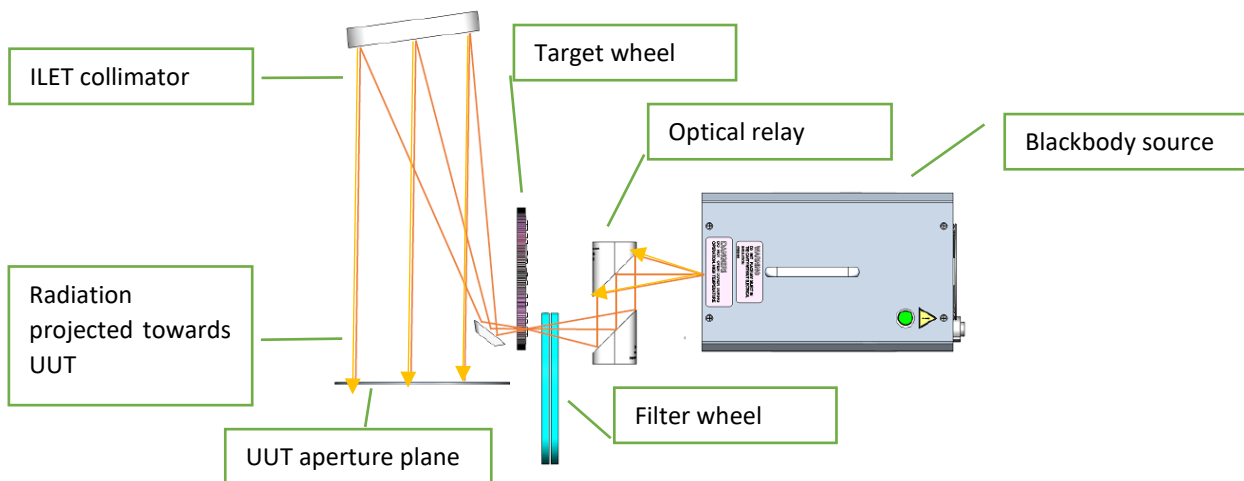

Figure 2. Optical diagram of the system of figure 1. The integrating sphere of figure 1 is not shown here.

3. SPECTRAL CAMERA TESTING WITH CONTINUOUS SPECTRAL SCAN

The system of the previous section can be used in discrete single wavelengths, but for certain applications the detector manufacturer is required to provide a continuos graph of response or normalized detectivity D* as a function of wavelength.

3.1 What is a CVF and how it works
Referring to reference 1:

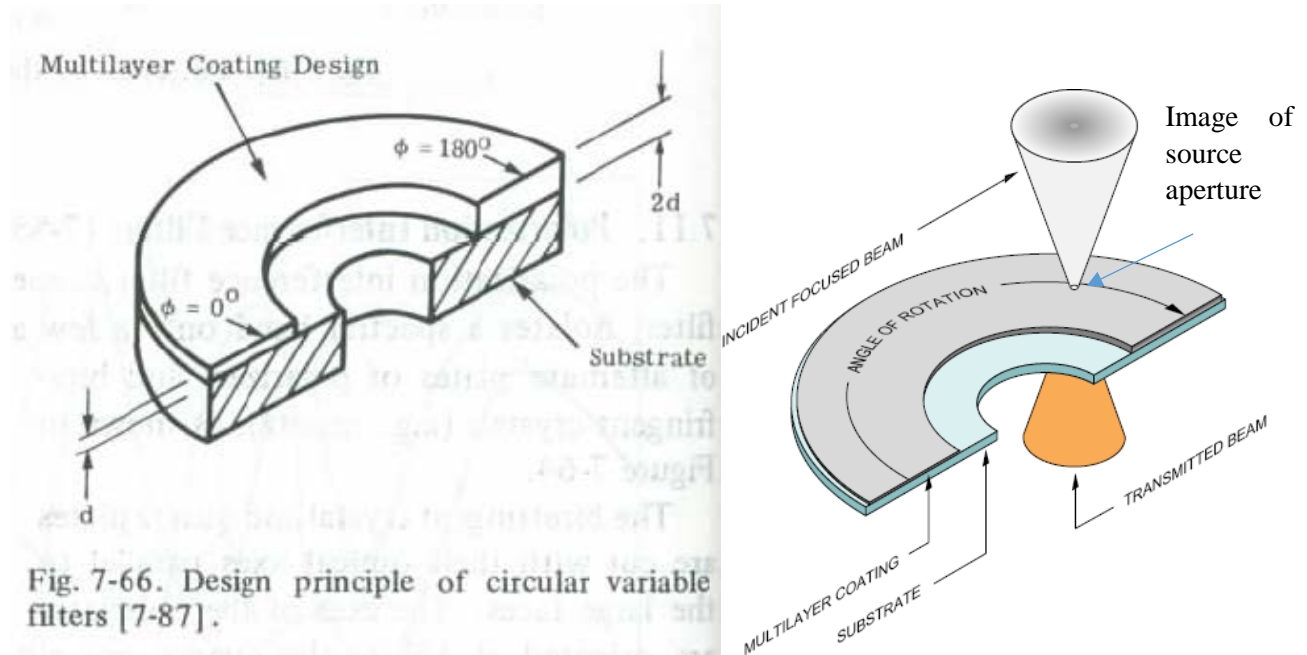Figure 3 illustrates a simple explanation given in the Infrared Handbook[2], page 7-118 of what a CVF is.

Fig. 7-66. Design principle of circular variable filters [7-87].

Figure 3: On the left side the CVF element is as shown in the Infrared Handbook of 1978[5] (figure scanned from the book). On the right side is a CVF wheel diagram showing the input and output focused light cones from the source. The CVF surface is on the focal plane of the system, to insure that the image of the small source aperture is transmitted at the specific wavelength corresponding to its position on the CVF's perimeter.

Quoting from the Handbook: "…a circular variable filter is an interference filter composed of a film deposited on a circular substrate. The thickness of the film, and hence the wavelength, varies linearly with angular position, ϕ, on the substrate…."
In addition, as shown in Figure 3 the film thickness varies from d at $\phi = 0^0$ to 2d at $\phi = 180^0$ so that also the peak wavelength of transmission in each position around the substrate varies from a certain λ_0 at 0^0 to $\lambda_{180} = 2 \lambda_0$ at 180^0. Other configurations of 90^0 or 360^0 segments can be manufactured. In order to obtain narrow peaks such as shown in Figure 5 the film thickness must be made of a large number of films of two alternating materials, one of low and one of high refractive index. The substrate and film materials must be suitable for the specific needed spectral range and resolution.

In all the CVF configurations manufactured by CI the peak wavelength varies approximately linearly with the angle around the substrate, as seen in Figure 3.
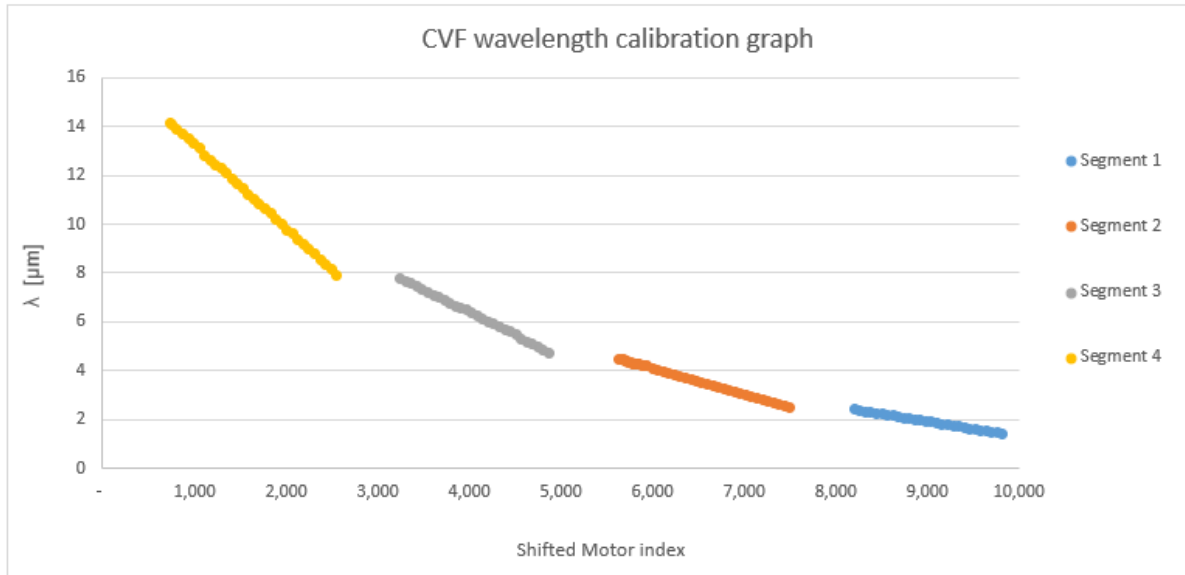
Figure 4: A typical peak wavelength of transmittance versus wheel motor index (a linear function of the angle variation) of a CI CVF, showing its linear behavior versus angle. Peak transmittance is typically between 50 and 80%. In this configuration four 90º segments are juxtaposed one in contact with the other, all with decreasing wavelength with motor index. The gaps between segments are index regions where the wheel is used to hold them in place, so that they are not in contact with each other, but there are no gaps in wavelength coverage.

In addition to the layers of graded thickness for variation of the peak wavelength, some layer deposition is needed to block spectral ranges outside the CVF region of interest. These layers are usually made of both graded and constant thickness over the whole substrate, and have to be carefully designed in order to provide the necessary spectral blocking to avoid unwanted wavelengths and harmonics to reach the UUT (see examples of blocking spectra in the reference[1]).

Example of Near Infrared CVF of CI is shown in the next figure.
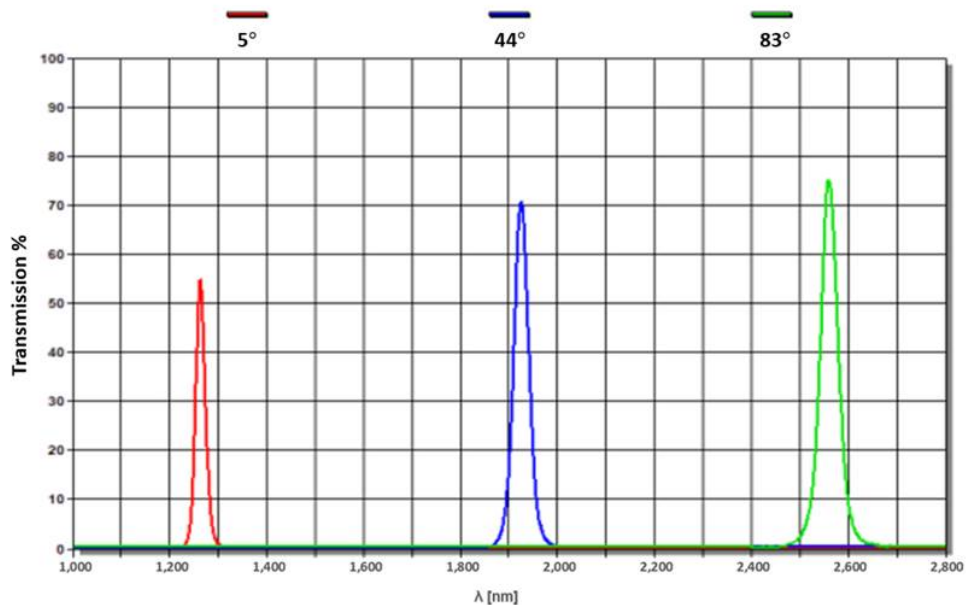
Figure 5: Example of Near Infrared CVF made of a 90^0 segment. The peaks are the transmittance spectra measured in three different wheel positions. The spectra shown are indicative of the CVF spectral resolution, typically being in the range of 2-2.5% of wavelength.

3.2 Continuous wavelength detector spectral response tester

The system described in this section is appropriate for testing single detectors or detector array engines, meaning detector arrays with digital signal output, such as a frame grabber. The basic feature of a single detector test system is that the projected radiation must be focused on a point, i.e. a region of small area, since the detector itself is a chip of usually a few microns or tens of microns on the side. The system of figure 6 represents just such a system.
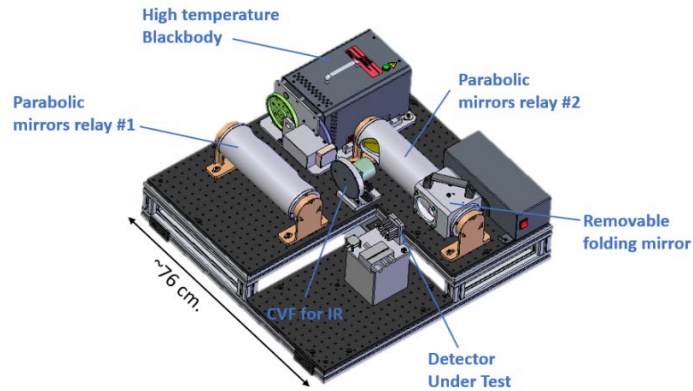
Figure 6. Detector array test station. The system is designed to allow testing of detector array pixels, and extract their spectral response by having the scanning wavelength CVF on an intermediate focal plane.

Figure 7 shows the optical layout of such a system.
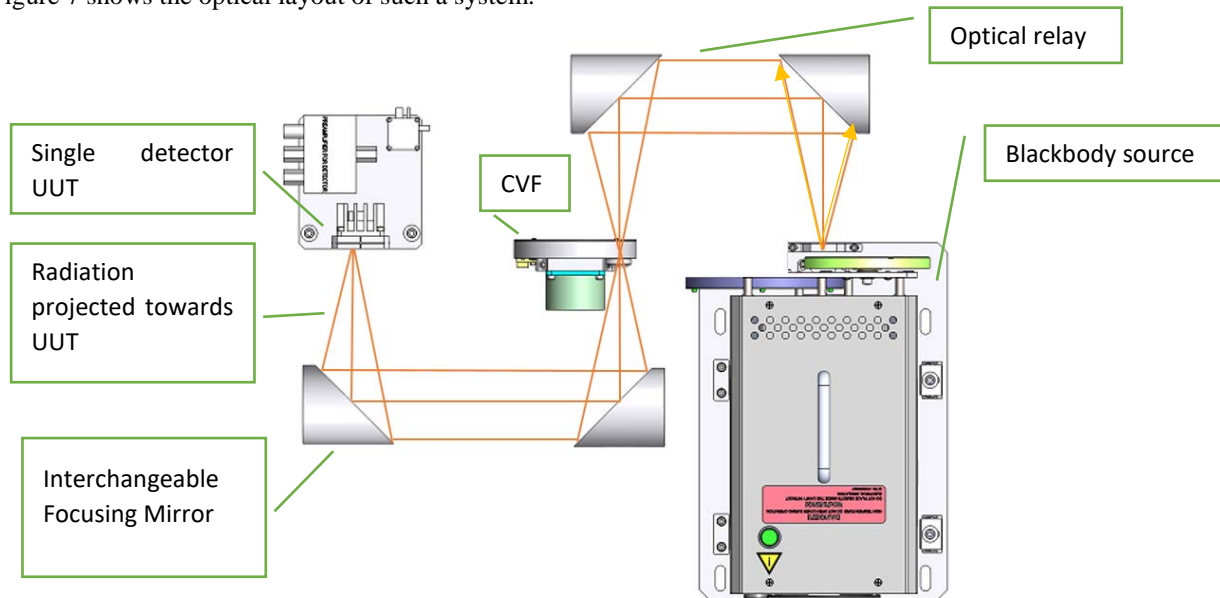
Figure 7. Optical layout of a single detector array spectral tester. The radiation form the source is first focused by the relay onto the CVF monochromator and then by a second relay with an interchangeable focusing mirror onto the single point detector.

If the detector is not a single one, but a detector array, the Interchangeable Focusing Mirror can be replaced by a flat mirror, and the radiation is projected onto the detector without being focused, in order to illuminate the whole array surface at once. Referring to figure 8, the radiation from the source is initially focused on the CVF for spectral selection, and then is projected onto the detector engine to be tested. As the CVF is scanned, the data from the detector are recorded by a computer through a frame grabber, and then used to calculate spectral response, uniformity, etc.
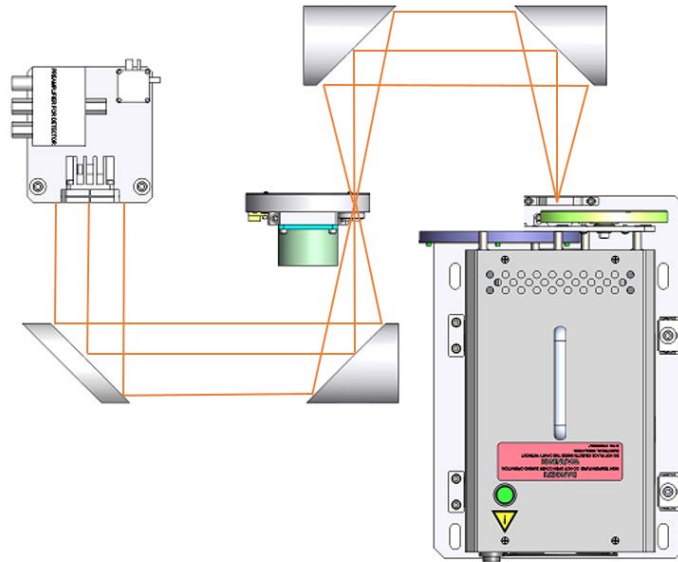
Figure 8. Optical layout of an array detector engine spectral tester. Note that the mirror in front of the detector is a flat folding mirror: this is to flood the detector with monochromatic radiation, of usually a few millimeter size, with the light from the source.

3.3 Large METS system for spectral camera response measurements (with CVF), without target wheel

The next tester type is a larger system. In many applications the UUT has a large objective optics, of the order of tens of centimeters, and the testing requires similarly large projection optics clear apertures. For this purpose CI has developed several large collimator products (called METS, or Modular Electro-optical Test Systems) with primary mirrors between 8" and 25" diameters, and F numbers between 5 and 10. Some applications of these large systems require the spectral information of response, MTF, etc. as in the small systems, except that now the size of most components scales up accordingly.

1 meter
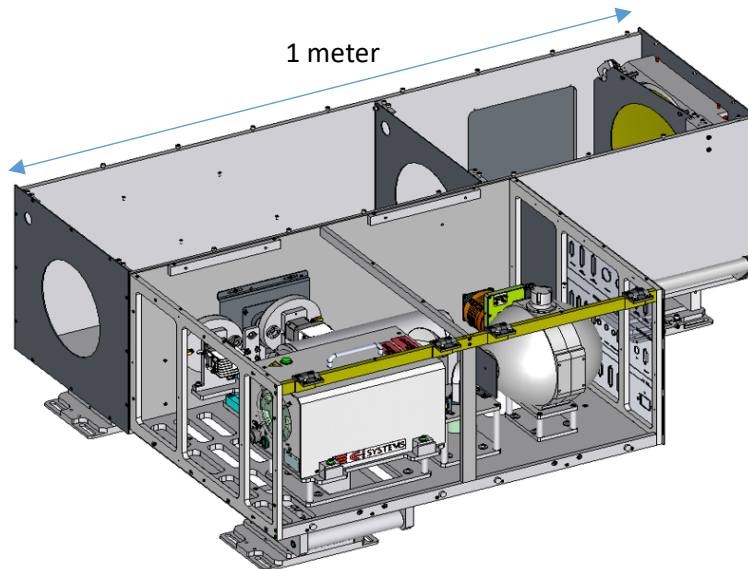
Figure 9. A typical METS based spectral test system is shown here. In this case the collimator is CI's METS8-VS, 8" clear aperture with 40" focal length, off-axis parabolic mirror, and IR & VIS sources with 2 matching CVF's covering the spectral range 0.4 – 14 um. The CVF's are located directly on the collimator's focal plane. In this version there are no targets for spatial resolution measurements.
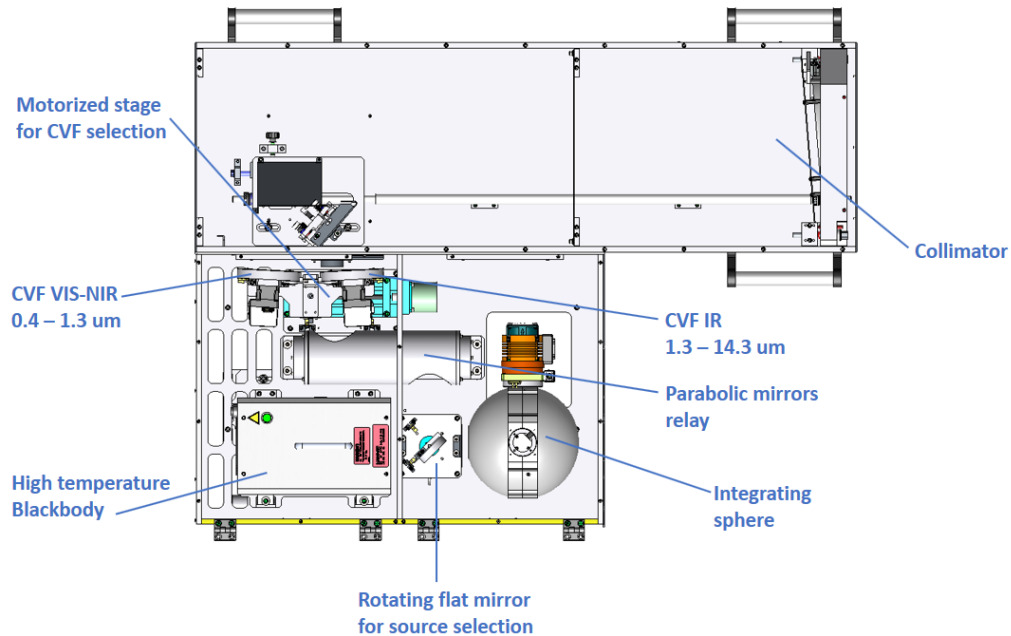
Figure 10 shows a top view of such system, showing all its elements. The rotating flat mirror switches between the blackbody source and the integrating sphere. The two CVF's are on a motorized stage for alternate use. The relay images the sources' apertures onto the CVF plane.
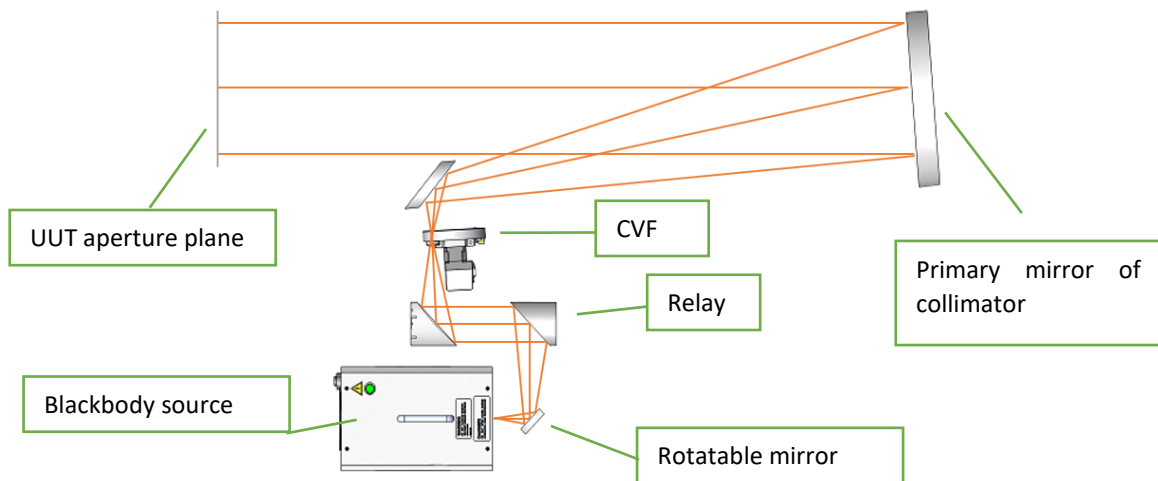

Figure 11. Optical diagram (partial) of the system of figure 10. By rotating the "Rotatable mirror" by 180 degrees the Integrating sphere of figure 10 (not shown here) is brought to act as the source, and the appropriate VIS/NIR CVF (not shown here) is automatically positioned in place of the infrared CVF, by the computer controlled motorized stage.

3.4 Large METS system for spectral / spatial resolution measurements (with CVF) and target wheel with patterns

The next tester type is a large METS based system, like in figure 10, but with additional target wheel for spectral-spatial resolution measurements. In this case a complete optical diagram is three-dimensional, so we show here only the general system view, and its optical diagram, after being collapsed into two dimensions.
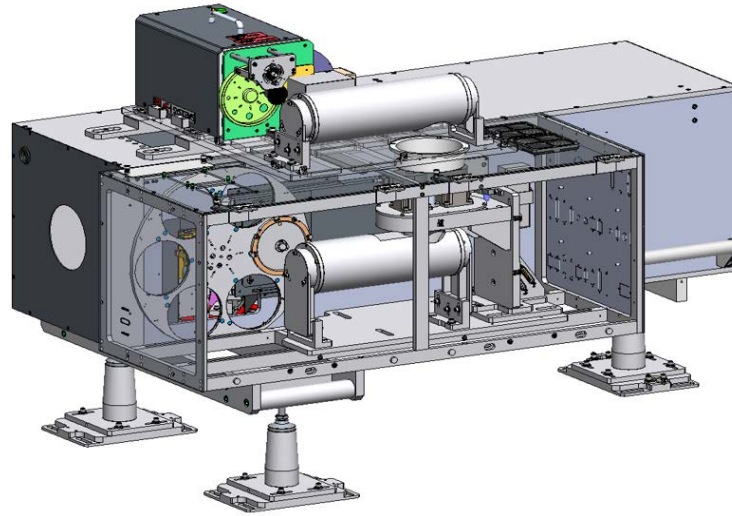
Figure 12. The largest and most complex spectral/spatial camera tester is shown here. The collimator is a CI's METS8-VS, 8" clear aperture with 40" focal length and off-axis parabolic mirror. The monochromatic beam, after passing through the CVF, is imaged on a standard target wheel, with a magnification of ~2.5, enabling projection of targets with up to 5 mm size.
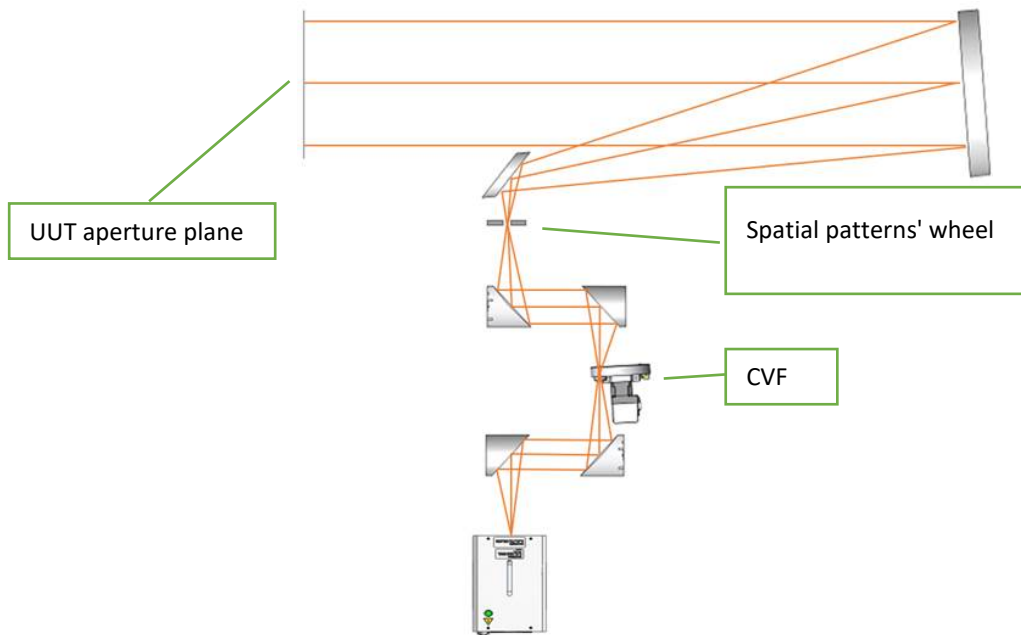

UUT aperture plane

Spatial patterns' wheel

CVF

Figure 13. Ray trace diagram of the system of figure 12, collapsed into two dimensions.
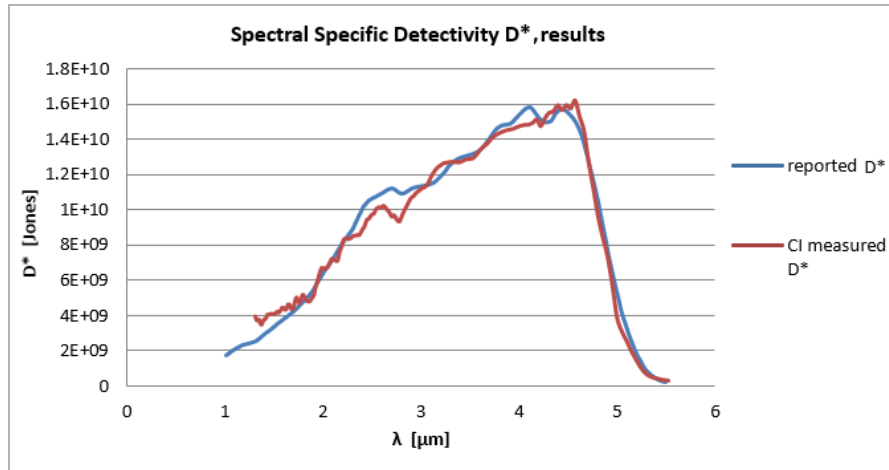
4. EXAMPLES OF SPECTRAL UUT MEASUREMENTS

Figure 14. Example of specific detectivity D* measurement of single detector, compared with the D* data provided by the manufacturer.
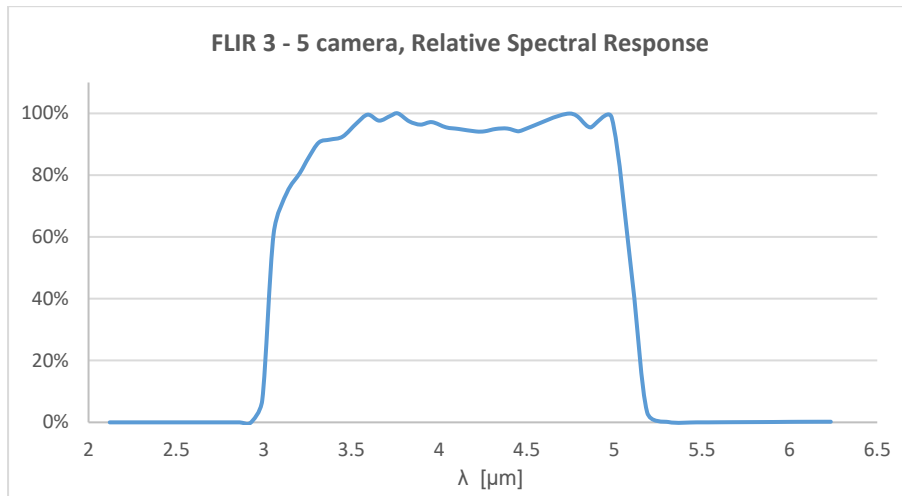

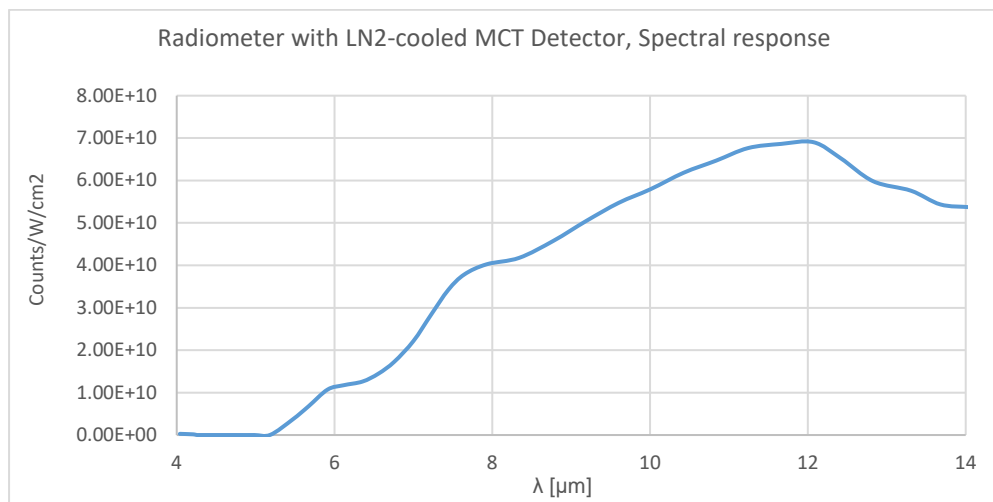Figure 15. Example of relative spectral response of a 3-5 micron camera, measured with our system.

Figure 16. Example of the spectral response of a radiometer with cooled MCT detector, measured with our system.

5. CONCLUSION

CI has developed a family of systems that are used to measure sensors' spectral and spatial parameters. The tests that can be done with such systems are separately their spectral response and spatial resolution, and combined into spectral dependence of spatial resolution.

REFERENCES

[1] Dario Cabib, Moshe Lavi and Henry Orr, Revival of Circular Variable Filters, Proc. of SPIE Vol. 7835, 783500O-1, 2010.

[2] William L. Wolfe and George J. Zissis, "The Infrared Handbook, page 7-118", Published by the Office of Naval Research, Department of the Navy, Washington DC, for a description of the CVF and pages 20-1 and following for a thorough treatment of radiometers of different kinds, (1978).

The design method of off-axis reflective stabilized zoom system based on deformable mirrors

Yuxuan Tian[a], Xuemin Cheng*[a], Qun Hao[b]

[a]Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055, China;
[b]School of Optics and Photonics, Beijing Institute of Technology, Beijing, 100081, China
*Corresponding author:chengxm@sz.tsinghua.edu.cn

ABSTRACT

Optical zoom system plays an important role application in high precision opto-electronic imaging for the high-speed moving target imaging. The novel stabilized zoom stabilization system based on deformable mirrors(DMs) has a future application prospect in integrating fast doubling, accurate focusing, image plane stabilization and aberration compensation. In this paper, we design a reflective zoom structure including the front group and the rear group. The double DMs and the reflective mirrors form the Combined Telephoto and Reversed-Telephoto structure, which achieve the zoom of the system and correct the off-axis aberration during the whole focal length. The fixed reflective mirror in rear group is used to compress the optical length and keep the image stabilized. We make full use of effective diameter range of DMs and the flexible deformations amount of actuator strokes to achieve the freeform surfaces in the system. It achieves a high zoom ratio of 14.52 and 5 milliseconds of zooming time from Wide-angle and Telephoto. This optical system is conducive to further achieving high zoom efficiency and high speed in the stabilized zoom system based on DMs.

Keywords: Deformable mirrors, zoom system, high zoom ratio

1. INTRODUCTION

The zoom systems can adjust the magnification of the imaging object, realizing the range of short focus at wide-angle and high resolution at telephoto while keeping the image plane stabilized. They play an important role in consumer electronics, digital cameras, surveillance cameras, biological detection, astronomical observation, reconnaissance systems and remote sensing systems and other fields[1-3]. The conventional mechanical zoom systems mainly consist of the zoom group, the compensation group and the fixed group, and they achieve a constant image plane position while zooming through the precise movement under the zoom group and the compensation group. However, the novel zoom systems require the unconventional new optical elements into the system, which are used as focal length variable devices as group elements to achieve zoom and image plane compensation function. The stabilized zoom systems based on DMs change the conjugate position of the object and image by deformation of device surface shapes, realizing elements position stabilized[4].

In 2007, researchers designed a reflective stabilized zoom system including two 59-channels OKO MDMMs and three fixed spherical mirrors.The surface shapes of the DMs were actuated and the system achieved a magnification increase of about 3.8 times[5]. In 2009, Krist of built an off-axis reflective four-mirrors stabilized zoom system by changing the surface shape of fixed mirrors[6]. The system achieved a high image quality and the focal length changes from 5.2mm to 15.6 mm. In 2014, Huang YW designed a 2x compact camera zoom module based on four sets of fixed lenses and double anamorphic mirrors[7], and the zoom ratio and the image quality should be further improved. In 2022, Tian proposed a catadioptric zoom design method based on transmission fixed lenses and reflection DMs and realized a high zoom ratio of 10. It is necessary to improve the zoom ratio and performance of reflective system based on DMs[8-9].

In this paper, we propose a design method of a reflective stabilized zoom system based on DMs. The front stabilized zoom system is designed as the Combined Telephoto and Reversed-Telephoto structure, and the fixed reflective mirror is used to control the optical length and realize the image stabilized. And we make full use of the flexible deformation in

the diameter of the aperture which the surfaces are actuated as freeform surfaces to correct the aberrations. The system achieves a high zoom ratio of 14.52 under the maximum center deformation of 10μm and the zoom time is measured 5ms from Wide-angle to Telephoto, which effectively improves the zoom ratio and the image speed of the system. This optical design method is helpful to achieve high zoom efficiency and high speed in the whole zoom focal length.

2. DESIGN METHOD

2.1 The Gaussian structure of stabilized zoom system

The optical system can be calculated as the thin lenses in Gaussian theory. The front stabilized zoom system has four elements. Based on the imaging model, the stabilized zoom equation is constructed and its Gaussian solution characteristics are analyzed:

$$Z = \left[\phi_1, -d_1, \phi_{DM_1}, -d_2, \phi_3, -d_3, \phi_{DM_2}, -d_4 \right] = 0 \tag{1}$$

Here, the ϕ_1 and ϕ_3 are fixed mirror, the ϕ_{DM_1} and ϕ_{DM_2} are DMs. And the d_i is the distance of the elements of ith to $(i+1)^{th}$. The relationship of the optical power change of the DM and the system are expressed as:

$$\Delta\phi_{DM_1} = \frac{d_4}{{}^1A_2 \, {}^2B_4} \Delta\Phi \tag{2}$$

$$\Delta\phi_{DM_2} = \frac{{}^1A_2}{d_4 \cdot {}^2B_4} \Delta\frac{1}{\Phi} \tag{3}$$

The focal length of the zoom system and the back focal length can be expressed :

$$f = \frac{1}{\Phi} = \frac{1}{{}^1C_4} = \frac{1}{\left[\phi_1, -d_1, \phi_{DM_1}, -d_2, \phi_3, -d_3, \phi_{DM_2} \right]} \tag{4}$$

$$d_4 = \frac{{}^1A_2}{\Phi_l} \sqrt{\frac{\Delta\phi_{DM_1}}{(\Gamma-1)\cdot\Delta\phi_{DM_2}}} \tag{5}$$

The Φ_l is the optical power of the Telephoto, and the Γ is the zoom ratio. When the first fixed mirror $\phi_1 = 0$, the DM1 is the first mirror, fixed mirror and DM2 are composed of the front stabilized zoom system. And we design the system structure as Telephoto system in the Wide-angle and Medium-angle and the Reversed-Telephoto system in the Telephoto.

2.2 The off-axis aberration in stabilized zoom system

The wave aberration of an off-axis optical system with surface decenter and tilt component is expressed as:

$$W = \sum_j W_{040j} (\vec{\rho}\cdot\vec{\rho})^2 + \sum_j W_{131j} \left[\left(\vec{H} - \vec{\sigma}_j \right)\cdot\vec{\rho} \right](\vec{\rho}\cdot\vec{\rho}) + \sum_j W_{222j} \left[\left(\vec{H} - \vec{\sigma}_j \right)\vec{\rho} \right]^2$$
$$+ \sum_j W_{220j} \left[\left(\vec{H} - \vec{\sigma}_j \right)\cdot\left(\vec{H} - \vec{\sigma}_j \right) \right](\vec{\rho}\cdot\vec{\rho}) + \sum_j W_{311j} \left[\left(\vec{H} - \vec{\sigma}_j \right)\cdot\left(\vec{H} - \vec{\sigma}_j \right) \right]\left(\left(\vec{H} - \vec{\sigma}_j \right)\vec{\rho} \right) \tag{6}$$

And the field vector of decenter[10-12] can be calculated as:

$$\vec{\sigma}_j = (0, \vec{\sigma}_{yj}^{sph}) = (0, \frac{-SRM_j}{\overline{u}_j + \overline{h}_j c_j}) = (0, \frac{-\sin(\alpha_j)}{\overline{u}_j + \overline{h}_j c_j}) \qquad (7)$$

$\overline{u}_i$ donates the principle ray angles, c_i donates the curvature, $\overline{h}_i$ donates the principle ray height. According to the eq.(6), we can get the the wave aberration of 3rd-order for off-axis optical system.

2.3 The global optimization design of optical stabilized zoom system

An evaluation function is used to describe the comprehensive description of the off-axis fixed zoom system based on deformable mirrors. This evaluation function describes the zero focal length and back focal length in the whole focus, and minimizes the third-order aberration items in the full field aberration distribution in the stabilized zoom system. The optimization variables includes the curvature, distance and tilt angle. The global optimization algorithm is used to solve the optimal value of nonlinear evaluation function and the initial optical configuration can be solved in Matlab. The final high-order freeform surfaces are optimized in CodeV.

3.   DESIGN AND RESULTS

We design an unobstructed reflective stabilized zoom system based on DMs, the Primary Mirror and the Tertiary Mirror are deformable mirror, and the Secondary Mirror and Fourth Mirror are fixed mirror. In the front stabilized zoom system, the DM_1 realizes the image zoom and the deformation of DM_2 compensates the image drift. We design the front stabilized zoom system as the Combined Telephoto and Reversed-Telephoto stabilized-zoom structure to achieve the zoom and correct the off-axis aberration during the whole focus. And the FM is spherical mirror which is used to compress the optical length and keep the image stabilized. The system focal length is 250mm-3630mm. The stop is on the SM which is used to expand the field angle.

The deformation of DM realizes the continuous zoom during the whole focal focus. In order to compress the optical length, we use the plane mirror to build the fold-reflective experimental optical layouts. The four-mirror optical layout is showed in Fig.1(a). In order to compress the optical length, we use the plane mirror to build the fold-reflective optical layouts in CODEV. And the experimental optical layouts of the stabilized zoom system is showed in Fig.1(b). And the surfaces are designed as the freeform surfaces to correct the off-axis non-rotational symmetric aberration.
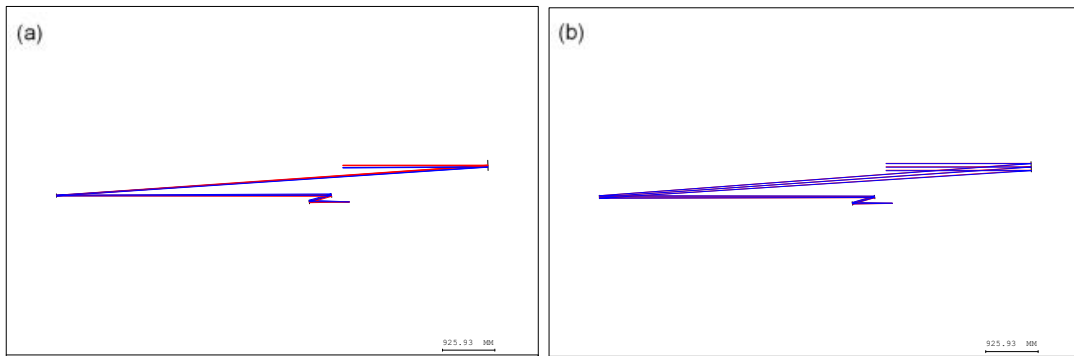

Figure 1. (a)The optical layout of the stabilized zoom system. (b)The optical layout of the fold-reflective experimental system.

There are two DMs in the system and the DM_1 (MMDM) is customized with Vision of Russian company in Fig.2(a), it has 52 channels and the aperture is 50mm which the maximum center deformation is 10μm . The DM2(PDM) is

from OKO Technologies in Fig.2(b), its has 37 channels and the aperture is 30mm which the maximum center deformation is 8μm . We use the 70% of the aperture of the deformable mirror which is benefical for the accuracy of actuation. And the surface is measured by Zygo Interferometer in Fig.2(c).

In Fig.3 the collimator is used to simulate the infinite objects, the high speed camera is used to get the image from the resolving test target from collimator. The camera can get 1000 pictures in 1 second, which is used to record the process of the zoom. And the stabilized zoom system achieve the image magnification without the movement of optical group.
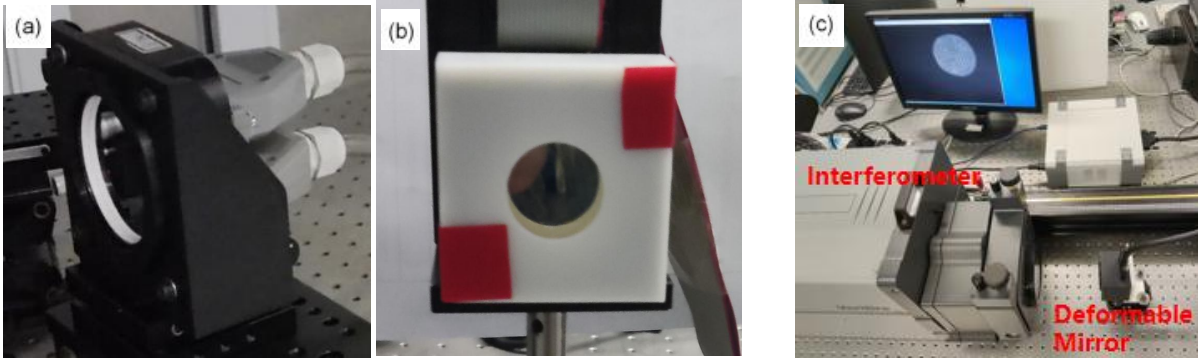

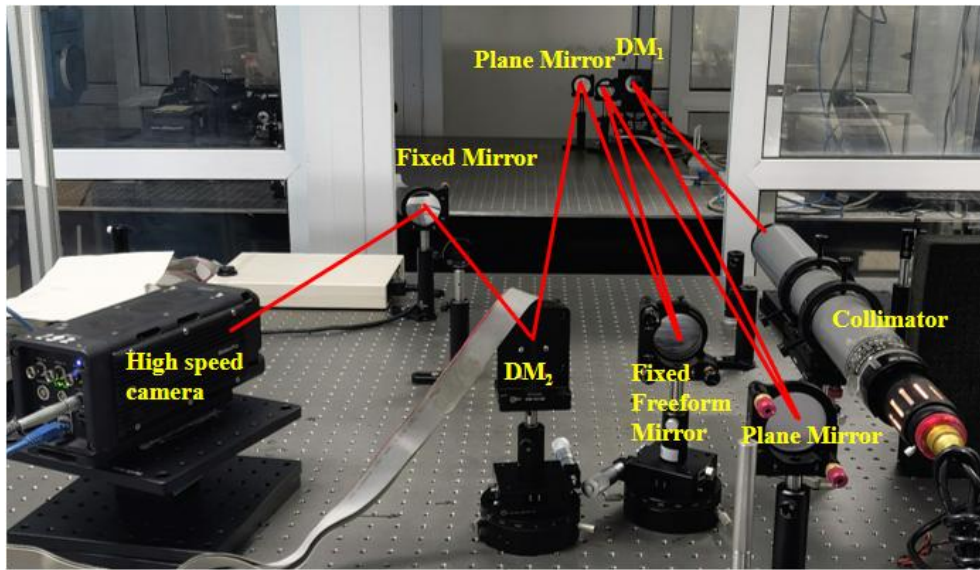Figure 2. (a)MMDM (b)PDM   (c)The test experimental system of the DMs .

Figure 3. The optical layout of the fold-reflective experimental system.
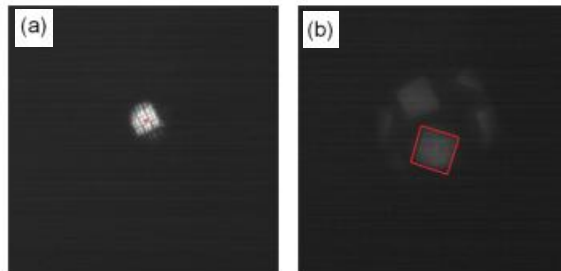

Figure 4. The optical layout of the fold-reflective experimental system.

In Fig.4, the picture of wide-angle and is the picture of Telephoto. The image intensity in Wide-angle is larger than Telephoto. We calculate the Target boundary length 1-4 and the average of the boundary length of the Telephoto and Wide-angle is the zoom ratio. The zoom ratio of stabilized zoom system based on two DMs is 14.52.

Table 1. The target boundary length of the resolving test target .

	Target boundary length 1	Target boundary length 2	Target boundary length 3	Target boundary length 4	Average boundary length
Wide-angle	10.066	9.5038	10.3463	9.7938	9.9275
Telephoto	141.8929	143.6562	143.1938	147.9345	144.1693

The Fig.5 show the 25 pictures of the high speed camera. They are obtained by camera in 25 milliseconds. We can see from the image Y4 Camera000132 of the Wide-angle. From the start of the zoom, the image is Y4 Camera000137. And the system switches to the Telephoto is in the picture of Y4 Camera000143. The time of zooming is 5 milliseconds.
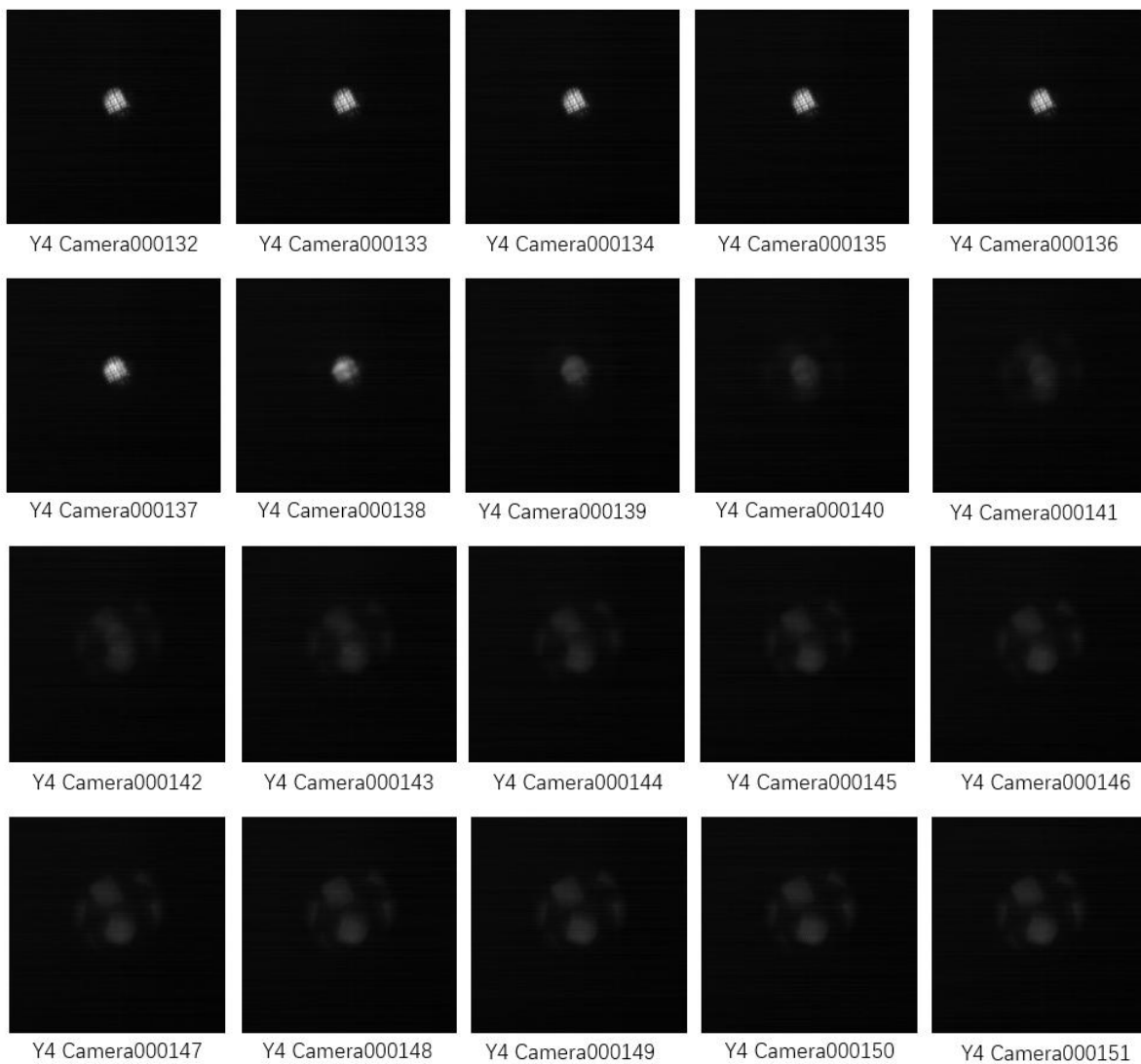

Figure 5. The optical layout of the fold-reflective experimental system.

4. CONCLUSION

In this paper, we propose a design method of a reflective stabilized zoom system based on deformable mirrors. The front optical structure is designed as the Combined Telephoto and Reversed-Telephoto structure to get the high efficiency zoom. And the rear group achieves the shorter optical length and decrease of image plane drift. And we make full use of the flexible deformation in the diameters of the aperture which the surfaces are actuated as freeforms to correct the aberration. The stabilized zoom system achieves a high zoom ratio of 14.52 under the maximum center deformation of 10μm and the zoom time is measured 5ms from Wide-angle to Telephoto, which effectively improves the zoom ratio and the image speed of the system. This optical system is helpful to achieve high zoom efficiency and high speed in the whole zoom focal length.

ACKNOWLEDGEMENT

REFERENCES

[1] Turner G. L. The history of optical instruments. A brief survey of sources and modern studies[J]. History of science; an annual review of literature, research and teaching,1969, 8(1):53.

[2] Clark A. D. Zoom lenses[M]. Monographs in Applied Optics. No.7. London: Adam Hilger Ltd. 1973:3-11.

[3] Hao Q , Cheng X , Du K . Four-group stabilized zoom lens design of two focal-length-variable elements[J]. Optics Express, 2013, 21(6):7758-7767.

[4] Brett E. Bagwell, David V. Wick, William D. Cowan, Olga Blum Spahn, William C. Sweatt, Ty Martinez, Sergio R. Restaino, Jonathan R. Andrews, Christopher C. Wilcox, Don M. Payne, and Robert Romeo "Active zoom imaging for operationally responsive space", Proc. SPIE 6467, MEMS Adaptive Optics, 64670D (14 February 2007); https://doi.org/10.1117/12.706820

[5] Heng Li, Xuemin, et al. An Electrically Tunable Zoom System Using Liquid Lenses[J]. Sensors, 2015.

[6] Kristof Seidl, Jens Knobbe, and Heinrich Grüger, "Design of an all-reflective unobscured optical-power zoom objective," Appl. Opt. 48, 4097-4107 (2009)

[7] Yu-Hung Huang, Hsiang-Chun Wei, Wei-Yao Hsu, Yuan-Chieh Cheng, and Guo-Dung John Su, "Optical zoom camera module using two poly-dimethylsiloxane deformable mirrors," Appl. Opt. 53, H248-H256 (2014)

[8] Xuemin Cheng, Hengzhi Ye, and Qun Hao, "Synthetic system design method for off-axis stabilized zoom systems with a high zoom ratio," Opt. Express 29, 10592-10612 (2021)

[9] Yuxuan Tian, Jindong Wang, Hengzhi Ye, Xuemin Cheng, Qun Hao, "The design of catadioptric-stabilized zoom system based on deformable mirrors," Proc. SPIE 12315, Optical Design and Testing XII, 123150D (20 December 2022); https://doi.org/10.1117/12.2646357

[10] R. A. Buchroeder, "Tilted component optical systems," Ph.D. dissertation (University of Arizona, 1976).

[11] R. V. Shack and K. Thompson, "Influence of alignment errors of a telescope system on its aberration field," Proc. SPIE 251, 146–153 (1980).

[12] K. P. Thompson, "Aberration Fields in Tilted and Decentered Optical Systems," Ph.D. dissertation (University of Arizona, 1980).

Design and development of an integrated LiDAR sensor for autonomous driving

A. Hofbauer*[a], R. Reichert[a], M. Hennecke[b], M. Baumgart[c],
A. Tortschanoff[c], B. Kirillov[b], M. Pfennigbauer[a]

[a]RIEGL Research Forschungsges.m.b.H, Riedenburgstraße 48, 3580 Horn, Austria;
[b]Infineon Technologies Austria AG, Babenbergerstraße 10, 8020 Graz, Austria;
[c]Silicon Austria Labs GmbH, High Tech Campus Villach - Europastraße 12, 9524 Villach, Austria

ABSTRACT

Automotive LiDAR systems are expected to play a crucial role in the future development of autonomous driving. In the course of the Austrian research project iLIDS4SAM an appropriate LiDAR sensor demonstrator has been designed and developed.

Based upon typical requirements for such sensors, e.g. the capability to detect objects of about 10 cm x 13 cm size at a distance of 80 m, a field-of-view of 20° x 90° (V x H) and an image rate of about 17 Hz, a highly innovative 3D laser scanner has been designed which combines state-of-the-art MEMS mirror beam deflection with a rather classical polygon mirror wheel.

Integrating a laser diode array of the newest generation, a multipixel APD detector array, waveform digitization as well as online waveform processing, 16 range measurement channels operating simultaneously are realized. The resulting LiDAR sensor offers a range measurement rate of 4.5 million measurements per second, each with the capability to resolve multiple targets.

The LiDAR sensor is manufactured as a prototype on the level of an elegant breadboard.

This contribution provides insight into the design of the LiDAR sensor and discusses the challenges identified during the design and development phase.

Keywords: 3D laser scanner, Multichannel LiDAR, Autonomous driving

1. INTRODUCTION

The Austrian research project iLIDS4SAM – Integrated LiDAR Sensors for Safe & Smart Automated Mobility – is a flagship project for future automated mobility (www.ilids4sam.at, [1]). It addresses the requirements and challenges of LiDAR systems suitable for autonomous driving, initiates the development of technologies required for processing LiDAR point clouds, and also focuses on their integration with other sensors.

A main part of the project is the design and development of an integrated LiDAR sensor which is capable of acquiring a 3D point cloud of the environment approaching the vehicle, not only for obstacle detection and alerting the driver but also meeting the challenging requirements with respect to data quality and point density of autonomous driving in an urban environment.

*ahofbauer@rieglresearch.com; phone +43-2982-4211; www.riegl.com

2. DEFINITION OF REQUIREMENTS

In a first step, several basic target specifications of the LiDAR system had to be considered and finally defined:

- The very first basic specification was the minimum size of an obstacle that should be detectable based on the LiDAR point cloud. For this project a reasonable size of such obstacles was defined to be 10 cm x 10 cm, which enables the detection of some typical children's toys like a ball or the seat of an overturned bicycle, but also to recognize a person lying flat on the road.
  In a later design phase, the minimum size has been slightly increased to 10 cm x 13.1 cm (V x H).

- The second basic specification was the distance at which the above-mentioned spatial resolution shall be met. Based on typical speed limits in urban environment of up to typ. 50 km/h, this basic distance was defined to be 80 m.
  A distance of 80 m would in principle allow a significantly higher speed of up to 75 km/h. However, the selected distance at only 50 km/h reserves some time for data acquisition and obstacle perception, i.e. transferring and processing the LiDAR data to recognize and identify an obstacle, but also for activating the vehicle's brakes.

- The third and fourth basic specifications were the horizontal and vertical fields of view (FOV). A vertical FOV of ±10° and a horizontal FOV of ±45° have been considered reasonable.

- The fifth basic specification was the frame rate, i.e. the number of points clouds of the vehicle's environment to be provided per second. A reasonable specification was found to be in the range of 10 Hz to 25 Hz, finally settling at the rather low image rate of early cinema films of about 16 2/3 frames per second (fps).
  This frame rate may be considered somewhat high in comparison to the time reserved for obstacle perception. However, a higher frame rate allows the collection of some possibly redundant information which can be used for consistency checks as well as sensor self-testing.

Based upon these basic specifications the following resulting requirements were found:

- The intended angular resolution is about 1.25 mrad x 1.64 mrad, equal to 0.0716° x 0.0938° (V x H).

- The number of single range measurements per frame is about 280 x 960 (V x H), thus about 269,000 measurement points per frame can be expected.

- Applying the defined frame rate, the rate of single range measurements is about 4.5 million measurements per second.

- The rate of range measurement results may be even slightly higher due to the detection of multiple targets for specific laser directions, however it is not expected to exceed ever $5 \cdot 10^6$ /sec.
  As the laser beam may be directed towards the sky for up to 50% of the vertical FOV, the expected rate of range results may be significantly lower than the numbers estimated above.

3. LIDAR SYSTEM DESIGN

3.1 Design challenge

To the best of the authors' knowledge no LiDAR system of such performance, i.e. delivering the huge amount of 269,000 range measurements per frame, evenly distributed over a FOV of ±10° x ±45° and at a frame rate of 16 Hz, had been published up to the year 2020, i.e. when this design was developed.

Due to the time limitation of the project the entire design had to rely on off-the-shelf components rather than components which would have to be designed for project specific requirements.

3.2 Selected solution

Several known designs and combinations of scan mechanisms have been simulated and analyzed for their performance. This investigation led to a highly innovative 3D LiDAR scanner concept combining:

- receiver channels with the full vertical FOV of ±10°;

- a MEMS mirror for the vertical scan of the laser beam, allowing a vertical scan frequency of about 2kHz;

- the usage of a - somewhat conventional - polygon mirror wheel for the horizontal scan mechanism for both, transmitter and receiver channel;

- 16 receiver channels and a fan-shaped laser beam operating all 16 channels simultaneously;

- simultaneous waveform digitization as well as online waveform processing for all 16 range measurement channels, each channel with the capability to resolve multiple targets per laser shot.

3.3 Detailed description

Receiver channels

Each receiver channel offers an instantaneous FOV of ±10° x 1.64 mrad (V x H). The aspect ratio is about 213:1 and thus extremely high. 16 receiver channels are arranged side by side, together covering a FOV of ±10° x 26.2 mrad equal to ±10° x 1.50° (V x H).
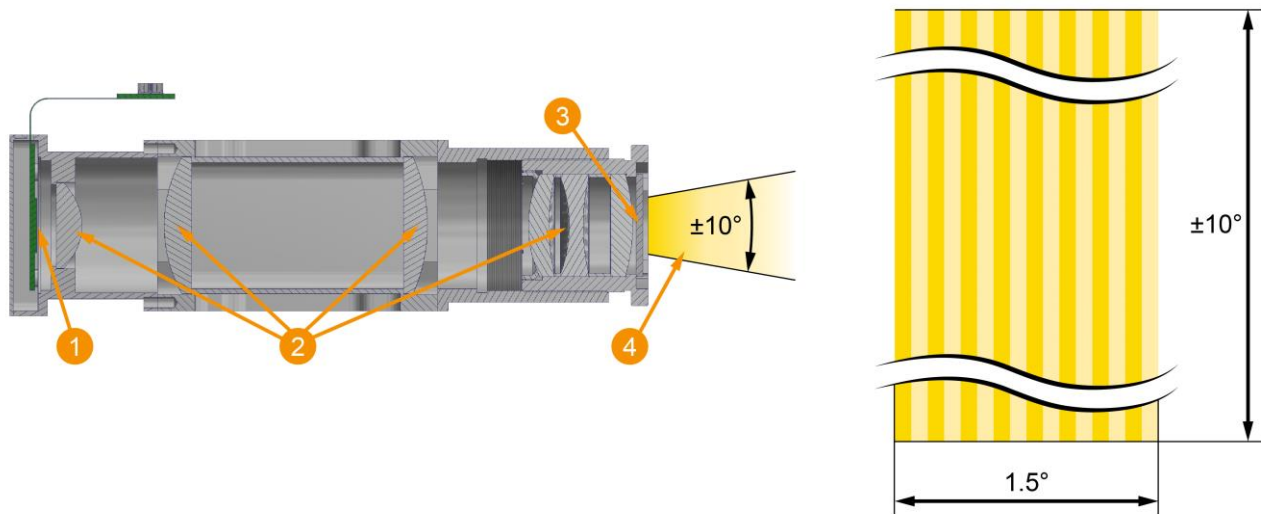

Figure 1.
    Left image: Cross section of the optical receiver assembly with APD array (1), lenses (2), optical band filter (3), and resulting vertical FOV (4);
    Right image: the instantaneous FOVs of the 16 receiver channels, arranged side by side, together covering a FOV of 20° x 1.5°

A main challenge of the receiver optical system design was the adaptation of the required extremely high aspect ratio of the FOV of a single LiDAR channel to the significantly lower aspect ratio of the corresponding APD pixel.

Laser transmitter

The laser transmitter assembly is of high complexity as it realizes not only the shaping of the laser diode array's raw beam to a laser fan with a divergence of 1.25mrad x 26.2mrad (V x H) but integrates also the MEMS mirror for the

vertical scan of the laser fan of ±10°. Figure 2 gives some insight into the internal design of the assembly. A detailed description of the optical setup and its properties has been published in [2].

The integrated MEMS mirror offers a usable diameter of 5mm and an angular amplitude of up to ±10°, thus being capable of covering a beam angular range of ±20°. Due to the design of the transmitter optics the resulting beam deflection is in the range of ±10°, as required. The MEMS mirror is operated at its resonance frequency of nominally 2 kHz.
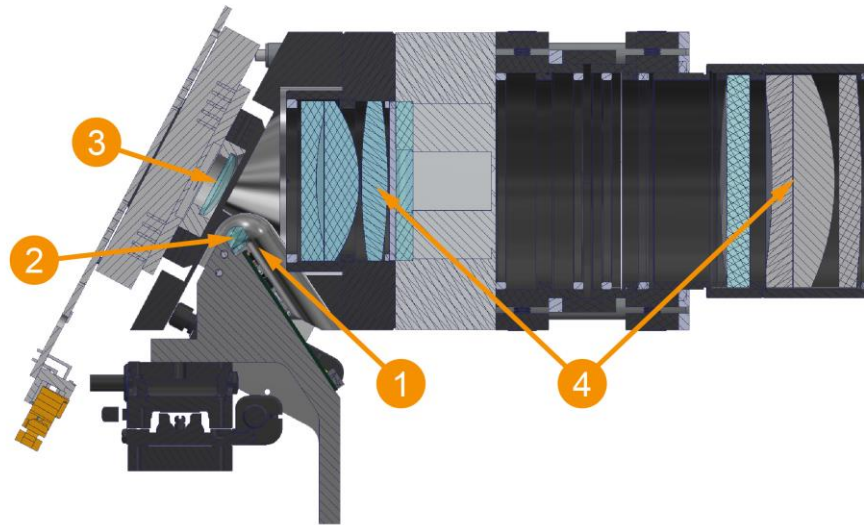

Figure 2. Cross section of the laser transmitter assembly with the laserdiode array (1), a first group of lenses (2), the oscillating MEMS-mirror (3), and a second and third group of lenses (4)

The laser source is an array of edge-emitting laser diodes with a center wavelength is 905nm. The array offers an optical peak power of about 600W at a pulse width (FWHM) of 6.7ns. It is driven by GaN FETs. A detailed description of the laser driver electronics and its properties has been published in [3].

Frame scan - polygon mirror wheel

Due to the specific design of the LiDAR scanner system with

- a given angular resolution,
- a fixed MEMS resonance frequency (defining the line scan rate),
- a typical usage of 50% of the angular range of a polygon mirror wheel, and
- a given diameter of the receiver´s optical channel (required for achieving the intended maximum range)

the selection of the number of facets of the polygon mirror wheel is a trade-off between

- the horizontal FOV,
- the framerate, and
- the diameter of the polygon mirror wheel.

Calculations for several numbers of facets have been accomplished; the results are listed in Table 1. Applying the limitations given above, a polygon mirror wheel of 4 facets has been found to be of smallest size.

Table 1. Properties depending on the number of facets of the polygon mirror wheel

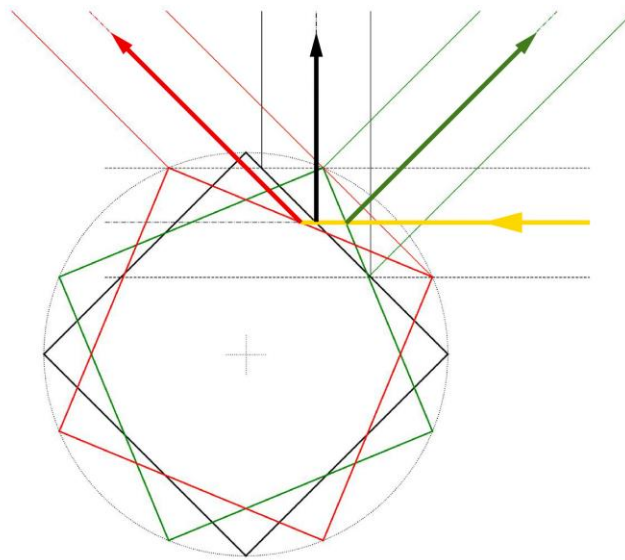| Number of facets	Horizontal FOV	Frame rate	Diameter of polygon mirror wheel (diagonal)
3	120°	12.5 fps	191 mm
4	90°	16.7 fps	158 mm
5	72°	20.8 fps	196 mm
6	60°	25 fps	234 mm

Figure 3. Top view of the 4-facet polygon mirror wheel in those rotational positions which limit the horizontal FOV, i.e. -45° (green), 0° (black) and 45° (red) with the corresponding paths of the laser beam

3.4 Point pattern

One of the main goals of the scanner system design is to achieve an evenly distributed and - as far as possible - regular point pattern.

Vertical scan

The vertical scan of the laser transmitter is realized by a MEMS mirror which is operated at its resonance frequency (about 2 kHz). Thus, the MEMS mirror's tilt angle is very close to a sinusoidal shape. As a consequence thereof, the vertical angle of the resulting laser beam at the output of the transmitter assembly still follows almost a sine function. The resulting angular amplitude is already 10°, thus covering the vertical FOV of ±10°.

Note that the receiver does not need any vertical scan as its instantaneous vertical FOV is already ±10°.

Horizontal Scan

The horizontal scan of transmitter and receiver is realized by the above-mentioned 4-facet polygon mirror wheel. The main axes of the optical paths of transmitter and receiver are in parallel. Due to the arrangement of the mirror wheel's axis of rotation with respect to the laser beam axis (which is in parallel to the axis of the receiver's optical path), a

change of the wheel's angle causes a change of the reflected laser beam of doubled value; thus, changing the mirror's angle by 1° causes a change of the reflected beam direction of 2°.

Based on these the sinusoidal change of the vertical beam angle and the linear increase of the horizontal beam angle, the laser's beam axis moves as shown in Figure 4.
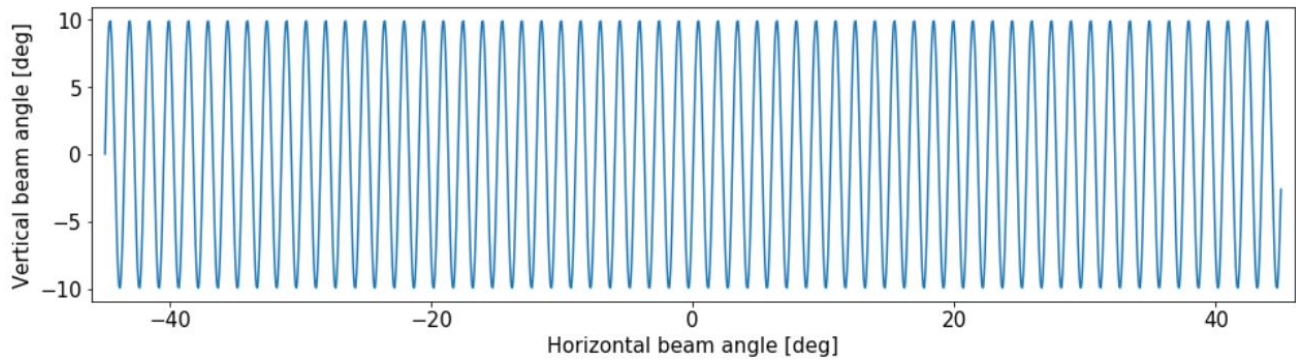

Figure 4. Path of the laser's beam axis in the plane of horizontal and vertical beam angle

The laser beam's footprint is similar to a horizontal fan with a divergence of 1.25mrad x 1.5° (V x H). This horizontal divergence is by design equal to the horizontal angular distance between two subsequent up-scans (or down-scans) of the MEMS mirror. When applying such a broad beam of 1.5° to Figure 4, the gaps between two neighboring up-scans but also between two subsequent down-scans are closed. This means the FOV is seamlessly covered without any gaps.

Laser triggering

To achieve an even point distribution over the entire vertical FOV, the laser needs to be triggered accordingly, i.e. at vertical laser beam angles with a constant step size in between. This requires also a constant step size between the corresponding tilt angles of the MEMS mirror. However, this even point distribution comes at the price of a strongly varying laser pulse repetition rate: the average value within a full period of the MEMS oscillation is about 560 kHz whereas the maximum value is even as high as 880 kHz. This maximum PRR occurs in the center of the vertical FOV when the MEMS as well as the laser beam are at their maximum angular speed.
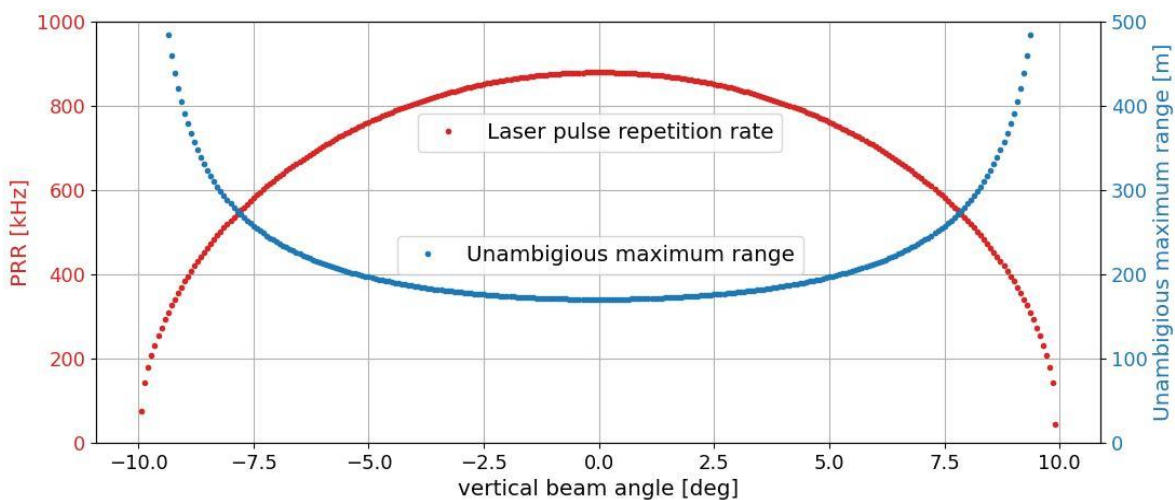

Figure 5. Characteristics of the laser PRR (red, left y-axis) with respect to the MEMS mirror's tilt angle
and the LiDAR channels' resulting unambiguous maximum range (blue, right y-axis)

One side effect of the rather high maximum PRR is that the unambiguous maximum range of the LiDAR channels is under worst case conditions limited to only 170 m.

Interleaving point sets

As explained above, the application of a laser fan of 1.5° horizontal divergence ensures that the entire vertical and horizontal FOV is covered without gaps. Even more, the entire FOV is covered not only once but twice per frame:

- once by the data points acquired during the up-scans and
- a second time by the data points acquired during the down-scans.

This feature of the scan mechanism is utilized to reduce the laser's pulse repetition rate by implementing an interleaving of the two sets of data points:

- during the up-scans the laser is triggered only at the even numbered vertical angular positions and
- during the down-scans the laser is triggered at the odd numbered vertical angular positions.
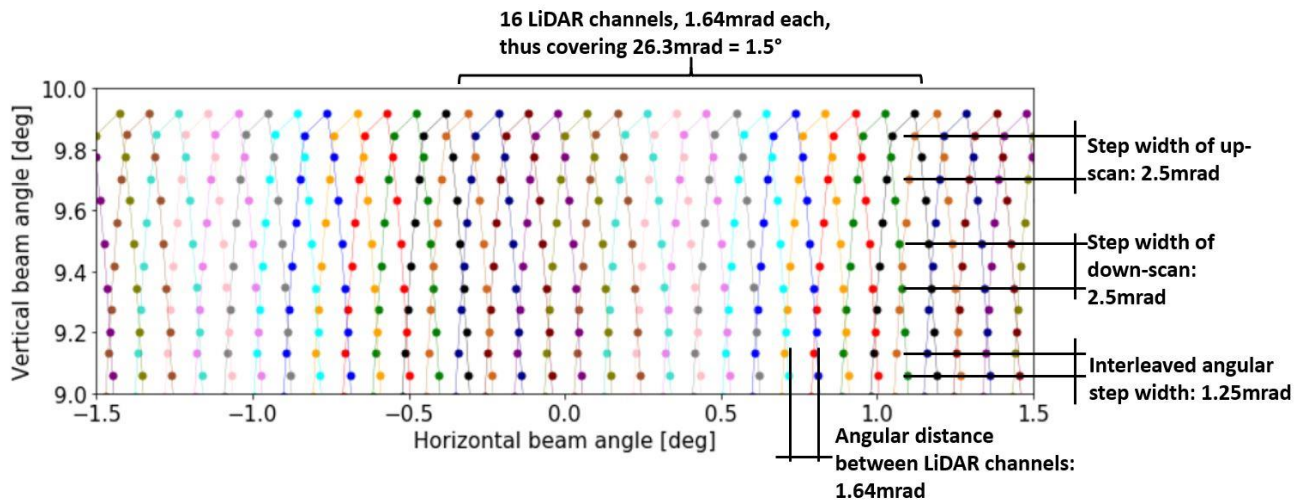
Figure 6. Zoom into the resulting point pattern of 16 LiDAR channels using one color per channel;
the colored bullets show the angular positions of the laser beams when the laser is triggered;
the dimensions highlight the equidistant and even distribution of the acquired data points

3.5 Multi-channel range measurement

Signal chain and digital signal processing

The signals within a LiDAR system are of different nature (Figure 7):

- the laser pulse emitted as well as the echo signals are optical (red);
- these signals are converted into analog electrical signals (blue);
- these still analog signals are conditioned appropriately and finally digitized (black).

For digitizing optical pulses width an FWHM-width of about 7 ns, the LiDAR system uses a sampling rate of 200 MSps. An appropriate anti-aliasing filter is applied.
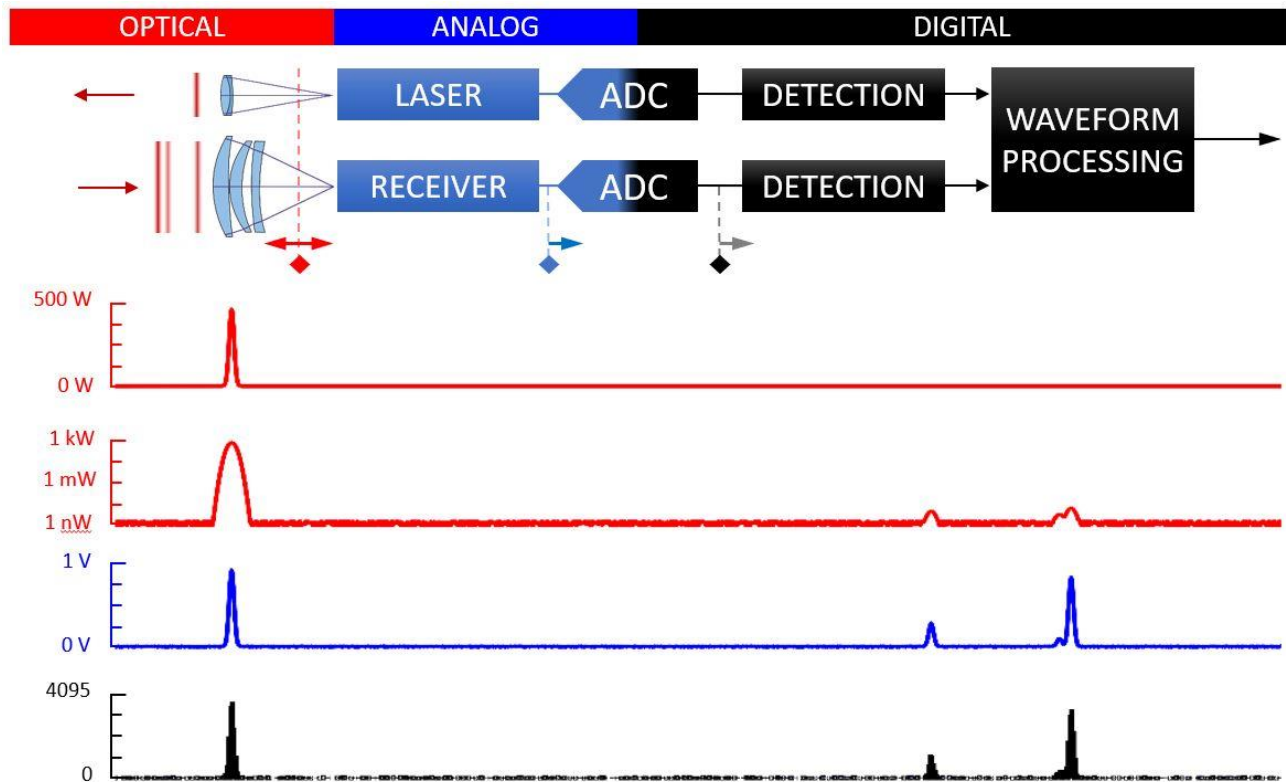
Figure 7. Signal chain applying early signal digitization and digital signal processing to the greatest possible extent

This early signal digitization allows the utilization of digital signal processing technology to the greatest possible extent:

- Detection of echo pulses

- Selection of the relevant samples of an echo pulse

- Waveform processing (Figure 8):

  - o The selected samples of the digitized pulse are compared to the system response waveforms to find the best fit.

  - o This best fit is used to estimate the signal strength / amplitude of the pulse and the temporal position to calculate the target´s range.
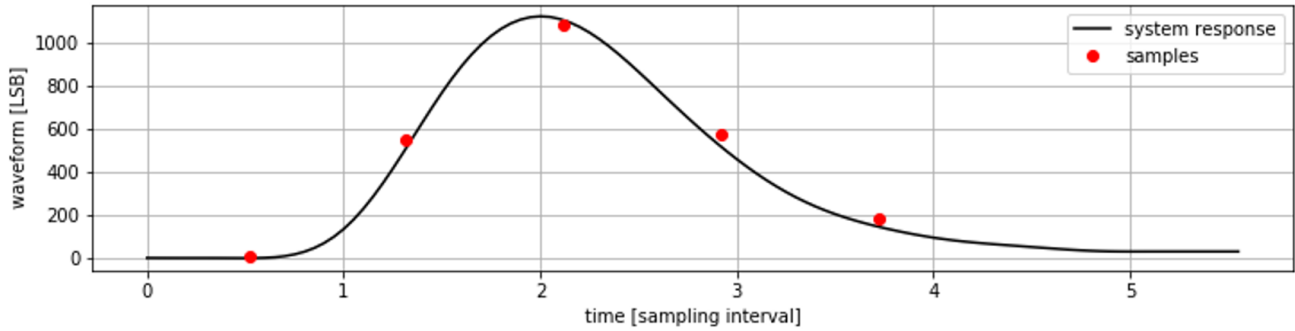
Figure 8. Principle of the applied waveform processing technology: the samples of the digitized
pulse are compared to system response waveforms to find the best fit

Range measurement accuracy

The application of the waveform processing method described above (see also[4]) enables the device to achieve a very good range measurement accuracy as well as a rather low range standard deviation of less than 2 cm over a wide amplitude range.

Figure 9 shows test results of a breadboard LiDAR system consisting of

- the receiver channel introduced in section 3.3,

- a single beam laser diode transmitter using a single lens transmitter optics and a driver electronics quite similar to that described in [3], and

- the digitization and data processing electronics developed in the course of project iLIDS4SAM.

This LiDAR system was directed towards a flat target at a fixed position and the amplitude of the echo signal arriving at the receiver was varied by inserting a neutral density filter of variable attenuation in front of the receiver. Many settings of the variable attenuator were chosen, and a set of distance measurement results was collected for each attenuator setting. Each single set was processed applying statistical methods to get the following results:

- range mean value

- range standard deviation

- minimum range

- maximum range

- amplitude mean value

The upper diagram of Figure 9 shows the mean value as well as minimum and maximum of the ranges vs. the amplitude mean value, and the lower diagram shows the standard deviation of the ranges vs. the amplitude mean value. Best range measurement accuracy (smallest span between minimum and maximum range) as well as the lowest range standard deviation are achieved at echo signal amplitudes of 10 dB to 25 dB. At lower amplitudes the accuracy gets worse due to a low SNR. At high amplitudes (> 30 dB) the receiver's signal chain limits the achievable accuracy due to compression effects.
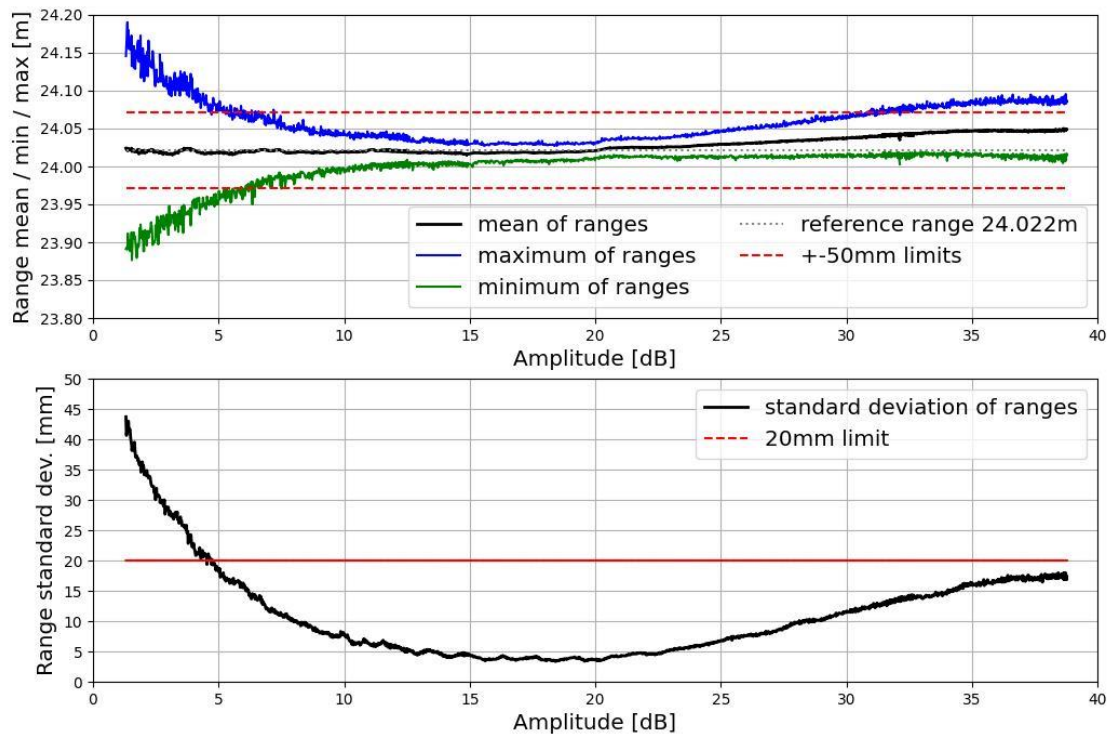
Figure 9. Resulting range measurement accuracy and range standard deviation when measuring to a plane
target which is placed at a distance of 24 m and oriented perpendicular with respect to the laser axis

Maximum range - dependencies

The potentially achievable maximum range of a LiDAR system depends first of all on the laser's pulse power, the sensitivity of the detector used in the receiver, and the size of the receiver's optical aperture. Besides these system inherent properties, several external / environmental conditions which are beyond the system operator's control need to be considered:

- the reflectance of a target's surface,

- the size of the target in comparison to the laser's footprint (at the distance in question),

- the clarity of the atmosphere, usually defined by the visibility, and

- the amount of background light within the receiver's FOV, whether coming from the sun, vehicle's head lamps, road lighting, or whatsoever.

In this special design, background light has some non-negligible influence on the achievable maximum range. Each receiver channel has a rather large FOV of 20° x 1.64 mrad. An optical bandpass filter reduces the amount of background light arriving at the detector pixel. However, the bandwidth of this filter had to be selected quite large due to:

- the laser-diode's spectral width of typical 7 nm (FWHM),

- the sample specific variation of the laser-diode's center wavelength of typical ±10 nm,

- the laser's thermal wavelength shift of about +0.28 nm/K, in combination with the laser-diode's self-heating of up to 60 K, thus causing a shift of up to 17 nm, and

- a shift of the filter's center wavelength with the beam's angle of incidence (±10°).

Based on the already known properties of the integrated components as well as the design of the optical channels, the achievable maximum range has been calculated for varying environmental conditions, see Figure 10.
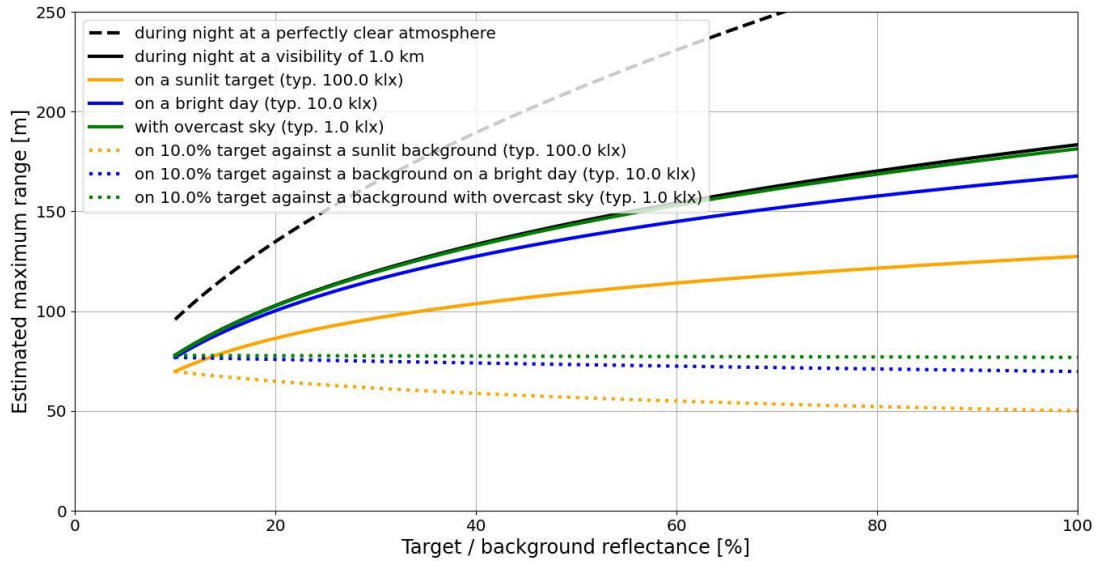

Figure 10. Estimation of the achievable maximum range under varying environmental conditions

All lines with exception of the black dashed line assume a visibility of about 1 km. (On a clear day the visibility may be significantly higher, e.g., even several 10 km.)

Furthermore, all estimations assume

- a flat foreground target which is only slightly larger than the laser footprint,

- a very large background target which covers the whole vertical FOV of the sensor behind the foreground target,

- that both targets are oriented perpendicular with respect to the laser axis, and

- each target's reflection characteristic is that of a Lambertian scatterer.

Due to the small vertical divergence of the laser fan (about 1.25 mrad) in comparison to a receiver pixel's vertical FOV (±10°), only a small part of a receiver pixel is hit by the echo pulse of the foreground target. However, the whole remaining pixel may be hit by reflected light of the background target. Therefore, the diagram shows two types of results:

- Solid lines show the estimated maximum range for a target situation where the foreground and the background target have the same reflectance and are hit by a varying amount of sunlight, i.e. starting at 0 klx (during night) and up to 100 klx (a bright day around noon). With increasing reflectance of both targets, the amplitude of the foreground target's echo signal increases by the same factor as the reflected light of the background target when arriving at the receiver.

- Dotted lines show the estimated maximum range for a foreground target with a constant reflectance of 10% and a background target with varying reflectance. Both are hit by the varying amount of sunlight (1 klx up to 100 klx), but the variation of the reflectance indicated on the x-axis of the diagram applies only to the background target. With increasing reflectance, only the reflected light of the background target arriving at the receiver increases, but the amplitude of the foreground target's echo remains constant (as its reflectance

remains constant at 10%). Thus, the achievable maximum range reduces with increasing reflectance of the background target.

4. CURRENT STATUS OF THE PROTOTYPE LIDAR SENSOR

A prototype of the LiDAR sensor based upon the design described above has been developed in course of the project iLIDS4SAM. This prototype is currently under realization on the level of an elegant breadboard, but behind schedule due to the delayed availability of several key components. Thus, test data are not available at the time of providing this report.

Figure 11. LiDAR sensor under development: view from outside

The main intention of this prototype is first of all the proof of concept, so some bulkiness has been accepted:

- Main dimensions: 424 mm x 225 mm x 232 mm (L x W x H)
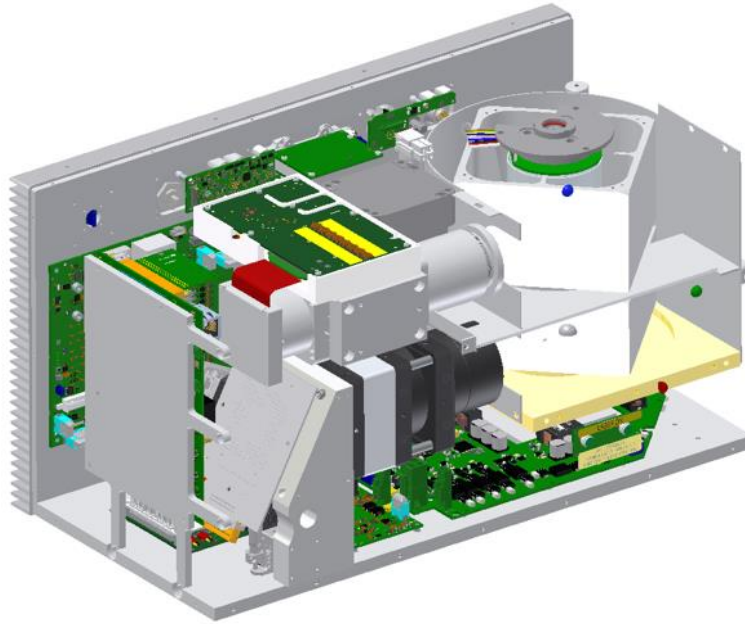- Mass (preliminary): 15 kg

Figure 12. LiDAR sensor under development: view of the interior

5. OUTLOOK

As the main purpose of the mentioned LiDAR sensor is basically the proof of concept and evaluation of its performance, there remains much room for further development and improvements. Two main directions of such further steps have been identified.

Reduction of SWaP

The design of the LiDAR sensor has not been optimized for compactness, mainly due to three reasons:

- The sub-assemblies have been designed by different project partners for fulfilling the required functionality and specifications. These sub-assemblies have not undergone a combined optimization for size and weight yet.

- To save costs and time, some of the electronics boards originate from earlier developments although neither size and shape nor functionality of these boards fit perfect. As a consequence, the power consumption is higher than necessary. Nevertheless, even the newly developed electronic boards are based on commercial-off-the-shelf (COTS) components and/or earlier developed ASICs.
  Thus, there is a significant potential for reducing the board count and their sizes when developing fully new set of electronic boards integrating purpose-specific ASICs.

- Several key components used for the LiDAR sensor had to be selected from COTS products, as their customization would have increased costs beyond the available budget.

  - In many cases the functionality of the COTS components does not fulfill the required specifications, thus compensation measures are necessary which require additional space, add weight and increase power consumption.

  - In some cases, the COTS components are larger than required.

A typical example of such a component is the APD array which has been carefully selected but still does not offer the required, very high pixel aspect ratio. To adapt the aspect ratio accordingly, the optical design of the receiver assembly had to be enhanced by additional lenses. This added weight, required space and reduced the overall optical efficiency of the receiver path. Thus, the overall performance (e.g. maximum range) is reduced and/or an increase of the laser transmitter power is needed for compensation.

As a consequence of the increased size of the optical receiver path even the height of the polygon mirror wheel had to be enlarged.

Design enhancements

In order to achieve longevity, many designs of automotive LiDAR sensors refrain from using moving parts. From that point of view the usage of a polygon mirror wheel may be considered somewhat outdated. However, this in particular offers some advantageous chances for functional enhancements for a favorable point distribution:

- Utilization of a non-regular polygon mirror wheel of different facet sizes, each of them offering a different horizontal FOV:
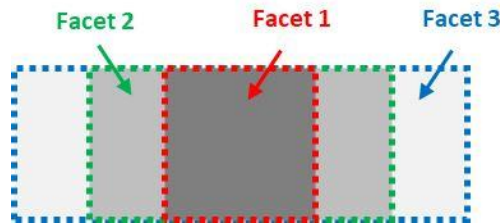
Figure 13. Sketch of acquired areas when applying different horizontal FOVs

- o The central area of the FOV which is of highest interest is acquired at a high frame rate and thus high spatial resolution whereas the boundary areas are scanned at a lower rate.

- o This configuration reduces the point density in the boundary areas in favor of the center area.

- Combining two sets of transmitter and receiver assemblies with one polygon mirror wheel:
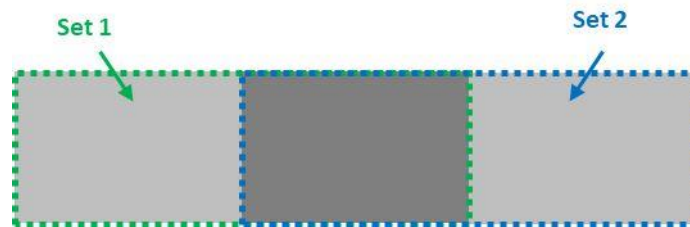
Figure 14. Sketch of acquired areas when combining two sets of transmitter and receiver assemblies

- o The horizontal FOVs of these two sets have the same size but are shifted horizontally by a certain angle. In the overlap region the resulting point density and/or frame rate is twice that of the boundary regions.

- o This configuration splits the required pulse repetition rate, thus reducing the average PRR and the thermal stress for each transmitter array.

- Integration of two laser diode arrays with vertical angle shift into the transmitter assembly:

Figure 15. Sketch of acquired areas when integrating two laser diode
arrays with a vertical angle shift into the transmitter assembly

- o The vertical FOVs of the two laser diode arrays have the same size but are shifted vertically by a certain angle. In the overlap region the resulting point density and/or frame rate is twice that of the boundary regions.
- o This configuration splits the required pulse repetition rate, thus reducing the average PRR and the thermal stress of each transmitter array.

6. ACKNOWLEDGMENT

REFERENCES

[1] "iLIDS4SAM: Integrated LiDAR sensors for safe & smart automated mobility", https://www.ilids4sam.at

[2] Baumgart, M., Reichert, R., Kirillov, B., Hennecke, M.E., Tortschanoff, A., Pfennigbauer, M., Hofbauer, A., "Beam-shaping for a Lidar system for urban scenarios", Submitted to IEEE Sensors 2023

[3] Jones, E., Hennecke, M.E., "GaN-Based Lidar Pulse Generator Achieving 320 A in 6 ns: Design and System Integration Considerations", APEC Industry Session 11.2 (March 2022), https://www.psma.com/sites/default/files/uploads/node/7227/IS11.2.pdf

[4] Pfennigbauer, M., Ullrich, A., "Improving quality of laser scanning data acquisition through calibrated amplitude and pulse deviation measurement", Proc. SPIE 7684, 7684-53, (2010). https://doi.org/10.1117/12.849641

[5] Briese, C., Höfle, B., Lehner, H., Wagner, W., Pfennigbauer, M., Ullrich, A., "Calibration of Full-Waveform Airborne Laser Scanning Data for Object Classification", Laser Radar Technology and Applications XIII, edited by Monte D. Turner, Gary W. Kamerman, Proc. of SPIE Vol. 6950, 69500H, (2008) · 0277-786X/08/$18 https://doi.org/10.1117/12.781086

[6] Jelalian, A.V., "Laser radar systems," Artech House, Boston London, 1992

Laser scanners with polygonal mirrors:
Optomechanical analysis and design

Virgil-Florin Duma[*,a,b], Maria-Alexandra Duma[c]

[a]3OM Optomechatronics Group, *Aurel Vlaicu* University of Arad, 77 Revolutiei Ave.,
310130 Arad, Romania

[b]Doctoral School, Polytechnic University of Timisoara, 1 Mihai Viteazu Ave.,
300222 Timisoara, Romania

[c]School of Science and Engineering, University of Groningen, W.F. Bathoornstraat,
9731 CG Groningen, The Netherlands

ABSTRACT

We report a fully-developed theory of laser scanners with rotational Polygon Mirrors (PMs). The deduced scanning function, velocity, and acceleration of PMs have been deduced and discussed in comparison to those of Galvanometer Scanners (GSs). All other characteristic parameters have been obtained, including angular and linear field-of-view (FOV), as well as duty cycle [Proc. of the Romanian Acad. Series A 18, 25-33, 2017]. Although this developed theory considered the laser beams reduced to a single ray (i.e., the center axis of the beam), the specific approach has allowed further on for a complete analysis for scanning laser beams with finite diameters. The multi-parameter optomechanical analysis of these PM functions was performed as well, considering all constructive and functional parameters [Appl. Sci. 12, 5592 (2022)]. The non-linearity of scanning functions (i.e., the non-constant scanning velocities) has been approached. In order to linearize the PM or GS scanning function, a two supplemental mirrors device was developed. This increases the distance between the PM and its objective lens within a reasonable dimension of the system, by folding the scanned laser beam. Rules-of-thumb have been obtained for the design of these scanning heads. The optical part has been completed with a Finite Element Analysis (FEA) of rotational PMs, assessing their structural integrity. An optomechanical design scheme completes the PM scanning heads study, highlighting the links between optical and mechanical aspects. This type of scheme can be utilized for other optomechanical scanners, as well.

Keywords: laser scanners, polygon mirrors (PMs), galvanometer scanners (GSs), geometrical optics, modeling, multi-parameter analysis, Finite Element Analysis (FEA), optomechatronics.

1. INTRODUCTION

Polygon mirrors (PMs)-based scanners are one of the fastest optomechatronic laser scanning devices [1,2]. Compared to the most utilized galvanometer-based scanners (GSs) or to MEMS devices, the purely rotational PMs avoid the necessity to stop-and-turn of scanners with oscillatory elements. Therefore, the scan frequency of PM scanners can surpass that of GSs and MEMS even if the latter are employed at resonance. Furthermore, as we have demonstrated [3-6], there is always a trade-off between scan frequency and amplitude of GSs. As the former is increased the latter is strongly limited, with an almost exponential decrease [4,6], which is also dependent on the scanning input signals (i.e., sinusoidal, triangular, and sawtooth, the latter with different input/theoretical duty cycles). In contrast, PM frequency and amplitudes are not coupled, although in order to increase the scan frequency by using PMs with a higher number of facets, larger PMs must be employed – in order to preserve the scan amplitude [1,7].

The present report points out briefly such (optomechanical) aspects, while providing an overview on our works on PM-based scanning heads [8-15] (compared to other scanning systems, as well, such as GSs [16], MEMS [17,18], and Risley prisms [19-21]), addressing both optical and mechanical aspects of PMs.

---

[*] Email: duma.virgil@osamember.org; phone: +40-751-511451; sites: http://3om-group-optomechatronics.ro/;
https://www.researchgate.net/lab/3OM-Optomechatronics-Group-Virgil-Florin-Duma

2. OPTICAL ANALYSES OF PM SCANNING HEADS

While GSs and MEMS with oscillatory mirrors have as a convenient feature the fact that the plane of their mirror is on their oscillatory axes, PMs are characterized by a certain distance from their rotational axis to each facet. This distance defines the PM apothem R (Fig. 1), while the number n of facets gives the characteristic angle $\alpha = \pi/n$, which corresponds to half of a facet. Other parameters for the most common case of a laser beam incident on the PM facets perpendicular to the optical axis (O.A.) of the post-objective lens (of diameter D) are the eccentricity e (of the incident laser beam with regard to the PM pivot O) and the distance L from this incident beam to the (object principal plane of the) lens.
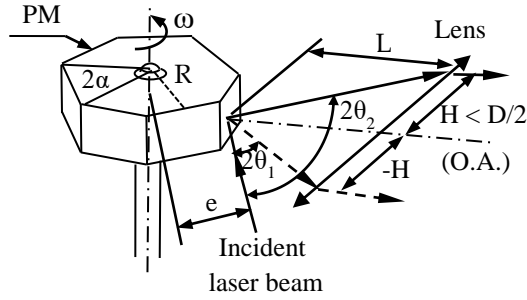

Figure 1. Characteristic parameters of a PM-based scanning head with the incident laser beam perpendicular to the O.A. of the objective lens – setup considered in the detailed studies in [8,9], as well as in the preliminary reports in [10-15].

By analyzing the movement of the PM, we have deduced and analyzed several characteristic function and parameters in [8,9], with regard to the above geometrical parameters and to the rotational velocity ω of the PM:

(i) **The scanning function $h(\theta)$** has been defined as the current distance from the O.A. to the beam deflected by the PM and refracted by the lens; it has been deduced in [8-11] as

$$h(\theta) = R\sqrt{2} - e - \frac{R}{\cos\theta} + e \cdot \tan\theta - \frac{L}{\tan 2\theta} . \tag{1}$$

With regard to the scanning function of a GS that has the incoming beam incident on the mirror on its oscillatory axis, the PM scanning function has two more terms (i.e., the first two ones in the above equation). While the GS scanning function depends only on the distance L to the lens or to the scanned plane (and on the rotation angle of the GS mirror), the PM function $h(\theta)$ also depends on e and R (i.e., parameters that are specific only to PMs). This complicates the mathematical discussion but does provide two additional DOFs in optimizing the scanning process.

(ii) **The scanning velocity and acceleration** have been deduced and compared in [8-11] to those of GSs. Similar to the $h(\theta)$ function, the velocity $v(\theta)$ and the acceleration $a(\theta)$ of PMs depend on all constructive parameters R, e, and L, but also on ω.

(iii) **The rotational angles θ_{min} and θ_{max}** of the PM that define what Beiser named down- and up-deflected beams [7], respectively are purely functions of e and R or of the unitless factor $e/r = \xi$ that we have defined in order to compare (and optimize) such scanning systems regardless of their dimensions. This pair of angles defines the angular field-of-view (FOV) of the scanner with the simple relationship

$$FOV = 2(\theta_{max} - \theta_{min}) = 4\alpha \tag{2}$$

It is worth mentioning that this maximum available angular FOV imposes the (maximum) dimension D, therefore the objective lens, as the distance L is approximately equal to the object focal length.

(iv) **The angles θ_1 and θ_2** characterize the rotation of the PM for which the reflected beam reaches the lower and the upper margins of the lens in the meridian plane, respectively. Usually

$$\theta_{min} < \theta_1 < \theta_2 < \theta_{max} , \tag{3}$$

as the lens is placed within the FOV of the PM. Ideally, one may want to have

$$\theta_{\min} = \theta_1 < \theta_2 = \theta_{\max}, \tag{4}$$

but this is a condition that can be fulfilled only in particular cases, as the PM's FOV is asymmetrical, as pointed out by several researchers [22-24].

It is worth mentioning that the case

$$\theta_1 < \theta_{\min} < \theta_{\max} < \theta_2 \tag{5}$$

may be of interest as well, although it may not seem convenient, as the FOV does not fill the entire lens aperture in this case. However, double pass PMs can be utilized in such a case, with the beam reflected on the next PM facet already beginning its pass, although the current beam has not yet completed its pass [2].

(v) **The duty cycle** η of the scanner is defined as the time efficiency of the scanning process [1,2]. Using the two pairs of angles defined above,

$$\eta = (\theta_2 - \theta_1) / (\theta_{\max} - \theta_{\min}), \tag{6}$$

therefore

$$\eta = (\theta_2 - \theta_1)/2\alpha = n \cdot (\theta_2 - \theta_1)/2\pi. \tag{7}$$

From these relationships one may observe that for PMs η is "n" times higher than for single mirror scanners, rotational (i.e., the monogon – often manufactured in its pyramidal configuration) or oscillatory (i.e., for GSs, MEMS, or other (including low-cost) variants with tilting mirrors [25]).

Also, from the discussion at the previous point,

$$\eta < 1 \text{ for Eq. (3)}$$

$$\eta = 1 \text{ for Eq. (4)}$$

$$\eta > 1 \text{ for Eq. (5)}$$

(vi) **The migration functions** were proposed, defined and characterized in our previous works [8-11]. They are caused by the translation of the laser spot on the PM facet during the PM rotations, due to the distance R from the PM facet to the pivot O. Essentially, it is as if the beam that is incident on the lens starts from a certain point A on the O.A. of the lens [8-11]. This point A is moving on the O.A., therefore it cannot stay in the object focal point of the lens, and this defines the **longitudinal/axial displacement function** $z(\theta)$. Also, there is a **transversal displacement function** $y(\theta)$, defined by the displacement of the reflection point P in Fig. 1 [8] on the PM facet.

Multi-parameter analyses of these two functions (that help understanding and designing PM scanning heads) have been performed in Fig. 3 [8] and in Fig. 4 [9].

One must highlight that the entire discussion above has been completed for the laser beam considered to be reduced to a single ray, i.e. to its center axis. However, the specific geometric approach and parametrization of this analysis allowed for solving the real-life case of finite diameter (2ρ) laser beams by using the eccentricity e of the incident laser. Thus, the characteristic functions and parameters of the PM scanning head can be written for $e - \rho$ and $e + \rho$, respectively - for the left and right margins of the laser beam, as we have developed in detail in [9], and as it was briefly approached for applications in [26], based on our previous theory and suggestions in [8].

3. OPTIMIZATION OF THE SCANNING FUNCTION AND VELOCITY

A main issue of PM scanning heads, similar to GSs or MEMS is the non-linearity of the scanning function $h(\theta)$, which produces a variable scanning velocity $v(\theta)$, lower around the O.A. and higher towards the margins of the lens. A common way of tackling with this issue is to employ F-theta lenses. However, this increases the cost of the system. Therefore, efforts have been made to design lens systems that are both appropriate and as simple as possible [27,28].

In this respect we have had a different approach, by increasing the distance L, which has the highest impact on the linearity of the scanning function $h(\theta)$, as demonstrated in [9]. This means using a lens with a longer focal distance. The issue in this case is related to the increase of the axial dimension of the system. Supplemental mirrors can be employed in order to fold the beam, with three additional fix mirrors [29], or even with two such mirrors, placed at a certain (small) angle with regard to each other. The latter solution has been explored [10] and developed in detail [9], in order to allow

for an optimized design of the system. This implies using an angle between the two additional mirrors that allows for maximizing *L*, as well as for minimizing the necessary dimensions of the mirrors.

4. MECHANICAL ASPECTS: FINITE ELEMENT ANALYSIS (FEA) OF THE POLYGONS

PMs have numerous mechanical issues that must be considered during the design and manufacturing stages. Such aspects involve for example motors, bearings, materials, tolerances, manufacturing process and roughness of (active) optical surfaces. One of their main issues is related to their structural integrity, due to the high centrifugal forces produced by their high rotational velocities. Therefore, in order to properly design PMs, a FEA must be carried out, usually in a multi-parameter approach. Both structural integrity issues and levels of deformations must be observed, as in the case of other optomechanical systems with fast moving (roational) elements, for example choppers and scanners [30-36] – in various configurations.

An example of such a multi-parameter FEA was made in our detailed study in [9], for a most common PM configuration. Regarding the conclusions of this study, the rotational velocity ω (considered up to 120 krpm) is the most impactful functional parameters that may contribute to the appearance of certain zones with critical levels of the mechanical stress σ_{max}, for example:

- The number *n* of facets does not impact the level of σ_{max}, although one can conclude that as *n* increases, σ_{max} decreases on the exterior part of PM;

- σ_{max} is maximum (for all levels of *n*) towards the central (mounting) cylindrical hole of the PM, i.e. towards the shaft. The higher the radius of this central hole is, the higher σ_{max} gets, with the possibility to surpass the yield limit;

- The apothem *R* significantly impacts σ_{max}, with a proportional relationship between the two parameters (as the centrifugal forces are dependent on the radial dimensions of the PM);

- The PM width increases the levels of σ_{max} and may also change the positions for which it reaches its maximum;

- The mounting holes of the PM on its support impact σ_{max} in what regards their number, position (towards the PM apex or corresponding to the middle of a PM facet) and radius. A trade-off must be made between such aspects and the constructive requirements of the mechanical assembly [9];

- The material (with its specific parameters, mainly density, yield stress and Young modulus) is an essential aspect, and there is roughly a limit of ω for which structural steel can still be utilized, as well as another, higher limit of ω from which beryllium alloys must be considered (and alluminium alloys are no longer appropriate).

5. DESIGNING SCHEMES OF PMS – OPTOMECHANICAL LINKS

Figure 2 presents a designing scheme that is a synthesis of calculus procedures presented in the literature [1,2]. Essentially, the requirements of the design theme regarding scan frequency and beam diameter must be correlated with the technologically achievable levels of ω in order to determine the number *n* of facets and then the PM apothem.

However, this leaves numerous other issues, both optical and mechanical that must be addressed [13]. Therefore, the second designing scheme that concluded our detailed study on this topic (Fig. 17 [9]) tries to address the necessary links between the various aspects, resulting in an optomechanical scheme that can support designers in this task – not only for PMs, but for other types of optomechanical scanners, as well.

6. CONCLUSIONS

PM design is a complex task for which both optical and mechanical aspects must be considered, as approached in detail in [9]. Regarding the purely optical aspects, these studies tried to offer an exact and rigorous, although as simple as possible approach regarding all characteristic parameters and functions of PMs. The applications range of these scanners is wide, but some of them stand out, including for Swept Sources (SSs) for Optical Coherence Tomography (OCT) [33-36] for both biomedical imaging [18,37,38] and Non-Destructive Testing (NDT) [39], for laser manufacturing (in a PM plus GS scanning head, as alternatives to dual axis 2D GS [6]) and in Remote Sensing, the latter in order to rapidly scan, with good resolution and FOV, from airborne platforms for example. Several of these applications are promising directions of work, including in our group.
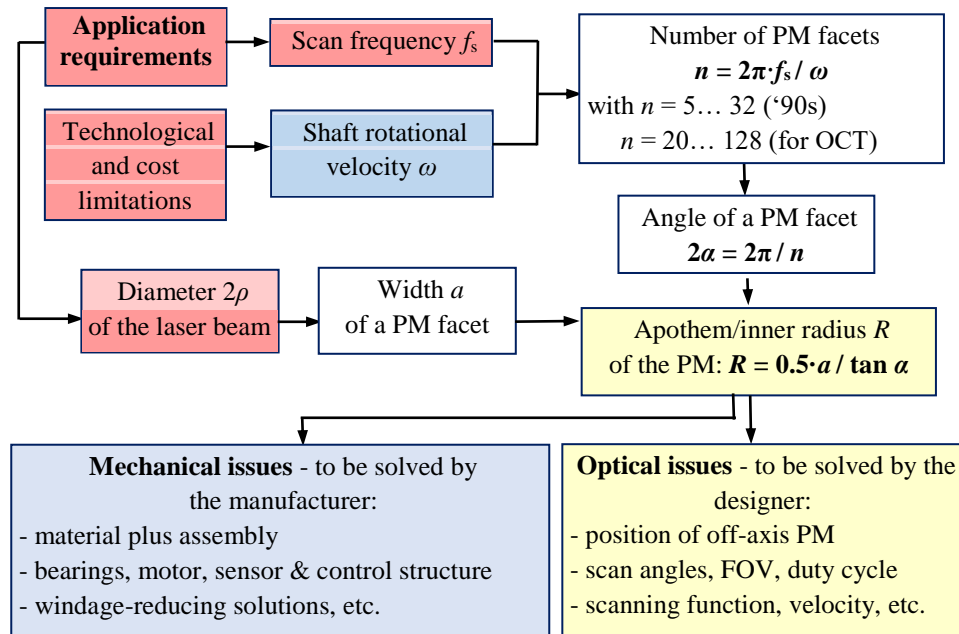
Figure 2. Optical scheme for the design of a PM, made in [9] following the steps of the classical calculus in [1], followed by mechanical issues that are left to the maunfacturers, as well as by optical aspects addressed in our preliminary studies in [10-15] and finalized in [8.9].

ACKNOWLEDGEMENTS

REFERENCES

[1] Marshall, G. F., [Handbook of Optical and Laser Scanning], CRC Press–Taylor and Francis, New York (2011).

[2] Bass, M., Ed., [Handbook of optics], Mc. Graw-Hill Inc., New York, (2010).

[3] Duma, V.-F., "Optimal scanning function of a galvanometer scanner for an increased duty cycle," Optical Engineering 49(10), 103001 (2010).

[4] Duma, V.-F., Lee, K.-S., Meemon, P., Rolland, J.P., "Experimental investigations of the scanning functions of galvanometer-based scanners with applications in OCT," Appl. Opt. 50(29), 5735-5749 (2011).

[5] Duma, V.-F., Tankam, P., Huang, J., Won, J.J., Rolland, J.P., "Optimization of galvanometer scanning for Optical Coherence Tomography," Appl. Opt. 54(17), 5495-5507 (2015).

[6] Duma, V.-F., "Laser scanners with oscillatory elements: Design and optimization of 1D and 2D scanning functions," Appl. Math. Modelling 67, 456-476 (2019).

[7] Beiser, L., "Design equations for a polygon laser scanner," Proc. SPIE 1454, 60-66 (1991).

[8] Duma, V.-F., "Polygonal mirror laser scanning heads: Characteristic functions," Proc. of the Romanian Academy Series A 18(1), (2017).

[9] Duma, V.-F., Duma, M.-A., "Optomechanical Analysis and Design of Polygon Mirror-Based Laser Scanners," Appl. Sci. 12, 5592 (2022).

[10] Duma, V.-F., "On-line measurements with optical scanners: metrological aspects," Proc. SPIE 5856, 606-617 (2005).

[11] Duma, V.-F., "Novel approaches in the designing of the polygon scanners," Proc. SPIE 6785, 6785-1Q (2007).

[12] Duma, V.-F., Nicolov, M., "Numerical and experimental study of the characteristic functions of polygon scanners," Proc. SPIE 7390, 7390-42 (2009).

[13] Duma, V.-F., Rolland, J.P., "Mechanical constraints and design considerations for polygon scanners," Mechanisms and Machine Science 5(8), 475-483 (2010).

[14] Duma, V.-F., Podoleanu, A., "Polygon mirror scanners in biomedical imaging: a review," Proc. SPIE 8621, 8621V (2013).

[15] Duma, V.-F., "Scanning in biomedical imaging: from classical devices to handheld heads and micro-systems," Proc. SPIE 8925, 8925 0L (2014) – Invited Paper.

[16] Montagu, J., Scanners - galvanometric and resonant, in: C. Hoffman, R. Driggers (Eds.) Encyclopedia of Optical and Photonic Engineering, 2nd ed. CRC Press (2015).

[17] Lu, C.D., Kraus, M.F., Potsaid, B., et al., "Handheld ultrahigh speed swept source optical coherence tomography instrument using a MEMS scanning mirror," Biomed. Opt. Express 5, 293-311 (2014).

[18] Cogliati, A., et al, Rolland, J. P., "MEMS-based handheld scanning probe with pre-shaped input signals for distortion-free images in Gabor-Domain Optical Coherence Microscopy," Opt. Express 24, 13365-13374 (2016).

[19] Marshall, G.F., "Risley Prism Scan Patterns," Proc. SPIE 3787, 74-86 (1999).

[20] Li, A., Liu, X., Sun, J., Lu, Z., "Risley-prism-based multi-beam scanning LiDAR for high-resolution three-dimensional imaging," Optics and Lasers in Engineering 150, 106836 (2022).

[21] Duma, V.-F., Dimb, A-L., "Exact Scan Patterns of Rotational Risley Prisms Obtained with a Graphical Method: Multi-Parameter Analysis and Design," Appl. Sci. 11(18), 8451 (2021).

[22] Li, Y., Katz, J., "Asymmetric distribution of the scanned field of a rotating reflective polygon," Appl. Opt. 36, 342-352 (1997).

[23] Li, Y., "Single-mirror beam steering system: analysis and synthesis of high-order conic-section scan patterns," Appl. Opt. 47, 386-398 (2008).

[24] Li, Y., "Beam deflection and scanning by two-mirror and two-axis systems of different architectures: a unified approach," Appl. Opt.47, 5976-5985 (2008).

[25] Shen, C.-K., Huang, Y.-N., Liu, G.-Y., Tsui, W.-A., Cheng, Y.-W., Yeh, P.-H., Tsai, J.-c., "Low-Cost 3D-Printed Electromagnetically Driven Large-Area 1-DOF Optical Scanners," Photonics 9, 484 (2022).

[26] H.-M. Hoang, S. Choi, C. Park, J. Choi, S.H. Ahn, J. Noh, "Non-back-reflecting polygon scanner with applications in surface cleaning," Opt. Express 29, 32939-32950 (2021).

[27] Varughese K.O.G. and Siva Rama Krishna, K., "Flattening the field of postobjective scanners by optimum choice and positioning of polygons," Appl. Opt. 32, 1104–1108 (1993).

[28] Walters, C.T., "Flat-field postobjective polygon scanner," Appl. Opt. 34, 2220-2225 (1995).

[29] Duma, M.-A., Duma, V.-F., "Theoretical approach on the linearity increase of scanning functions using supplemental mirrors," Proc. SPIE 11028, 1102817 (2019).

[30] Duma, V.-F., "Theoretical approach on optical choppers for top-hat light beam distributions," J. of Opt. A: Pure and Appl. Opt. 10, 064008 (2008).

[31] Duma, V.-F., "Optical choppers with circular-shaped windows: Modulation functions," Communications in Nonlinear Science and Numerical Simulation 16, 2218-2224 (2011).

[32] Duma, V.-F., et al, "Optical choppers with high speed rotating elements," Proc. SPIE 9131, 9131OE (2014).

[33] Oh, W.Y., Yun, S.H., Tearney, G.J., Bouma, B.E., "115 kHz tuning repetition rate ultrahigh-speed wavelength-swept semiconductor laser," Opt. Lett. 30, 3159-3161 (2005).

[34] Yun, S.H., Boudoux, C., Tearney, G. J., Bouma, B.E., "High-speed wavelength-swept semiconductor laser with a polygon-scanner-based wavelength filter," Opt. Lett. 28, 1981-1983 (2003).

[35] Mao, Y., Flueraru, C., Sherif, S., Chang, S., "High performance wavelength-swept laser with mode-locking technique for optical coherence tomography," Opt. Communications 282, 88–92 (2009).

[36] Leung, M.K.K., et al, "High-power wavelength-swept laser in Littman telescope-less polygon filter and dual-amplifier configuration for multichannel optical coherence tomography," Opt. Lett. 34, 2814-2816 (2009).

[37] Erdelyi, R.A., Duma, V.-F., Sinescu, C., Dobre, G.M., Bradu, A., Podoleanu, A., "Dental Diagnosis and Treatment Assessments: Between X-rays Radiography and Optical Coherence Tomography," Materials 13(21), 4825 (2020).

[38] Erdelyi, R.-A., Duma, V.-F., Sinescu, C., Dobre, G.M., Bradu, A., Podoleanu, A., "Optimization of X-ray Investigations in Dentistry Using Optical Coherence Tomography," Sensors 21(13), 4554 (2021).

[39] Cioabla, A., et al, "Effect of an Anaerobic Fermentation Process on 3D-printed PLA Materials of a Biogas-generating Reactor," Materials 15(23), 8571 (2022).

Laser Doppler Multi-Beam Differential Vibrometers for Detection of Buried Objects

Vyacheslav Aranchuk*[a], Boyang Zhang[a], Stanley Johnson[b], Ina Aranchuk[a], and Craig Hickey[a]

[a]University of Mississippi, 145 Hill Drive, University, MS, USA 38677-1848; [b]Coherent Light Science LLC, 7040 N. Montecatina Dr., Tucson, AZ 85704, USA

ABSTRACT

The laser-acoustic detection of buried objects, such as landmines, is based on excitation of elastic waves in the ground and creating a vibration image of the ground surface by using a laser Doppler vibrometer (LDV). The technique provides high probability of detection and low false alarm rate. However, traditional LDVs require operation from a stable stationary platform due to their sensitivity to the motion of the vibrometer itself. Recently developed laser Doppler multi-beam differential vibration sensor has low sensitivity to the motion of the sensor itself, while measuring vibration velocity difference between points on the object with interferometric sensitivity. Low sensitivity to the sensor motion allows for vibration measurements from a moving vehicle. Two configurations of the developed sensor: the linear array and the 2D array sensors, are discussed in the paper. The linear array sensor measures velocity difference between points on the object illuminated with a linear array of 30 laser beams, and creates a vibration image of the object by scanning the array of beams in a transverse direction. The 2D array sensor employs an array of 34 x 23 laser beams and measures velocity difference between corresponding points on the object over the whole illuminated area simultaneously. Simultaneous measurements at all points allow for the fast recording of the vibration image of the area of interest, and makes possible calculation of the vibration phase and instantaneous velocity images. Description of the sensors and the experimental results are presented in the paper.

Keywords: laser vibrometer, landmine detection, buried object, laser-acoustic detection, sensor.

1. INTRODUCTION

Laser-acoustic detection of buried objects, such as landmines, has proven itself as a technique that provides high probability of detection and low false alarm rate[1-4]. The method consists of exciting vibrations in the ground and measuring vibration of the ground surface at multiple points with a laser Doppler vibrometer (LDV) to create a vibration image of the ground surface. Vibrations of the ground in the frequency range from about 50 Hz to 300 Hz are excited by using airborne sound created by loudspeakers or seismic waves created by mechanical shakers. The interaction of a buried object with the elastic waves in the ground causes the object to vibrate. Due to the mechanical resonances and the higher mechanical compliance of the buried object compared to the neighboring soil, the vibration amplitude of the ground surface above the object at some vibration frequencies is higher than the vibration amplitude of the surrounding area. Therefore, a buried object can be detected by the area with higher vibration amplitude in the vibration image of the interrogated area. Single beam and multiple beam LDVs have been traditionally used for ground vibration sensing in laser-acoustic detection of buried objects[4-8]. However, since a LDV measures the relative velocity between the LDV and the object, the measurement results depend not only on the object vibration, but also on the motion of the LDV itself. As a result, motion of the LDV due to environmental vibration and acoustic noise can cause LDV signals to be significantly higher than, and indistinguishable from signals caused by the object vibration. Usually, LDVs operate from a stationary mechanically stable platform in order to withstand the ambient vibration. LDV operation from a moving vehicle with laser beams looking forward imposes another challenge for buried object detection due to Doppler shift in the LDV signal proportional to the speed of the vehicle. This Doppler shift caused by the vehicle motion can be several orders of magnitude higher than the Doppler shift caused by the ground vibration, and often exceeding the modulation bandwidth of the LDV, thereby making measurements impossible. A recently developed Laser Multi-Beam Differential Interferometric Sensor (LAMBDIS) has low sensitivity to the motion of the sensor itself, while measuring vibration velocity difference between points on the object with interferometric sensitivity[9-11].

*aranchuk@olemiss.edu; phone 1 662 915-5610

The principle of operation of the LAMBDIS is based on measuring the velocity difference between points on the object by employing interference of light reflected from the object points, without using a reference beam. Due to the absence of a reference beam, the Doppler shift induced by the sensor motion is canceled out, that results in low sensitivity to the sensor motion. Low sensitivity to the sensor motion allowed for the application of the LAMBDIS for detection of buried objects from a moving vehicle. Two configurations of the LAMBDIS: the linear array and the 2D array sensors, have been developed and are discussed in the paper. The linear array LAMBDIS measures velocity difference between points on the object illuminated with a linear array of 30 laser beams, and creates a vibration image of the interrogated area by scanning the array of beams in a transverse direction. The 2D array LAMBDIS employs a matrix array of 34 x 23 laser beams and measures velocity difference between corresponding points on the object over the whole illuminated area simultaneously. Simultaneous measurements at all points allow for the fast recording of the vibration image of the area of interest, and makes possible calculation of the vibration phase and instantaneous velocity images.

2. LINEAR ARRAY LAMBDIS

The principle of operation of the linear array LAMBDIS is based on the interference of light reflected from different points on the object surface illuminated with a linear array of laser beams, as shown in Figure 1.
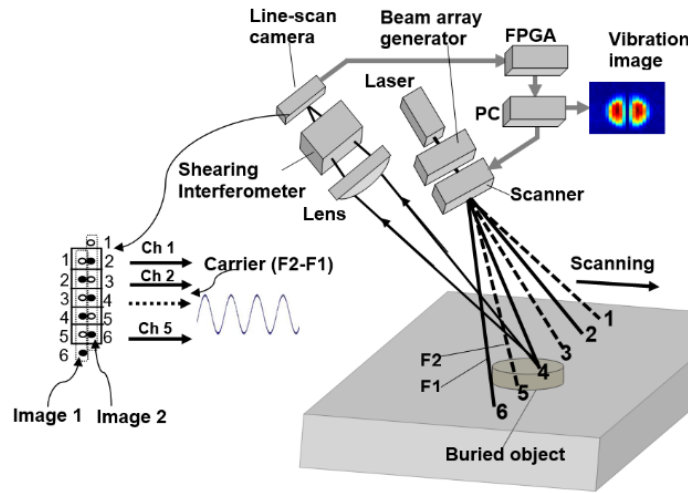

Figure 1. Functional layout of linear-array LAMBDIS based on digital line-scan CMOS camera and FPGA-based real-time signal processor

A linear array of 30 laser beams (only 6 beams are shown for clarity), is generated by the beam array generator of the sensor and focused onto an object (ground surface). Adjacent beams in the array have different optical frequencies, F_1 and F_2. Specifically, each beam of frequency F_1 is positioned on the object surface exactly between two beams of frequency F_2, so that the optical frequency of even number beams is different from the optical frequency of odd number beams. A receiver lens and a shearing interferometer create two sheared images of the laser spots on the object on the digital line-scan CMOS camera. The two images are sheared relative to each other in the direction of the array of points by an odd number of intervals between neighboring laser spots. In the image plane, the light from each pair of corresponding laser spots are mixed on the CMOS sensor, producing heterodyne signals with the carrier frequency $F_C = F_2 - F_1$. For example, referring to Figure 1, for the case of one interval shear between the two images, Image 1 and Image 2 of laser spots on the object surface, the Image 1 of spot 1 is overlapped with the sheared Image 2 of spot 2, the Image 1 of spot 2 is overlapped with the sheared Image 2 of the spot 3, and so on.

Vibration of the object causes frequency shifts of reflected light due to the Doppler effect, resulting in frequency modulated signals with the carrier frequency $F_C = F_2 - F_1$ on the CMOS camera. The vibration velocity difference between the corresponding points on the object, for example between points 1 and 2, points 2 and 3, and so on, are computed in real-time through demodulation of these heterodyne signals by a FPGA based processor. A vibration image of the measured

area of the ground can be obtained by either scanning the array of beams over the area from a stationary platform using a scanner, or by moving the sensor.

Motion of the sensor causes approximately the same Doppler shift for all beams, and is automatically subtracted from the measurements. So, the LAMBDIS principle allows for measuring vibration velocity between points on the object surface with interferometric sensitivity, while having low sensitivity to the motion of the sensor itself. Experiments on detection of buried objects have been conducted with the linear array LAMBDIS mounted on a vehicle that demonstrated the ability of the linear array LAMBDIS to obtain vibration images of the ground surface and detect buried objects from a moving vehicle in real-time. Figure 2 shows an example of a frequency response (a) of a buried object and vibration images (b) of the buried object obtained for different vibration frequencies.
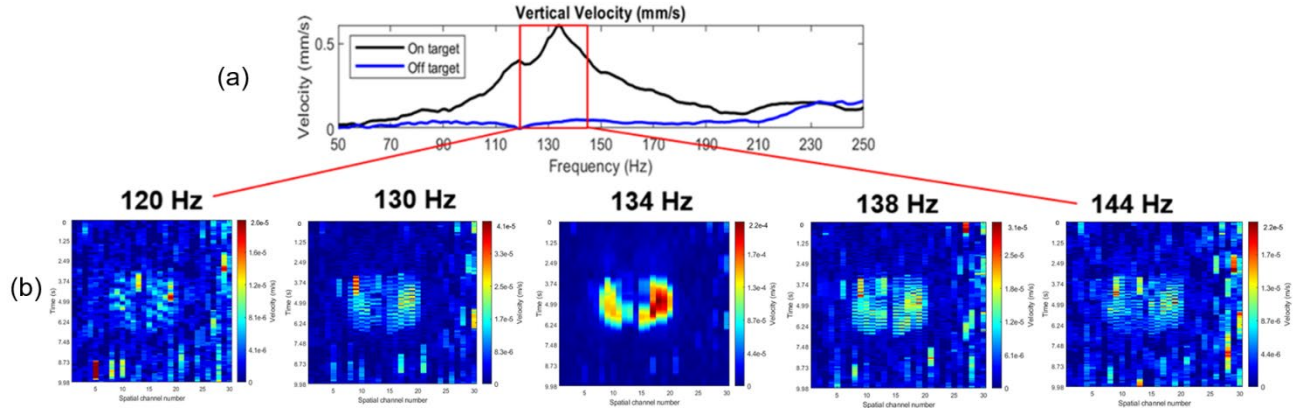

Figure 2. Frequency response (a) and vibration images (b) of a buried object for different frequencies

However, while the linear array LAMBDIS allows for fast simultaneous vibration measurement in a linear array of points on the object, increasing the speed of scanning results in reduced spatial resolution due to smaller number of sampling points in the scanning direction across the scanned area. As the result, the vibration image for a selected frequency band is getting blurred with the increase in the scanning speed. This is illustrated by Figure 3, which shows a vibration image of an object obtained with the different speed of beams: 0.1 m/s (a), 0.2 m/s (b), and 0.4 m/s (c).
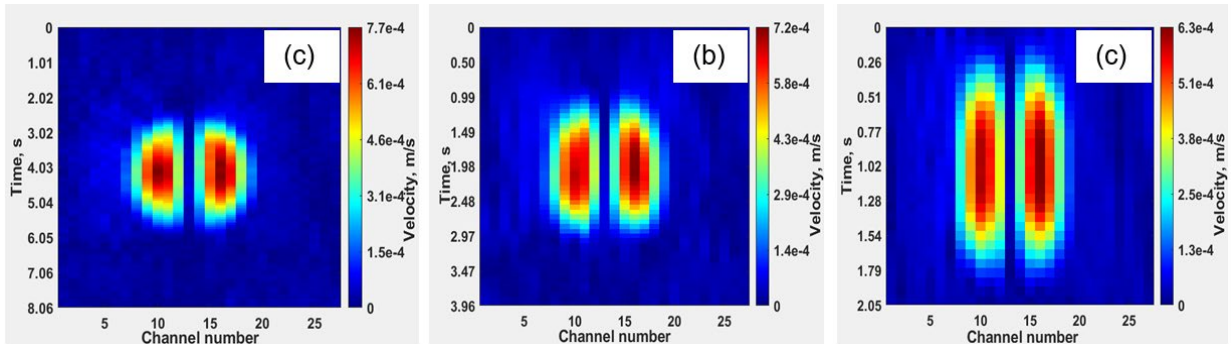

Figure 3. Vibration image of an object obtained with the different speed of beams: (a) – 0.1 m/s, (b) – 0.2 m/s, (c) – 0.4 m/s.

In order to overcome this limitation of the linear array LAMBDIS caused by the speed of scanning, we developed a 2D array LAMBDIS that employs a two-dimensional matrix array of laser beams and measures the whole area of interest at all points simultaneously.

3. 2D ARRAY LAMBDIS

The functional layout of the 2D array LAMBDIS is shown in Figure 4. The principle of operation of the 2D array LAMBDIS is based on the interference of light reflected from different points on the object surface illuminated with a matrix array of laser beams. A matrix array of 34 x 23 laser beams (only 7 x 5 beams are shown for clarity), is generated by a beam array generator of the sensor and focused onto an object, for example the ground surface.
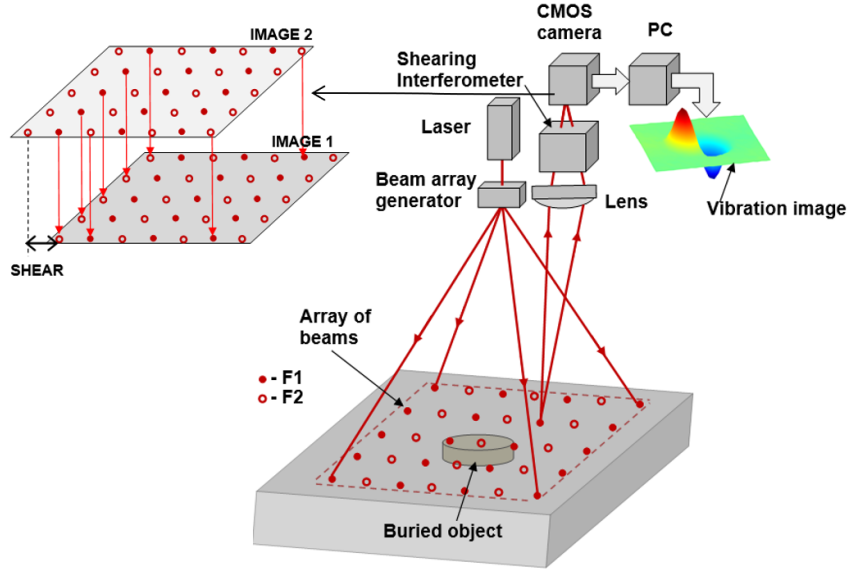

Figure 4. Functional layout of 2D array LAMBDIS

Beams in adjacent rows of the array have different optical frequencies, F_1 and F_2. A receiver lens and a shearing interferometer create two sheared images of the laser spots which are captured using a high-speed CMOS digital camera. The two images are sheared relative to each other by an odd number of intervals between neighboring rows of laser spots, so that the images of laser spots with frequency F_1 are overlapped with the images of laser spots with frequency F_2. The light from each pair of overlapped laser spots interfere on the high-speed digital CMOS camera, producing heterodyne signals with the carrier frequency $F_C = F_2 - F_1$. Digital demodulation of the heterodyne signals recorded by the camera provides computation of the velocity difference between the corresponding points on the object, and creates a vibration image of the measured area illuminated with the 2D array of beams. Motion of the sensor causes approximately the same Doppler shift for all beams, which is automatically subtracted from the measurements. Due to the simultaneous measurements of the vibration at all points of the interrogated area, the vibration phase between object points is preserved in the measurement results, which allows for computation of a vibration phase and the instantaneous velocity image. Measurement time can be significantly shorter as compared to measurements with a scanning linear array of beams of the linear array LAMBDIS.

3.1 Optical Layout

The optical schematic of the 2D array LAMBDIS is shown in Figure 5. The schematic works as follows. The laser beam is divided by a non-polarizing beamsplitter NPBS1 into two beams, each of them is frequency shifted by a different amount of F_1 = 110 MHz and F_2 = 110.01 MHz respectively by using acousto-optical modulators AOM1 and AOM2. Half-wave plates HWP1 and HWP2 are used to adjust the polarization direction of the laser beams. Frequency shifted beams are then combined together on the focusable beam expander BE using mirrors M1-M2 and a polarizing beamsplitter PBS. Frequency shifted beams pass through the beam expander, and are incident onto a diffractive optical element beam splitter DOE. The DOE splits frequency shifted beams into two identical linear arrays of beams. The DOE used in the design splits each incident beam into a matrix of 17 (X) x 23(Y) beams with the full pattern angles of 17.3 x 13.2 degrees. The angle between the two frequency shifted beams combined on the BE is adjusted in such a way that the two arrays of beams are sheared on the object surface in X-direction relative to each other by a half of an inter-beam spacing in the X-direction of the array, producing a combined 34 x 23 beam array, in which adjacent rows of beams in the X-direction have different optical frequencies F_1 and F_2. The beams are focused on the object surface by adjusting focus of the beam expander. The

light reflected back by the object surface enters a Michelson interferometer formed by non-polarizing beam splitting cube NPBS2 and mirrors M3 and M4. A lens and the Michelson interferometer create two laterally sheared images of laser spots on the object surface on the sensor of the CMOS camera. The two images are sheared relative to each other in X-direction by the odd number of intervals between neighboring rows of laser spots, so that the images of laser spots with frequency F_1 are overlapped with the images of laser spots with frequency F_2.
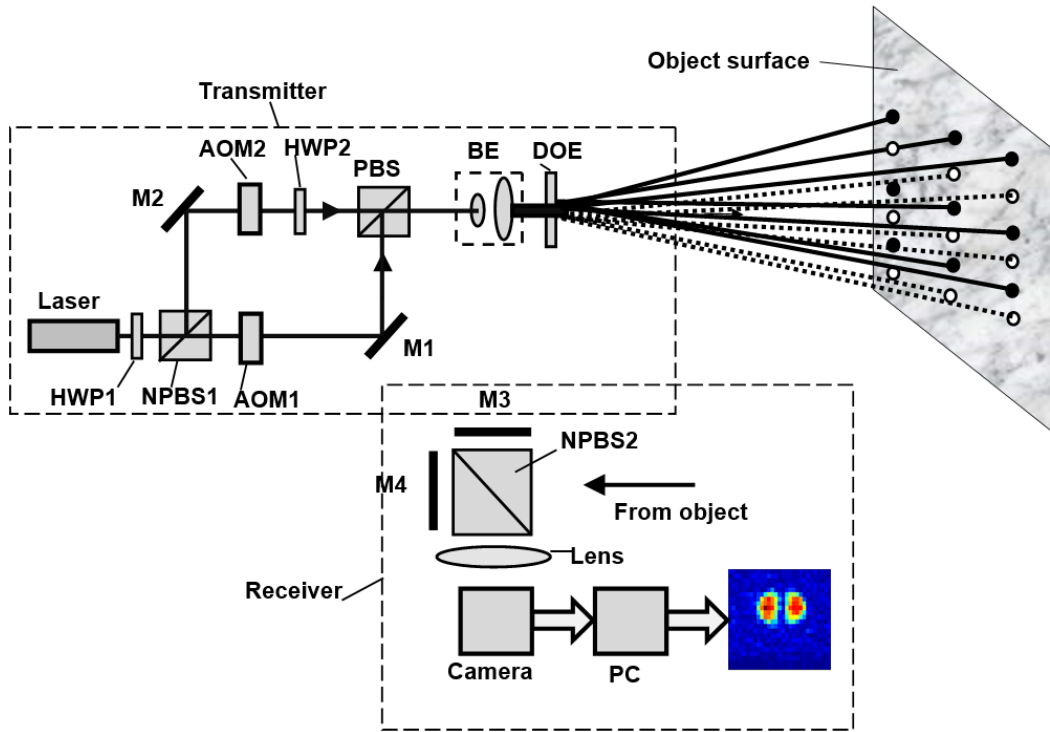

Figure 5. Optical layout of 2D array LAMBDIS

As a result, the light from each pair of corresponding laser spots on the object, which have different frequencies are mixed to interfere on the CMOS camera, producing heterodyne signals with the carrier frequency $F_C = 10$ kHz. The signals are recorded with the digital CMOS camera and demodulated on a personal computer (PC), which calculates the vibration velocity and phase difference for each pair of overlapped object points, and creates a vibration image of the measured area.

3.2 Signal processing

The block-diagram of the signal processing system is shown in Figure 6. The signal processing system consists of a high-speed digital CMOS camera and a PC running MATLAB. A digital high-speed CMOS monochrome camera (Phantom VEO 1310) was used in the 2D array LAMBDIS design. The camera has the following major features: 1280 x 960 pixels in the standard and 640 x 480 pixels in the binning mode, 18 μm pixel size, and bit depth 12 bits. The camera was operated at 100,000 fps frame rate, capturing images with a resolution of 320 x 240 pixels in the 2 x 2 square binning mode. The high-speed CMOS camera captures images and stores a selected quantity of frames in its internal flash memory. These saved images are subsequently transferred to a computer PC. MATLAB, operating on the PC, manages tasks such as image reading, pixel selection, and the generation of carrier signals for the selected pixels. These carrier signals then undergo frequency demodulation by using an I&Q demodulation technique[4,11], providing the computation of vibration velocity magnitude and phase. These steps contribute to the process of generating vibration images, which are finally presented on the computer monitor.
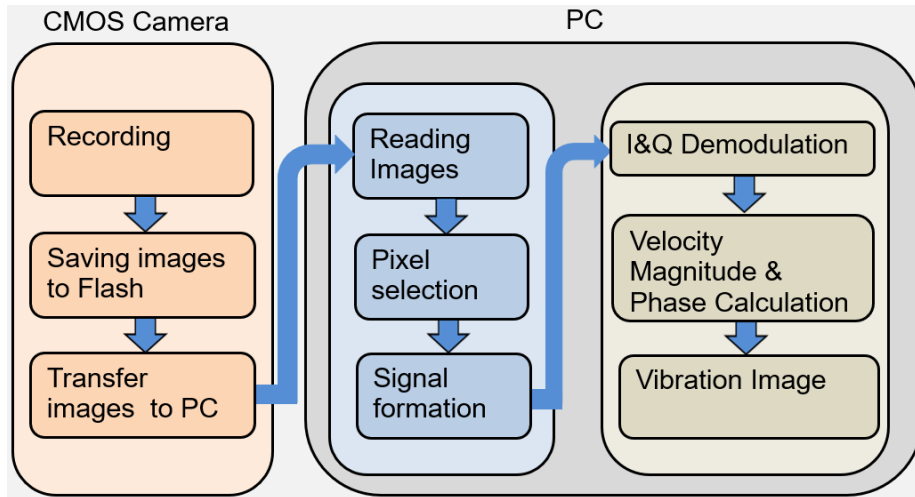
Figure 6. Block diagram of the signal processing system.

3.3 Experimental results

Performance of the 2D array LAMBDIS was experimentally verified in laboratory by measuring the vibration of a circumferentially clamped 250 mm diameter plate vibrating in a first axial-symmetric spatial mode at 100 Hz frequency. Figure 7 shows an example of a vibration velocity image using the velocity magnitude (a), vibration phase (b), and instantaneous vibration velocity (c) obtained by recording signals for 100 ms duration. The 2D array LAMBDIS measures vibration velocity difference between points on the object and provides a dynamic deformation gradient of the vibrating object surface. Simultaneous measurements at all points over the interrogated area allows for shorter measurement time in comparison to the linear array LAMBDIS. Figure 8 shows an example of vibration images of the vibrating plate for different measurement time: 100 ms, 10 ms, 5 ms, and 2.5 ms, and corresponding segments of the time-domain vibration signal.
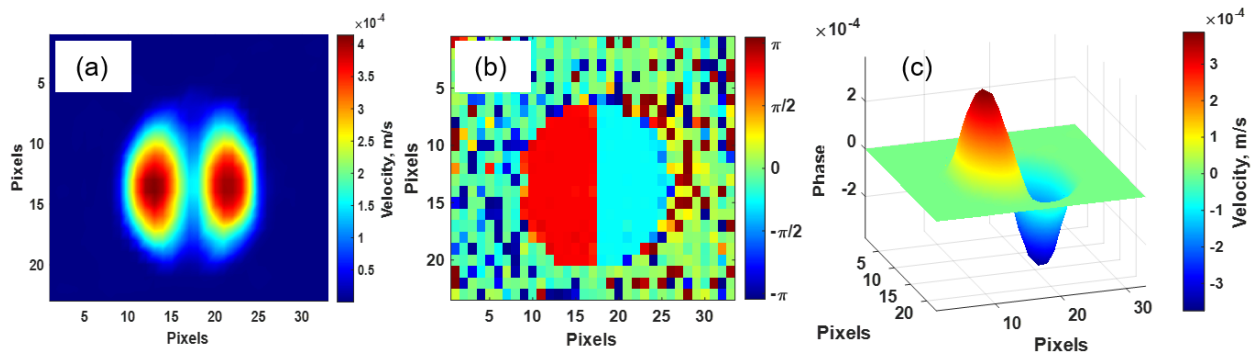

Figure 7. Vibration velocity image of a circumferentially clamped circular plate obtained with the 2D array LAMBDIS: (a) – velocity magnitude, (b) – vibration phase, (c) – instantaneous velocity.
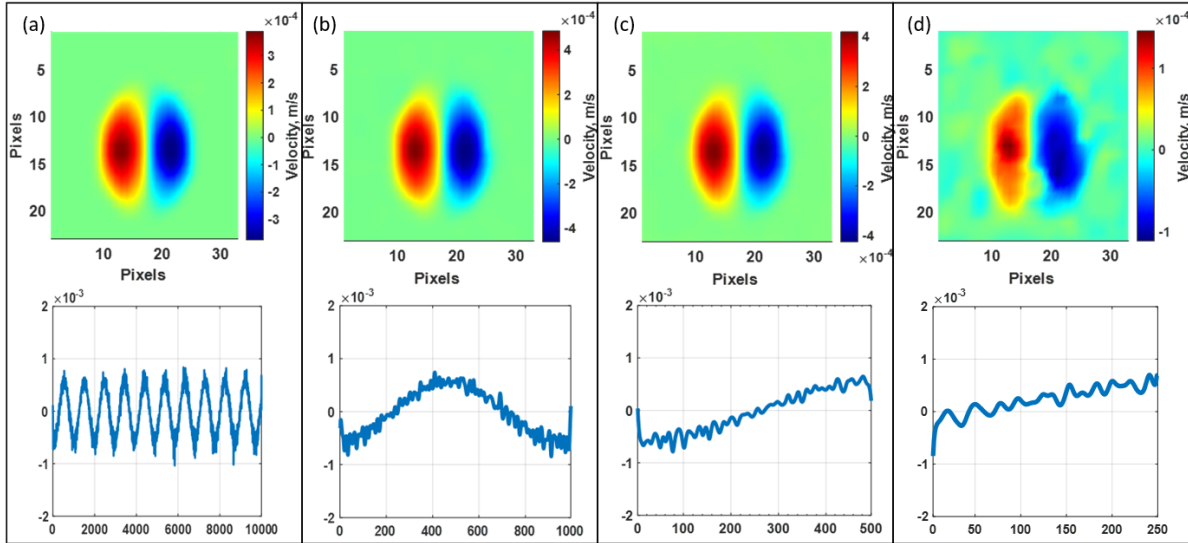
Figure 8. Vibration images of the circular plate vibrating at 100 Hz frequency and corresponding segments of the time-domain signal for different measurement times: (a) – 100 ms, (b) – 10 ms, (c) – 5 ms, and (d) – 2.5 ms,

One can see from Figure 8 that the vibration image quality stayed the same for the number of periods of the recorded vibration signal decreasing from 10 (a) down to 0.5 (c) periods. With the further reduction in the measurement time, for example down to 0.25 periods of vibration (d), the contrast of the vibration image decreases. So, the vibration image could be obtained for the time shorter than one period of vibration. Experiments on detection of buried objects were conducted in laboratory. A schematic of the experimental setup is shown in Figure 9 (a).
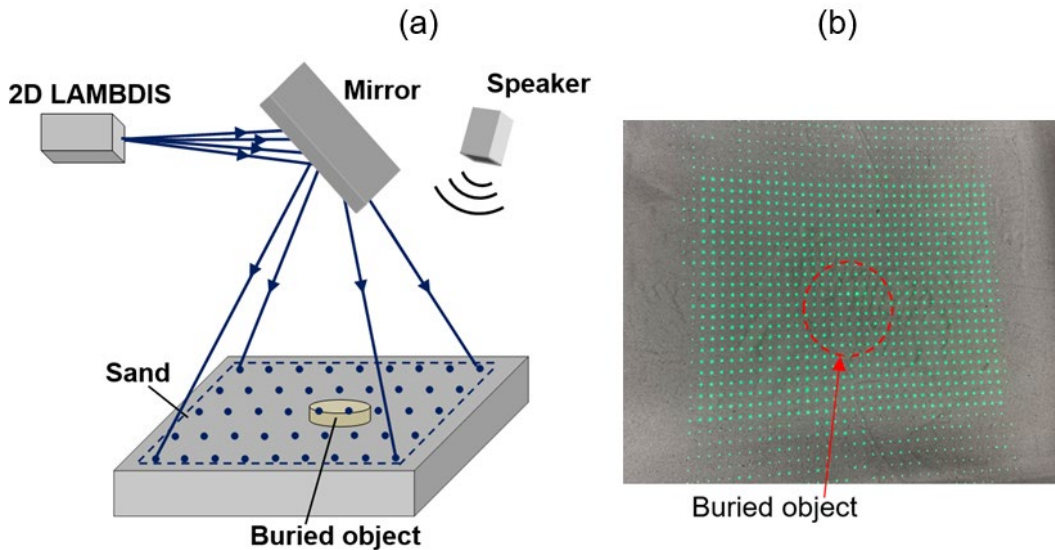
Figure 9. Schematic of the experimental setup for detection of a buried object (a) and a photograph of the array of beams on the sand surface over the buried object (b).

A fan of beams from the 2D array LAMBDIS setup was directed to the surface of a sand box with a buried object by using a folding mirror. The 34 x 23 array of beams covers 77 x 56 centimeters area on the sand surface. A photograph of the array of beams on the sand surface over the buried object is shown in Figure 9 (b). A loudspeaker was used to excite vibration of sand by using a broadband excitation signal in the frequency band from 40 to 200 Hz. A frequency response of the buried object measured with a scanning LDV is shown in Figure 10. One can see from Figure 10, that the buried object has a resonance at 100 Hz frequency.

Figure 10. Frequency response of the buried object

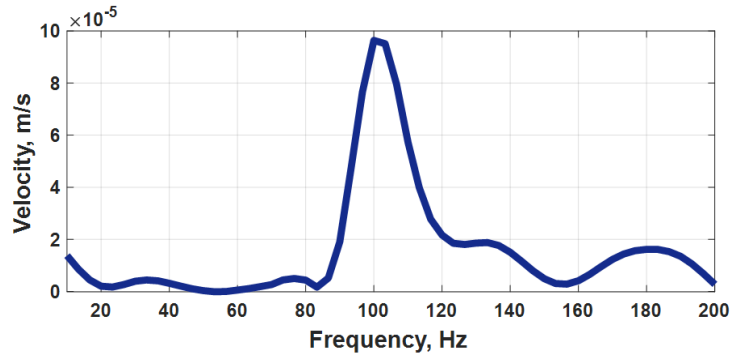The 2D array LAMBDIS measurements were completed for different measurement times. Figure 11 shows instantaneous vibration velocity images of the buried object for 100 Hz frequency obtained for different measurement times: (a) – 1 s, (b) – 100 ms, (c) – 40 ms, and (d) – 10 ms.
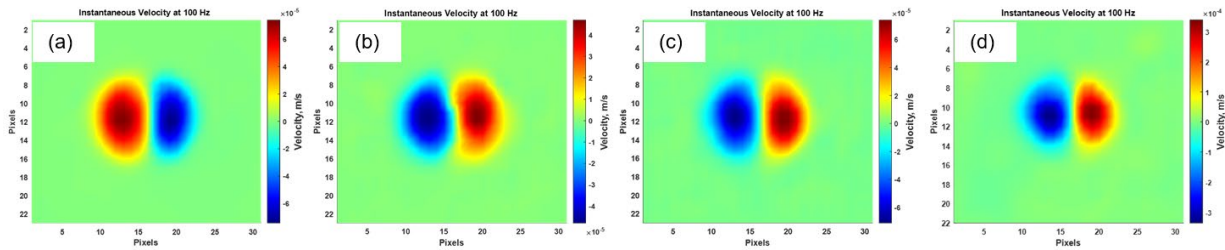

Figure 11. Instantaneous vibration velocity image of the buried object for 100 Hz frequency obtained for different measurement times: (a) – 1 s, (b) – 100 ms, (c) – 40 ms, and (d) – 10 ms.

One can see from Figure 11 that the buried object could be detected for the measurement time as short as 10 ms, which corresponds to one period of vibration.

4. CONCLUSIONS

A linear and a 2D array Laser Multi-Beam Differential Interferometric Sensors for whole-field vibration measurements have been developed. The sensors have low sensitivity to the motion of the sensor itself and allow operation from a moving platform. The linear array sensor measures velocity difference between points on the object illuminated with a linear array of 30 laser beams, and creates a vibration image of the object by scanning the linear array of beams in a transverse direction. The 2D array sensor simultaneously measures velocity difference between points on the object over the whole area illuminated with an array of 34 x 23 laser beams. Simultaneous measurements at all points allow for the fast recording of the vibration image, and allow the calculation of the vibration phase and instantaneous velocity. The 2D array sensor is capable of whole-field vibration measurements using times shorter than one period of vibration. The experiments demonstrated the ability of the 2D array sensor to provide a vibration image of the ground surface for laser-acoustic detection of buried objects for the measurement time as short as the period of vibration.

ACKNOWLEDGMENTS

REFERENCES

[1] Sabatier, J.M. and Xiang, N., "An Investigation of Acoustic-to-Seismic Coupling to Detect Buried Antitank Landmines", IEEE Transactions on Geoscience and Remote Sensing, 39(6), 1146-1154 (2001).

[2] Xiang, N. and Sabatier, J.M., "An Experimental Study on Antipersonnel Landmine Detection using Acoustic-to-Seismic Coupling", J. Acoust. Soc. Am., 113(3), 1333-1341 (2003).

[3] Xiang, N. and Sabatier, J.M., "Laser Doppler Vibrometer-Based Acoustic Landmine Detection Using the Fast M-Sequence Transform", IEEE Geoscience and Remote Sensing Letters, 1, 292-294 (2004).

[4] Aranchuk, V., Lal, A., Hess, C., and Sabatier, J.M., "Multi-beam laser Doppler vibrometer for landmine detection", Optical Engineering, 45 (10), 104302 (2006).

[5] Lal, A., Aranchuk, V., Doushkina, V, Hurtado, E., Hess, C., Kilpatrick, J., L'Esperance, D., Luo, N., Markov, V., Sabatier, J., Scott, E. "Advanced LDV instruments for buried landmine detection", Proc. SPIE 6217, 621715 (2006).

[6] Aranchuk, V., Lal, A. K., Zhang, H., Hess, C. F., and Sabatier, J.M., "Acoustic Sensor for Landmine Detection using a Continuously Scanning Multi-Beam LDV," Proc. SPIE 5415, 61-69 (2004).

[7] Aranchuk, V., Sabatier, J.M., Lal, A.K., Hess, C.F., Burgett, R.D., and O'Neill, M., "Multi-beam laser Doppler vibrometry for acoustic landmine detection using airborne and mechanically-coupled vibration," Proc. SPIE 5794, 624-631 (2005).

[8] Lal, A.K., Hess, C. F., Zhang, H., Hurtado, E., Aranchuk, V., Markov, V. B., and Mayo, W. T., "Whole-field laser vibrometer for buried landmine detection". Proc. SPIE 4742, 640-648 (2002).

[9] Aranchuk V., Johnson S., Aranchuk I., and Hickey C., "Laser Doppler multi-beam differential vibration sensor based on a line-scan CMOS digital camera for real-time buried object detection", Optics Express, 31(1), 235-247 (2023).

[10] Aranchuk V., Johnson S., Aranchuk I., and Hickey C., "Laser Doppler multi-beam differential vibration sensor for real-time continuous visualization of buried objects in multiple frequency bands", Applied Optics, 62(23), G12-G17 (2023).

[11] Aranchuk V., Aranchuk I., Carpenter B., and Hickey C., "Laser multibeam differential interferometric sensor for acoustic detection of buried objects," Opt. Eng. 62(2), 024103 (2023), doi: 10.1117/1.OE.62.2.024103.

Optical instrumentation development for tracking and reconnaissance of satellites

Fabrizio Silvestri[1], Eugenio di Iorio[1], Ralph C. Snel[1], Ivan Ferrario[1], Breno Perlingeiro[1], David L. Bakker[1], and Gustavo Castro do Amaral[1]

[1]High Tech Industry Unit, TNO, Stieltjesweg 1, Delft, 2628 CK, The Netherlands

ABSTRACT

The number of satellites is rapidly growing, hence the demand for increasingly precise knowledge of the satellites' orbital parameters is essential to avoid collisions, debris, and efficient use of the orbits. Recognizing, cataloging, and measuring with better confidence are actions crucial to preserve the health of crewed and uncrewed flying objects. Moreover, strategies to distinguish them may vary: TNO is developing suitable optical instrumentation for flying object reconnaissance along these two main paths. The satellite license plate (SLP) is a collaborative method based on a tag mounted on the satellite before launch. This plate consists of retroreflectors and wisely arranged bandpass filters. Therefore, it is passive and needs no power as opposed to an onboard radio beacon. Once a ground-based laser terminal illuminates the tag attached to the satellite, it sends back to Earth a signal encoding a unique identifier in the spectral domain. The current activities of TNO focus on proof-of-principle experiments in relevant environments (free-space tests over 2.5 km distances) and system design.

Keywords: Space Situational Awareness, Satellite Identification, Space Debris, Optical Ground Station

1. INTRODUCTION

Artificial satellites are essential in modern society, from television signal transmission to complex weather monitoring systems. As the deployment costs for satellite decrease, their number in orbit increases.[1] The forecast is to have more crowded orbits with tighter requirements for orbital station keeping. This leads to a traffic in space, which can be a risk for the missions due to collisions or need for frequent manouvering to avoid these collisions. Moreover, these spacecraft have a lifetime for their operation. After this, they become obsolete objects which can collide with others leaving fragments in space adrift.[2] Orbit-keeping is crucial to avoid collisions, and identifying the satellites allows the management of the overall courses. The current methodologies to identify satellites are based on technologies already available, like RADAR and optical observation.[3–5]

Currently at TNO we are developing several systems to favor Space Situational Awareness.[6] Amongst these we are looking into non-cooperative and cooperative methods. The non-cooperative method exploits the property of the light backscattered from the satellite, to retrieve information on shape and material of the satellite under observation. This method is described in more detail in the connected paper.[7] In this work, we present a cooperative solution for satellite identification. Our technique is based on a passive component, a license plate, containing an array of retroreflectors and a specific amount of band-pass filters. This device is attached to the surface of the satellite without being latched to the power-bus. To recognize the satellite, a ground station shall illuminate the license plate on the satellite using frequency multiplexed modulated light containing the center wavelengths of the band-pass filters. By demultiplexing and detecting the reflected light, the ground station can distinguish the number of filters for each wavelength. Therefore, it can identify the satellite license plate. Figure 1 illustrates the concept of the technique. Although the method presents limitations due to the cooperation necessity, it requires the same processing level as a satellite communication system and provides, in principle, more codes to identify satellites than other techniques.[8–10] Furthermore, the elements to build it are technologies available in the market.

Further author information:
Fabrizio Silvestri: E-mail: fabrizio.silvestri@tno.nl
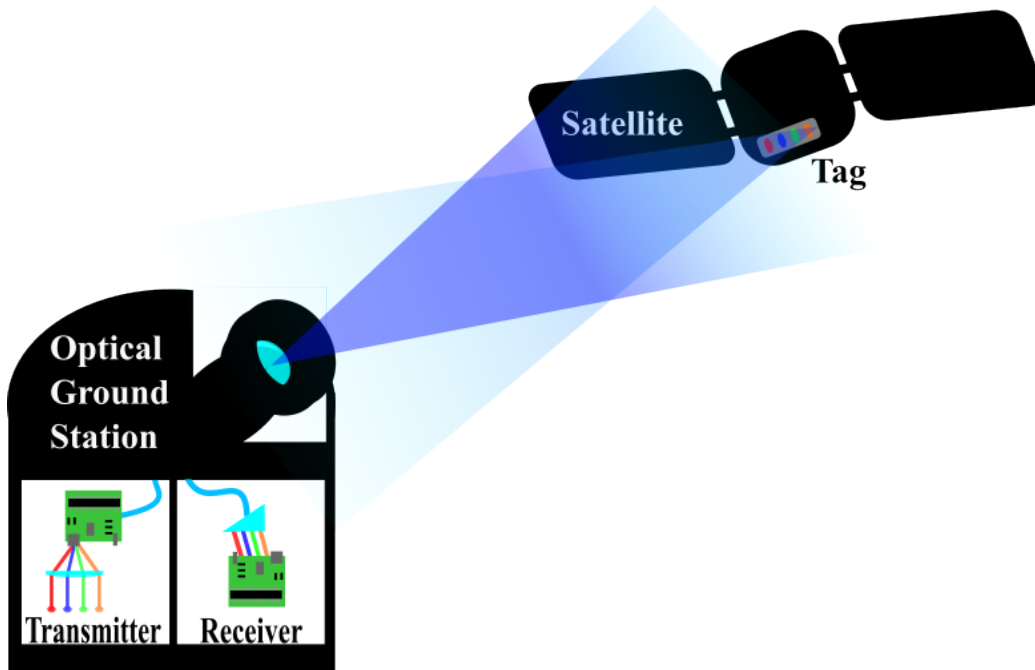Breno Perlingeiro: E-mail: breno.perlingeirocorea@tno.nl

Figure 1. The architecture for tracking and reconnaissance of satellites. Different laser sources produce light at different wavelengths, matching the center wavelengths of the filters in the license plate. The telescope illuminates the satellite and measures the reflected light. The detection system demultiplexes the incoming light and measures the spectral components allowing the system to decode and distinguish the number of filters in the license plate.

2. SATELLITE LICENSE PLATE

The recognition process of a satellite is divided into three steps: light multiplexing, modulation and transmission, reflected light reception, demultiplexing and decoding. The first step starts when the line of sight with a cooperative satellite is established. The ground station must be capable of generating multiple wavelengths, which are matched to the class of band-pass filters. The generated spectral channels are modulated simultaneously with a pulse to select in time the measurement window and to be able to calculate the distance. Then, these signals are combined and sent via a telescope to illuminate the cooperative satellite.

Then, the pulsed multiplexed light reflects off the license plate and propagates in the same direction as the transmitted light since the mirrors in the tags are retroreflectors. Note that the reflected intensity for each wavelength depends on the number of filters centered at that wavelength. This light is received by the same telescope used for transmission. The received signal is demultiplexed in its spectral channels. Subsequently, they are individually detected.

With the detected signal, we can calculate the average power on each channel. To decouple the effect of the attenuation to this measurement, we compare the power ratio for each channel to the total. Then, the result is compared to a constellation of nominal signatures that relates the ratios to the code. Using the maximum likelihood approach, a decoding software maps the measured signal to a specific code related to the number of filters in the license plate, identifying the satellite.

3. CONCLUSION AND FUTURE WORK

Recently, we have been developing a proof-of-principle experiment on a limited scale system with only 5 unique combinations. The test setup consists of a free space optical link of 2.45 km, with a prototype of the optical ground station on one end, and the retroreflecting tag on the other. This test, whose results are currently under processing, constitutes a representative case for the technology employed. Moreover, it serves as a validation of the current model used to decode the tag configuration from the actual optical signals receivded.

The future developments will be focused on the development of the full optical ground station, on the basis of the knowledge acquired during the ground to ground test. The optical ground station would be targeted to the tracking and identification of Low-Earth-Orbit (LEO) satellites, which potentially will mount a compatible license plate.

ACKNOWLEDGMENTS

REFERENCES

[1] Archive, E. S. A. . U. N. C. S., "Space safety - we're launching more than ever," (2021). https://www.esa.int/Space_Safety/Clean_Space/(archive)/0/(type)/image.

[2] Archive, E. S. A. . U. N. C. S., "Space care - satellite vs debris," (2021). https://www.esa.int/Space_Safety/Clean_Space/(archive)/0/(type)/image.

[3] D.A. Vallado, J. G., "Simulating space surveillance networks," (2011). AAS/AIAA Astrodynamic Specialist Conference.

[4] T. Flohrer, H. K., "Space surveillance and tracking in esa's ssa programme," (2017). 7th European Conference on Space Debris.

[5] Safety, E. S. A. . U. N. S., "Space surveillance and tracking - sst segment," (2021). https://www.esa.int/Space_Safety/Space_Surveillance_and_Tracking_-_SST_Segment.

[6] M.Noppen, I., "A dutch perspective on space situational awareness (ssa) for defence," (2023). SPIE Sensors+Imaging 2023.

[7] R. C. Snel, B. Vasilescu, E. D. I. P. P. J. L. I. F. F. S., "Spectropolarimetry for space object identification," (2023). SPIE Sensors+Imaging 2023.

[8] G. Kirchner, L. Grunwaldt, R. N. F. K. M. B. Z. Y. H. F. C. H., "Laser ranging to nano-satellites in leo orbits: Plans, issues, simulations," (2013). 18th Intenational Workshop on Laser Ranging.

[9] R.M. Holmes, C.T. Weaver, D. P., "Elroi: A license plate for satellites that anyone can read," (2018). 32nd Annual AIAA/USU - Conference on Small Satellites.

[10] N.Bartels, P.Allenspacher, D. B. D. E. W., "Space object identification via polarimetric satellite laser ranging," *Communications Engineering* **1**(5), 1–10 (2022).

Criterion for quantifying the amount of spatial and spectral information acquired by hyperspectral cameras

Andrei Fridman*[a], Torbjørn Skauli[b], Philippe Déliot[c], John R Gilchrist[d], Trond Løke[a]

[a]Norsk Elektro Optikk AS, Østensjøveien 34, N-0667 Oslo, Norway; [b]University of Oslo, Department of Technology Systems, Gunnar Randers vei 19, 2007 Kjeller, Norway; [c]ONERA, Theoretical and Applied Optics Department, ONERA centre de Toulouse, BP 74025, 2 avenue Edouard Belin, 31055 Toulouse Cedex 4, France; [d]Clyde HyperSpectral Imaging and Technology Ltd., 1 Aurora Avenue, Clydebank, G81 1BF, UK

ABSTRACT

Hyperspectral cameras are optical instruments that are designed for capturing spatial information from a scene in such a way that each pixel contains the spectrum of the corresponding small scene area. One of the important factors when assessing camera performance, is the amount of spatial and spectral information in the acquired hyperspectral data. Traditionally, these are directly communicated to users as spatial pixel count and spectral band count. However, depending on the width of the sampling point spread function (SPSF) and of the spectral response function (SRF), the amount of acquired information may be significantly different for two cameras – even if the specified pixel and band counts are the same. As a better indication of the amount of acquired information, the authors suggest using two new specifications in the camera specification sheet: equivalent pixel count (EPC) and equivalent band count (EBC). Both specifications are derived from an optical resolution criterion such as full width at half-maximum (FWHM) of the SPSF and SRF. With the pixel count being a universally known and intuitive concept, and FWHM being a well-established resolution criterion, EPC and EBC specifications would allow for a quick and easy comparison between cameras with significantly different degree of optical blur, pixel count, and band count. EBC and EPC are drafted to be included in the upcoming standard dedicated to hyperspectral imaging devices. The standard is currently being finalized by P4001 working group, sponsored by the IEEE Geoscience and Remote Sensing Society standards committee.

Keywords: hyperspectral, camera, imaging, spectroscopy, EBC, EPC, equivalent pixel count, equivalent band count

1. INTRODUCTION

Hyperspectral cameras are optical instruments that are designed for capturing spatial information from a scene in such a way that each pixel contains the spectrum of the corresponding small scene area[1]. The quality of the hyperspectral data acquired by such a camera, depends on many factors, and many parameters are usually needed to adequately describe camera performance.

Different manufacturers of hyperspectral cameras take different approaches in specifying their cameras. Recognizing shortcomings in current practices, the IEEE Standards Association has established Project 4001 "Standard for Characterization and Calibration of Ultraviolet through Shortwave Infrared (250 nm to 2500 nm) Hyperspectral Imaging Devices" (P4001), sponsored by the IEEE Geoscience and Remote Sensing Society standards committee. A main aspect of the P4001 standard[2], which is currently being finalized, is to define a set of characteristics that give a full description of hyperspectral camera performance. The criterion presented in this paper, is drafted to be a part of that set.

* fridman@neo.no

One of the important factors when assessing camera performance, is the amount of spatial and spectral information in the acquired hyperspectral data. Traditionally, these are directly communicated to users as spatial pixel count and spectral band count. In frame cameras, the pixel count is specified in x and y directions, whereas in pushbroom cameras, naturally, only the cross-track pixel count is specified. However, depending on the width of the sampling point spread function (SPSF) and of the spectral response function (SRF), the amount of the acquired information may be significantly different for any two cameras – even if the specified pixel and band counts are the same. Note, that in this context, when talking about the width of the SPSF or of the SRF, we are talking about the width of the function that contains the main part of the energy – such as full width at half-maximum (FWHM) or similar[2,3,4,5,6]. As the majority of users consider the specifications for pixel and band counts important when comparing cameras, there is a need for additional specifications that would more clearly indicate the amount of acquired spatial and spectral information (not merely the amount of data recorded on the data storage system during the image acquisition). This paper proposes such specifications.

2. WHY IS IT NEEDED?

Let us examine two simulated images (Fig.1). Both images have the same spatial pixel count 138x200 pixels, however, image (a) clearly contains significantly more spatial information than image (b). The difference is due to a much wider SPSF in the camera that captured image (b) compared to the SPSF of the camera that captured image (a) (a possibility of deblurring in post-processing is briefly discussed in Chapter 7).
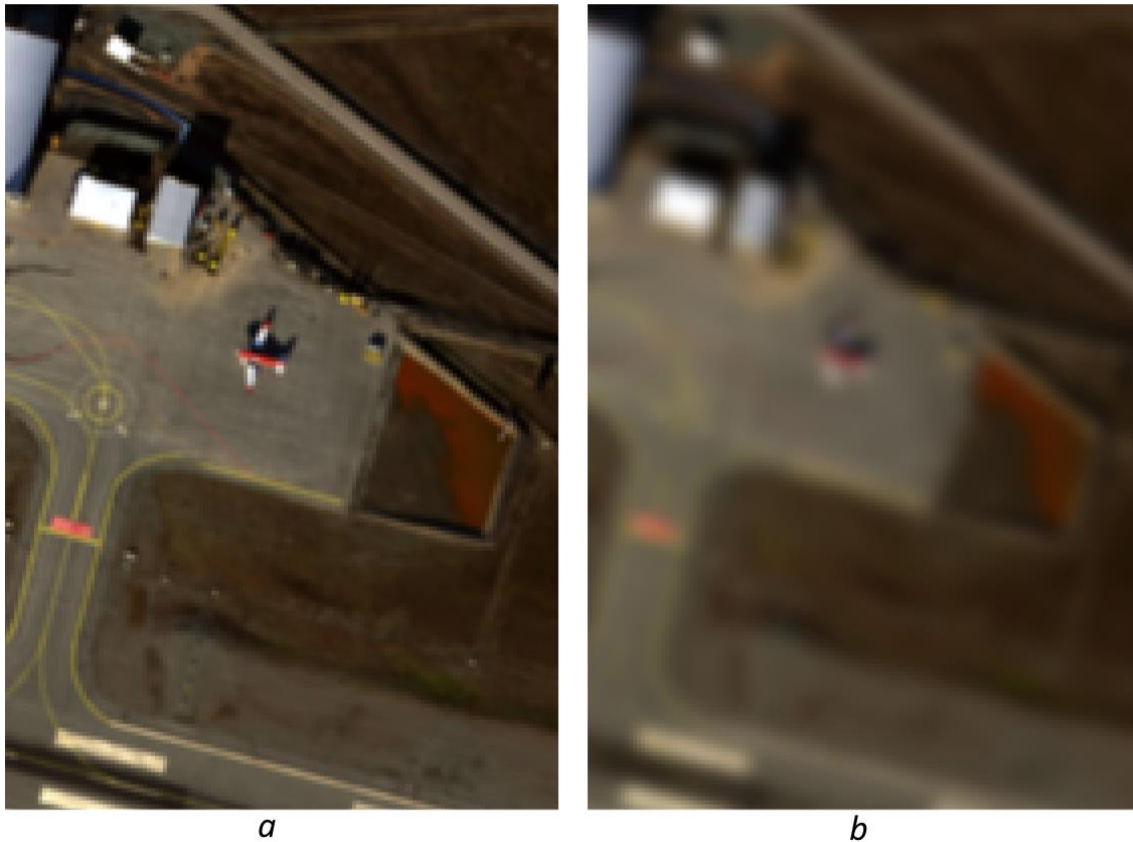

a   *b*

Fig.1. Two images 138x200 pixels each. Although the spatial pixel count is the same for both images, there is a lot more spatial information in image (a) due to a narrower SPSF in that image compared to image (b) that has a significantly wider SPSF. The optical blur is simulated.

Due to a much larger optical blur in image (b), the SPSFs of adjacent spatial pixels overlap significantly. This overlap makes the signals in two adjacent pixels quite similar, masking spatial variations in the scene. Therefore, the two images (a) and (b) from two different cameras are significantly different in terms of the amount of spatial information, despite having exactly the same spatial pixel count 138x200 pixels. It would be convenient to express the difference between these two cameras with a quantity that resembles the pixel count, since the pixel count is what many (perhaps most?) people rate cameras with.

The same is valid for the spectral dimension of hyperspectral images. Two cameras with the same spectral band count may acquire significantly different spectra in terms of the amount of spectral information, if these two cameras have significantly different SRF widths. Similarly to the situation with spatial pixels, for the spectral dimension it would be convenient to express the difference between the cameras with a quantity that resembles the spectral band count. Since the spectral dimension in a hyperspectral image is a single dimension, let us first discuss this quantity for the spectral data.

3. EQUIVALENT SPECTRAL BAND COUNT AS A CRITERION FOR THE AMOUNT OF SPECTRAL INFORMATION

The SRF describes the sensitivity of a spectral band (i.e., the sensitivity of a spectral pixel) as a function of wavelength[7]. The SRF takes into account the size of the light sensitive area of the sensor pixel, the optical PSF, and other factors (such as the slit width in pushbroom hyperspectral cameras). The SRF width relative to the spectral pixel pitch (i.e., relative to the spectral sampling) may vary significantly between hyperspectral cameras. Just like with spatial pixels (see Fig.1), for a defined spectral range, a camera with a wider SRF will show less detail in the spectra than a camera with a narrower SRF – even if the spectral band count is the same in both cameras. When SRFs are significantly wider than the spectral pixel pitch, the SRFs of two adjacent spectral pixels have a large overlap. Because of this, the signal in two adjacent spectral pixels (i.e., bands) becomes similar, masking narrow spectral features; see Fig.2 for the examples of narrow and wide SRFs, and their overlap for two adjacent spectral bands.

If two adjacent SRFs were far enough from each other, so that the overlap would be acceptably small, there would be less crosstalk between the two adjacent spectral channels, they would be more independent from one another, and narrow spectral features would be more visible. But what is far enough? There is a known and widely accepted criterion for spectral resolution – SRF Full Width at Half-Maximum (FWHM). Note that SRF FWHM is used either as the actual measured width at half-maximum, or as a more robust equivalent of FWHM[2,3,4,5,6] – the actual measured width at half-maximum depends on only three points on the SRF curve and is therefore not robust with respect to peak shape. In this paper, we refer to the width measure simply as FWHM, noting that it is beneficial to use a more robust measure of width than the basic FWHM.

Let us first consider the case when SRF FWHM is larger than the spectral pixel pitch – this is by far the most usual case. There will be a significant overlap between SRF FWHMs for adjacent spectral pixels. It can be argued that in this case the bands, i.e., the spectral pixels, are positioned too close to each other. An SPSF can be approximated with a rectangular function of the same width as the SRF FWHM. Referring to this approximation, we can adopt a pragmatic criterion for independence of two neighboring bands, namely that their FWHM width is not larger than the spectral sampling interval. According to this criterion, two SRFs are reasonably independent (i.e., two SRFs have small enough overlap to be considered independent) when their rectangular function approximations are positioned next to each other with no gap between them (Fig.3).
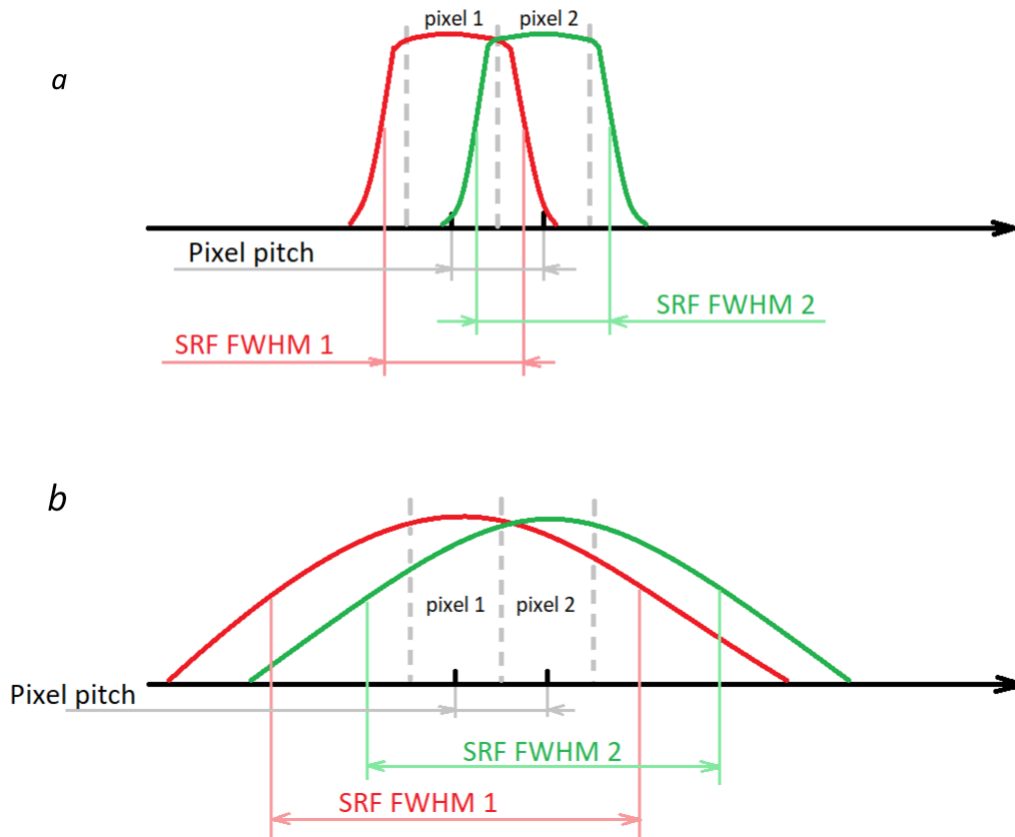
Fig.2 SRFs (red and green curves) of two adjacent spectral pixels (i.e., bands) are shown for two cameras. The camera shown in (a) has a quite narrow SRF compared to the pixel pitch. Because of this, the crosstalk between the two adjacent spectral pixels, that is indicated by the small overlap between the two SRFs, will be small. If the crosstalk between the bands is small, the camera will be capable of resolving narrow spectral features. The camera shown in (b), however, has very strong crosstalk between the two adjacent spectral pixels – the overlapping region for the two SRFs is very large. Therefore, the camera (b) will not be able to resolve narrow spectral features – the two adjacent pixels will capture more or less the same signal.

Fig.3. Two SRFs from Fig.2a approximated with rectangular functions of the same widths as the SRF FWHMs. These two rectangular functions are placed next to each other so that there is no gap or overlap between them.

Let us consider a narrow portion of the camera's spectral range $\Delta\lambda$ where the SRF FWHM for all bands is approximately the same. Let us, for now, denote the SRF FWHM for band k as $W_{SRF(k)}$ (valid for an S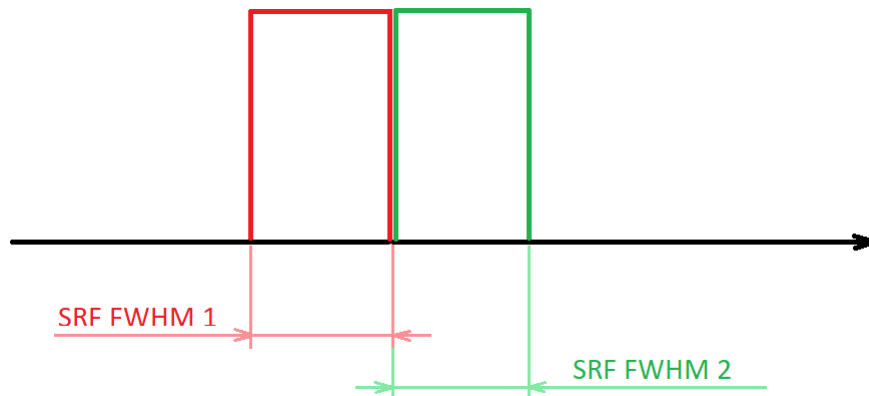RF FWHM larger than the spectral pixel pitch, see Eq.3 for details). To meet the condition of independence discussed above, the density of SRFs (the number of SRFs within a narrow $\Delta\lambda$) would need to be lower by a factor $1/W_{SRF(k)}$. By using this factor and the number of spectral bands that the camera has, we can derive an *equivalent band count* (EBC) that would indicate the amount of spectral information that the camera captures. Knowing the SRF FWHM for each spectral band, the equivalent band count can be calculated as

$$EBC = \sum_{k=1}^{L} \frac{1}{W_{SRF(k)}} \tag{1}$$

where L is the spectral band count and $W_{SRF(k)}$ is the SRF FWHM for band k. The SRF FWHM is expressed relative to the spectral pixel pitch, i.e., to the spectral sampling in the corresponding spectral band. $1/W_{SRF(k)}$ can be seen as a property of each spectral pixel that indicates how optically sharp that spectral pixel is. The equivalent band count can also be interpreted as the camera's band count L multiplied by the average inverse SRF FWHM:

$$EBC = \sum_{k=1}^{L} \frac{1}{W_{SRF(k)}} = L \times \frac{\sum_{k=1}^{L} \frac{1}{W_{SRF(k)}}}{L} \tag{2}$$

In some cases (for example, due to the sensor pixel design) the SRF FWHM may be narrower than the spectral pixel pitch. In such cases Eq.1 would generate a value for the equivalent band count that is larger than the actual band count in the recorded hyperspectral data. This would be illogical, since such a camera will actually have worse performance ("not seeing" parts of the spectral range very well) and will not provide more spectral information than the number of spectral samples (i.e., the band count) for a given spatial pixel. Therefore, $W_{SRF(k)}$ shall be interpreted as

$$W_{SRF(k)} = \begin{cases} 1, & SRF\ FWHM_k < 1 \\ SRF\ FWHM_k, & SRF\ FWHM_k \geq 1 \end{cases} \tag{3}$$

In a real camera, the SRF FWHM for band k is typically field dependent, i.e., it is different for different spatial pixels. The SRF for the band k in this chapter is the average of all SRFs for that band for all pixels within the camera's field of view. Note that small changes of the central wavelength in the band k across the camera's field of view (known as smile in pushbroom cameras) shall be interpreted as widening of the average SRF_k. Although, it can be argued that within a single spatial pixel such a shift does not blur the spectrum, it is common to assume that any band k in the final hyperspectral datacube (possibly after resampling) has the same central wavelength across the entire field of view. The situation may change in the future but, considering the current practices, it is correct to interpret these changes as widening of the average SRF_k.

The equivalent band count EBC shown in this chapter, is a single number that indicates the amount of spectral information in the hyperspectral data. It is derived from the SRF FWHM which is widely accepted as a criterion for spectral resolution.

All the equations shall be applied to the final datacube, i.e., after all intended processing (dark-frame correction, radiometric calibration, resampling, binning, etc.) of the acquired raw sensor data has been done.

4. EQUIVALENT PIXEL COUNT AS A CRITERION FOR THE AMOUNT OF SPATIAL INFORMATION

The SPSF describes the sensitivity of a spatial pixel as a function of the spatial coordinate. In several publications[7,8,9] the SPSF is also called the Spatial Response Function (SiRF) or the Angular Response Function (ARF). The SPSF takes into account the size of the light sensitive area of the sensor pixel, the optical PSF, and possible other factors specific for a camera architecture. The SPSF width compared to the spatial pixel pitch may vary significantly between hyperspectral cameras. A two-dimensional SPSF can be approximated[10] by its width in two orthogonal directions x and y. If, in a given camera, all SPSFs for all spatial pixels were the same, it would be possible to imagine the equivalent pixel count EPC that is derived from SPSF FWHM, as counting identical rectangular boxes fitted within the camera's field of view – each box having x- and y-dimensions SPSF $FWHM_x$ and SPSF $FWHM_y$ respectively. However, a typical camera will have different SPSF $FWHM_x$ and SPSF $FWHM_y$ for each pixel, and that direct geometrical interpretation will become somewhat chaotic. This can be fixed by treating the x- and y- directions independently from each other, and using the same averaging as in Chapter 3 for one of the dimensions. This approach is also useful for another reason: for a two-dimensional image, it is more practical to know the x- and y- pixel counts separately, rather than the total number of pixels in the image.

Let us consider a hyperspectral image MxN pixels. The equivalent pixel count *EPC* in each direction can then be calculated as:

$$EPC_x = \frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{M} \frac{1}{W_{SPSFx(i,j)}} \tag{4}$$

$$EPC_y = \frac{1}{M} \sum_{j=1}^{N} \sum_{i=1}^{M} \frac{1}{W_{SPSFy(i,j)}} \tag{5}$$

where *M* is the spatial pixel count in the x-direction, *N* is the spatial pixel count in the y-direction, $W_{SPSFx(i,j)}$ is the SPSF FWHM in the x-direction for the pixel *i,j* (valid for a SPSF FWHM larger than the spatial pixel pitch, see Eq.8 for details), and $W_{SPSFy(i,j)}$ is the SPSF FWHM in the y-direction for the pixel *i,j* (valid for a SPSF FWHM larger than the spatial pixel pitch, see Eq.8 for details). The SPSF FWHM is expressed relative to the spatial pixel pitch in the corresponding pixel. As it was with the spectral pixels, $1/W_{SPSF}$ can be seen as a property of each spatial pixel that indicates how optically sharp this pixel is in the x- or y-direction. Similarly to the spectral bands in the previous chapter, the equivalent pixel count in the x-direction can be interpreted as the camera's pixel count in the x-direction *M* multiplied by the average inverse SPSF $FWHM_x$:

$$EPC_x = M \times \frac{1}{MN} \sum_{j=1}^{N} \sum_{i=1}^{M} \frac{1}{W_{SPSFx(i,j)}} \tag{6}$$

And in the y-direction, the equivalent pixel count can be interpreted as the camera's pixel count in the y-direction *N* multiplied by the average inverse SPSF $FWHM_y$:

$$EPC_y = N \times \frac{1}{MN} \sum_{j=1}^{N} \sum_{i=1}^{M} \frac{1}{W_{SPSFy(i,j)}} \tag{7}$$

In some cases (for example, due to the sensor pixel design in combination with a very narrow optical point spread function) the SPSF FWHM may be narrower than the spatial pixel pitch. In such cases Eq.4 – Eq.7 would generate a value for the equivalent pixel count that is larger than the actual spatial pixel count in the recorded hyperspectral data. This would be illogical since a camera will not provide more spatial information than the number of spatial samples (i.e., the number of spatial pixels), and would actually have worse performance ("not seeing" parts of the scene very well). Therefore, W_{SPSF} for both the x- and y-directions shall be interpreted as:

$$W_{SPSF} = \begin{cases} 1, & SPSF\ FWHM < 1 \\ SPSF\ FWHM, & SPSF\ FWHM \geq 1 \end{cases} \tag{8}$$

where SPSF FWHM is taken in the corresponding direction for each corresponding spatial pixel.

In a real camera, the SPSF FWHM for any spatial pixel i,j is wavelength dependent, i.e., it is different in different spectral bands. The SPSF for any spatial pixel i,j in this chapter is the average of all SPSFs for that spatial pixel across the camera's spectral range. Note that small changes of the SPSF peak position as a function of wavelength (known as keystone in pushbroom cameras) shall be interpreted as widening of the average $SPSF_{i,j}$. That is because the vast majority of processing methods assume that keystone is absent in the final hyperspectral datacube. The situation may change in the future but, considering the current practices, it is correct to interpret these small spatial shifts of the SPSF peaks as widening of the average $SPSF_{i,j}$.

Similarly to the discussion in Chapter 3, SPSF FWHM can be used either as the actual measured width at half-maximum, or as a more robust equivalent of FWHM[2,3,4,5,6]. The authors recommend using a robust equivalent of FWHM for SPSF FWHM.

The equivalent pixel counts EPC_x and EPC_y shown in this chapter, are single numbers (one for the x-direction and the other one for the y-direction of the image) that indicate the amount of spatial information in the hyperspectral data in the respective spatial dimensions. It is derived from the SPSF FWHM which is widely accepted as a criterion for spatial resolution.

All the equations shall be applied to the final datacube, i.e., after all intended processing (dark-frame correction, radiometric calibration, resampling, binning, etc.) of the acquired raw sensor data has been done.

5. APPLYING EQUIVALENT PIXEL COUNT TO PUSHBROOM HYPERSPECTRAL CAMERAS

Pushbroom architecture is very important in hyperspectral imaging. Cameras that are built using this architecture, are common in many hyperspectral applications. These cameras require a scanning motion in one direction (called along-track direction) for forming a 2D image, because in a static position they have several pixels in the cross-track direction but only a single pixel in the along-track direction[1,7,9]. Moreover, the along-track pixel pitch is defined by the scanning speed and the camera's framerate, i.e., it is not locked to a single value in the camera hardware.

Let us first address the cross-track (i.e., the x-direction) EPC. This is a one-dimensional case, and the equation that describes it, will be identical to that for the spectral direction as described in Chapter 3:

$$EPC_x = \sum_{i=1}^{M} \frac{1}{W_{SPSFx(i)}} \tag{9}$$

where M is the spatial pixel count in the x-direction (i.e., cross-track), $W_{SPSFx(i)}$ is the SPSF FWHM in the x-direction for the pixel i (valid for a SPSF FWHM larger than the spatial pixel pitch, see Eq.11 for details). The SPSF FWHM is expressed relative to the spatial pixel pitch in the corresponding pixel.

The equivalent pixel count in the cross-track direction can be interpreted as the camera's pixel count M multiplied by the average inverse SPSF FWHM in the cross-track direction:

$$EPC_x = \sum_{i=1}^{M} \frac{1}{W_{SPSFx(i)}} = M \times \frac{\sum_{i=1}^{M} \frac{1}{W_{SPSFx(i)}}}{M} \tag{10}$$

To handle the cases where the SPSF FWHM is narrower than the pixel pitch, W_{SPSFx} shall be interpreted as:

$$W_{SPSFx(i)} = \begin{cases} 1, & SPSF\ FWHM_{x(i)} < 1 \\ SPSF\ FWHM_{x(i)}, & SPSF\ FWHM_{x(i)} \geq 1 \end{cases} \tag{11}$$

For the along-track EPC, Eq.5 can be used. However, there are two concepts that need to be addressed: the along-track pixel pitch and the along-track pixel count. Neither of these values are locked in the pushbroom camera hardware.

Let us first address the along-track pixel pitch. Although the along-track pixel pitch in a pushbroom camera can be set relatively freely (some pushbroom cameras, due to the predefined scanning motion, are exceptions), in reality most users prefer, and even expect, square pixels in the image. Also, comparison of spatial performance in the along- and cross-track directions is easier and more intuitive if the pixel pitch in these two directions is the same. So, for the purpose of calculating the equivalent pixel count in the along-track direction, we will assume square pixels, i.e., the same pixel pitch in the along- and cross-track directions (unless a different along-track pixel pitch is specified by the camera manufacturer).

Regarding the number of along-track pixels in the image, this number can be arbitrary: the longer a pushbroom camera runs uninterrupted, the longer image it manages to acquire in the along-track direction. Therefore, for pushbroom cameras, instead of setting a fixed along-track pixel count in an acquired image, it is more useful to specify a coefficient P that establishes a relationship between the number of the recorded along-track pixels N and the equivalent pixel count EPC_y:

$$EPC_y = P \times N \tag{12}$$

Of course, we already know what that coefficient is – for one along-track pixel line (formed by a single spatial camera pixel during the scanning motion) it is $1/W_{SPSFy}$. In a real camera $1/W_{SPSFy}$ is different for every along-track pixel line, because SPSF FWHM$_y$ is different for every camera pixel i. Utilizing a similar approach that was used for the spectral direction (Eq.2) and the spatial directions (Eq.6 and Eq.7), this can be addressed by averaging the inverse along-track SPSF FWHM for each pixel:

$$P = \frac{1}{M} \sum_{i=1}^{M} \frac{1}{W_{SPSFy(i)}} \tag{13}$$

where

$$W_{SPSFy(i)} = \begin{cases} 1, & SPSF\ FWHM_{y(i)} < 1 \\ SPSF\ FWHM_{y(i)}, & SPSF\ FWHM_{y(i)} \geq 1 \end{cases} \tag{14}$$

with SPSF FWHM taken in the along-track direction for each corresponding spatial pixel.

Now, if a pushbroom camera is specified as having P=0.7, for example, then a camera user will know that a 5000 pixels long image will have 3500 equivalent pixels along-track, a 10000 pixels long image will have 7000 equivalent pixels along-track, and so on.

6. CALCULATING EQUIVALENT PIXEL COUNT WHEN USING ABSOLUTE UNITS FOR SRF FWHM AND SPSF FWHM

So far, the SRF FWHM has been expressed relative to the spectral pixel pitch, i.e., to the spectral sampling in the corresponding spectral band. Similarly, SPSF FWHM has been expressed relative to the spatial pixel pitch in the corresponding spatial pixel. Depending on the context, it may be more convenient to express SRF FWHM and SPSF FWHM in absolute units (such as nm and mrad respectively) instead of expressing it relative to the spectral sampling (as in Eq.1 – Eq.3) and relative to the spatial pixel pitch (as in Eq.4 – Eq.14). Simply by substituting the relative FWHM in those equations with the ratio of the FWHM in absolute units to the sampling interval in the corresponding pixel, Eq.1 – Eq.14 in Chapters 3 – 5 can be adapted for using absolute units.

Let us first consider the spectral direction described in Chapter 3. $W_{SRF(k)}$ is:

$$W_{SRF(k)} = \frac{W'_{SRF(k)}}{S_k} \tag{15}$$

where $W'_{SRF(k)}$ is the SRF FWHM in the band k, and S_k is the spectral sampling in band k – both expressed in absolute units such as nanometers. Then, Eq.1 – Eq.3 can be rewritten as Eq.16 – Eq.18 respectively:

$$EBC = \sum_{k=1}^{L} \frac{S_k}{W'_{SRF(k)}} \tag{16}$$

$$EBC = \sum_{k=1}^{L} \frac{S_k}{W'_{SRF(k)}} = L \times \frac{\sum_{k=1}^{L} \frac{S_k}{W'_{SRF(k)}}}{L} \tag{17}$$

$$W'_{SRF(k)} = \begin{cases} S_k, & SRF\ FWHM'_k < S_k \\ SRF\ FWHM'_k, & SRF\ FWHM'_k \geq S_k \end{cases} \tag{18}$$

Note that the *SRF FWHM'* in Eq.18 is SRF FWHM expressed in absolute units.

Depending on the circumstances (the task at hand, provided camera specifications, etc.) it may be more convenient either:

a) expressing SRF FWHM relative to the spectral sampling in the corresponding band and using Eq.1 – Eq.3, or

b) expressing SRF FWHM in absolute units (such as nm) and using Eq.16 – Eq.18.

Let us now consider the spatial directions. In the x-direction, $W_{SPSFx(i,j)}$ is:

$$W_{SPSFx(i,j)} = \frac{W'_{SPSFx(i,j)}}{S_{x(i,j)}} \tag{19}$$

where $W'_{SPSFx(i,j)}$ is the SPSF FWHM in the x-direction in pixel i,j, and $S_{x(i,j)}$ is the spatial pixel sampling in the corresponding spatial pixel in the x-direction – both expressed in absolute units such as milliradians. Similarly, in the y-direction:

$$W_{SPSFy(i,j)} = \frac{W'_{SPSFy(i,j)}}{S_{y(i,j)}} \tag{20}$$

where $W'_{SPSFy(i,j)}$ is the SPSF FWHM in the y-direction in pixel i,j, and $S_{y(i,j)}$ is the spatial pixel sampling in the corresponding spatial pixel in the y-direction – both expressed in absolute units.

Then, Eq.4 – Eq.8 can be rewritten as Eq.21 – Eq.26:

$$EPC_x = \frac{1}{N} \sum_{j=1}^{N} \sum_{i=1}^{M} \frac{S_{x(i,j)}}{W'_{SPSFx(i,j)}} \tag{21}$$

$$EPC_y = \frac{1}{M} \sum_{j=1}^{N} \sum_{i=1}^{M} \frac{S_{y(i,j)}}{W'_{SPSFy(i,j)}} \tag{22}$$

The equivalent pixel count in the x-direction can be interpreted as the camera's pixel count in the x-direction M multiplied by the average ratio of the pixel pitch to the SPSF FWHM$_x$:

$$EPC_x = M \times \frac{1}{MN} \sum_{j=1}^{N} \sum_{i=1}^{M} \frac{S_{x(i,j)}}{W'_{SPSFx(i,j)}} \tag{23}$$

where

$$W'_{SPSFx(i,j)} = \begin{cases} S_{x(i,j)}, & SPSF\ FWHM'_{x(i,j)} < S_{x(i,j)} \\ SPSF\ FWHM'_{x(i,j)}, & SPSF\ FWHM'_{x(i,j)} \geq S_{x(i,j)} \end{cases} \tag{24}$$

with the *SPSF FWHM$_x$'* in Eq.24 expressed in absolute units.

In the y-direction, the equivalent pixel count can be interpreted as the camera's pixel count in the y-direction N multiplied by the average ratio of the pixel pitch to SPSF FWHM$_y$:

$$EPC_y = N \times \frac{1}{MN} \sum_{j=1}^{N} \sum_{i=1}^{M} \frac{S_{y(i,j)}}{W'_{SPSFy(i,j)}} \tag{25}$$

where

$$W'_{SPSFy(i,j)} = \begin{cases} S_{y(i,j)}, & SPSF\ FWHM'_{y(i,j)} < S_{y(i,j)} \\ SPSF\ FWHM'_{y(i,j)}, & SPSF\ FWHM'_{y(i,j)} \geq S_{y(i,j)} \end{cases} \tag{26}$$

with the *SPSF FWHM$_y$'* in Eq.26 expressed in absolute units.

The EPC in pushbroom cameras can be addressed in the same way – by using the substitutions Eq.19 and Eq.20. Then Eq.9 – Eq.11 become Eq.27 – Eq.29 respectively, and Eq.13, Eq.14 become Eq.30, Eq.31 respectively:

$$EPC_x = \sum_{i=1}^{M} \frac{S_{x(i)}}{W'_{SPSFx(i)}} \tag{27}$$

$$EPC_x = \sum_{i=1}^{M} \frac{S_{x(i)}}{W'_{SPSFx(i)}} = M \times \frac{\sum_{i=1}^{M} \frac{S_{x(i)}}{W'_{SPSFx(i)}}}{M} \tag{28}$$

$$W'_{SPSFx(i)} = \begin{cases} S_{x(i)}, & SPSF\ FWHM'_{x(i)} < S_{x(i)} \\ SPSF\ FWHM'_{x(i)}, & SPSF\ FWHM'_{x(i)} \geq S_{x(i)} \end{cases} \tag{29}$$

with the *SPSF FWHM$_x$'* in Eq.29 expressed in absolute units.

$$P = \frac{1}{M} \sum_{i=1}^{M} \frac{S_{y(i)}}{W'_{SPSFy(i)}} \tag{30}$$

$$W'_{SPSFy(i)} = \begin{cases} S_{y(i)}, & SPSF\ FWHM'_{y(i)} < S_{y(i)} \\ SPSF\ FWHM'_{y(i)}, & SPSF\ FWHM'_{y(i)} \geq S_{y(i)} \end{cases} \tag{31}$$

with the *SPSF FWHM$_y$'* in Eq.31 taken in the along-track direction for each corresponding spatial pixel and expressed in absolute units.

After calculating Eq.30, the along-track *EPC$_y$* can be calculated using Eq.12, as described in Chapter 5.

7. DISCUSSION

FWHM of SPSF and SRF are not only defined by the camera hardware. FWHM can be increased or decreased in post-processing. Such processing will change the signal-to-noise ratio (SNR) in the data and may introduce artifacts. A camera manufacturer may choose to use the native FWHM (as created by the optics and the sensor pixels) or to post-process the data for achieving the desired balance between the complexity of the optical design, FWHM, and SNR. When evaluating camera performance (or comparing two cameras) it is important to assess the quality of the final hyperspectral datacube, when all the intended processing has been applied to the acquired raw data. This approach has been recommended throughout this paper.

Intuitively, we may perceive a camera with large SPSFs or SRFs (compared to the pixel pitch) as inferior to another camera with significantly smaller SPSFs and SRFs. It is important to note that the reality is more nuanced. Cameras with large SRFs (relative to spectral sampling) can detect a position of a narrow absorption or emission line more precisely. Cameras with large SPSFs (relative to pixel pitch) allow for more precise resampling and better tracking of changes in the scene. It may also be easier to compare hyperspectral images acquired by different cameras if these cameras have wide SPSFs and SRFs. This is not surprising, since wide SPSFs and SRFs act as low pass filters for spatial frequencies, and we know from the Nyquist theorem that the pixel pitch needs to be small relative to the period of the highest spatial frequency if we need to fully sample a signal. So, there are advantages associated with optically blurry cameras.

On the other hand, a wide SPSF or SRF means that a signal from a small spatial object or a narrow spectral feature is distributed over more sensor pixels, and the detection limit will suffer compared to an optically sharp camera. And most importantly, we shall remember that, for a given number of pixels, the advantages associated with optically blurry cameras come at the cost of spatial and spectral details in the image. It is possible to compensate for the loss of the details by having more spatial pixels and more spectral bands, and by designing new optics for the higher pixel count. This is why the equivalent pixel count and the equivalent band count are very useful: they adequately quantify the amount of details in the image for sharp and blurry cameras with different physical pixel counts, making a comparison between cameras quick and easy.

8. CONCLUSION

Due to different SPSF and SRF widths in different hyperspectral cameras, pixel count and band count are not reliable indicators for the amount of spatial and spectral information that a hyperspectral camera captures. The authors suggest using the equivalent pixel count (EPC) and the equivalent band count (EBC) in camera specifications. With the pixel count being a universally known and intuitive concept, these specifications would allow quick and easy comparison between cameras with significantly different degrees of optical blur, pixel count, and band count.

If SRF FWHM and SPSF FWHM are expressed as a fraction of band pitch or pixel pitch, the following equations shall be used:

- For calculating EBC, Eq.1 or Eq.2, using the condition described in Eq.3;

- For calculating EPC, Eq.4, Eq.5 (or Eq.6, Eq.7), using the condition described in Eq.8;

- Additionally, for calculating EPC for pushbroom cameras

  o in the cross-track direction, Eq.9 or Eq.10, using the condition described in Eq.11;

  o in the along-track direction, Eq.12 and Eq.13, using the condition described in Eq.14.

If SRF FWHM and SPSF FWHM are expressed in absolute units (such as nanometers and milliradians respectively), the following equations shall be used:

- For calculating EBC, Eq.16 or Eq.17, using the condition described in Eq.18;

- For calculating EPC, Eq.21, Eq.22 or Eq.23, Eq.25 using the conditions described in Eq.24 and Eq.26;

- Additionally, for calculating EPC for pushbroom cameras:

  o in the cross-track direction for pushbroom cameras, Eq.27 or Eq.28 using the condition Eq. 29;

  o in the along-track direction for pushbroom cameras, Eq.30 and Eq.12 using the condition described in Eq.31.

REFERENCES

[1] Eismann M. T., [Hyperspectral Remote Sensing], SPIE Press, Bellingham, Washington USA, 2-7 and 331-334, (2012)

[2] IEEE-SA Project P4001 "Standard for Characterization and Calibration of Ultraviolet through Shortwave Infrared (250 nm to 2500 nm) Hyperspectral Imaging Devices"; https://standards.ieee.org/ieee/4001/7314/

[3] Palmer J. M. and Tomasko M. G., "Broadband radiometry with spectrally selective detectors", Optics Letters, Vol. 5, No. 5, (May 1980)

[4] Palmer J. M., "Effective bandwidths for LANDSAT-4 and LANDSAT-D' multispectral scanner and thematic mapper subsystems", IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-22, No. 3, (May 1984)

[5] CIE233:2019 "Calibration, Characterization and Use of Array Spectroradiometers"

[6] Ewald F., Kölling T., Baumgartner A., Zinner T., and Mayer B., "Design and characterization of specMACS, a multipurpose hyperspectral cloud and sky imager", Atmos. Meas. Tech., 9, 2015–2042, (2016), doi:10.5194/amt-9-2015-2016

[7] Mouroulis P., Green R.O., and Chrien T.G., "Design of pushbroom imaging spectrometers for optimum recovery of spectroscopic and spatial information", Applied Optics, Vol. 39, No. 13, (1 May 2000)

[8] Skauli T., "An upper-bound metric for characterizing spectral and spatial coregistration errors in spectral imaging", Optics Express, Vol. 20, No. 2, (16 January 2012)

[9] Baumgartner A., "Traceable imaging spectrometer calibration and transformation of geometric and spectral pixel properties", PhD-thesis, Osnabrück University, (2021)

[10] Torkildsen H.E. and Skauli T., "Full characterization of spatial coregistration errors and spatial resolution in spectral imagers", Optics Letters, Vol. 43, No. 16, (15 August 2018)

A New Remote Sensing Electro-Optical System on Low-light Target

Ge Jingjing*[a]

[a] Beijing Institute of Space Mechanics & Electricity, Beijing, China 100094

ABSTRACT

In view of the application of optical remote sensing disaster emergency rescue under the complex environment of low illumination at night, the optical remote sensing imaging technology under the condition of low illumination and low signal-to-noise ratio was studied. The compressed sensing technology of thin film diffraction grating primary mirror is used to realize large-aperture optical acquisition. The sensing ability of large dynamic range is increased by Geiger pattern imaging technology, and the dim and weak targets are identified by semantic sensing algorithm. The system realizes target recognition under the condition of extremely low image signal-to-noise ratio through the design of the new system's main mirror flattening and the aliasing compression and decoupling of spatial information and spectral information. The technology has completed space-based scheme design and desktop principle verification tests, and the spectral resolution reaches 5nm, realizing fast target search and recognition.

Keywords: Low-light level imaging, Computational remote sensing, Diffraction grating primary mirror, Emergency disaster reduction

1. INTRODUCTION

Space remote sensing satellites can cover, identify, measure, search and discover a large area of the ground by flying in high orbit. Especially in complex environments, optical remote sensing plays an important role in disaster monitoring and emergency rescue [1]. However, the traditional target recognition technology is mainly realized by extracting the target image from the visible light image of the measured area, and this method has certain limitations, because the acquisition of visible light image is often affected by the environment, and can not realize all-weather work. At present, the on-orbit imaging mainly relies on the solar reflection spectrum during the day for detection. Under the conditions of low illumination and low signal-to-noise ratio at night and dawn and dusk, there is not enough light energy to image, and its detection ability is greatly limited. There are a lot of reports that night detection can be used for major event assessment, urbanization detection, application of remote sensing data of night light [2], research on night city light comparison before and after the Syrian war [3], fire source detection [4] and other applications have received more and more attention.

Luminous remote sensing originated in the 1970s, after nearly half a century of development, at present, there are mainly three types of low-light level loads operating on orbit abroad. They are the Line of Service Scanning System (OLS) [5], the National Polar Orbit Operational Environmental Satellite System Preparation Project (NPP) [6], and the Visible Infrared Imager/Radiometer Group (VIIRS) [7] on NOAA-20, which is mounted on the U.S. Defense Meteorological Satellite (DMSP). At present, the only instrument in the world that can obtain 1/4 moon illumination at night and visible light cloud images under dawn and dusk conditions is the Line of Business scanner (OLS) loaded on the U.S. Defense Meteorological Satellite (DMSP), whose low-light channel can obtain 1/4 moonlight illumination and snow, sea ice, low cloud fog and surface feature images under dawn and dusk conditions, and can also be used under lower low-light conditions. The acquisition of remote sensing images such as lights, lightning, and wildfires, extending the detection capability during the day to the night, has also achieved great potential in Earth observation and other applications, but due to the spatial resolution of 2.7km (smooth mode) and 0.55km (fine mode), the effect cannot meet the large-scale fine detection at the 100-meter scale.

As we all know, the target recognition needs to be locked and discriminated. On the basis of discovery, the target identification needs to be completed, and the geometric size and contour information of the target need to be further extracted to improve the target recognition and positioning accuracy. Compared with natural targets such as landslides, volcanoes, earthquakes, and sudden increases in floods, which are characterized by small target changes (the size is about ten meters to 30 meters), the resolution required to confirm the target needs to reach about 16 times that of the

target, so in order to achieve further identification of the target, the load needs to have a high-resolution detection capability within 5m. It has the ability to extract the geometric size and contour information of the general natural disaster target movement.

In addition, the recognition ability of the system will be significantly affected when there is an occluded object. Therefore, in order to extract effective target signal light from the region to be measured, it is impossible to use only the image acquisition target recognition based on light intensity. The object can be detected and identified through the spectral reflection "fingerprint" characteristics of the object [8]. The target recognition technology using spectral characteristics is a method to establish the spectral image data cube of the two-dimensional image data and the spectral distribution data at the corresponding point position. The spectral imaging method can realize all-weather detection and has stronger feature recognition ability.

2. SYSTEM DESIGN

In order to meet the detection application requirements of high-resolution imaging and feature target recognition under low illumination conditions, the imaging system has the comprehensive detection capability of high spatial resolution, high spectral resolution, large dynamic range and high sensitivity (photon level), which is designed for all aspects of the imaging link such as optical collection, photoelectric conversion and data processing. Achieve high-precision information inversion under low illumination conditions. The specific detection link is shown in the figure below.
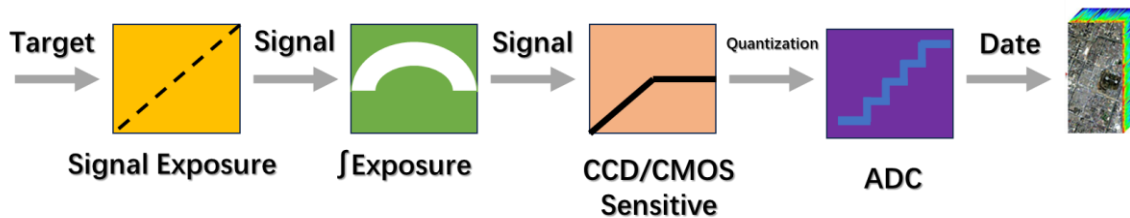

Figure 1. High-resolution imaging, feature target recognition and detection system

We use a long thin grating primary mirror with an optical aperture of 3m×20m to collect energy (in which the surface density of the grating is <5kg/m2 and the axial compression ratio is ≤0.25). The grating primary mirror can be folded in one direction during emission and unfolded in orbit. At the same time, in the imaging aspect, a large-dynamic range ultra-low noise ultra-high gain SPAD device is used for low-light level detection, and then the target under the condition of extremely low image signal-to-noise ratio is identified stably, and the detection capability of high spatial resolution and high spectral resolution is achieved under the condition of 1/4 month of 4×10-9 W/cm2/Sr.

The system light path of the imaging principle of the light grating primary mirror is shown in the figure below. The light emitted by the object is received by the grating and dispersed backward, and is imaged to the primary image plane by the objective mirror. A braid slit is placed on the image surface to modulate the light at different spatial positions. After that, the light is collimated by the collimator, incident on the prism-raster-prism (PGP) component for secondary dispersion, and finally imaged by the imaging mirror at the focal plane of the detector.
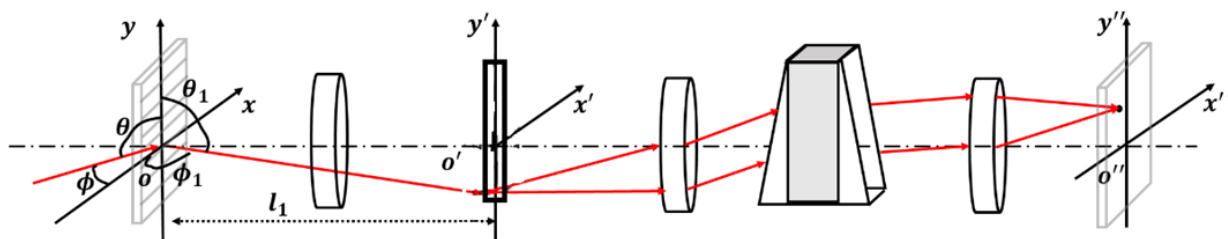

Figure 2. Schematic diagram of system optical path

Assuming that the direction of grating and prism dispersion is opposite, that is, the direction of grating dispersion is vertical upward dispersion, and the direction of prism dispersion is vertical downward dispersion, and the dispersion coefficient of grating is twice that of prism dispersion coefficient, then each spectral band in the spectral data is migrated upward through the grating pixel and downward through the prism pixel, and the corresponding spectral band number in the spectral data is marked. Based on the above assumptions, the mathematical model is simplified. The schematic diagram of the simplified version of spectral data and dispersion is shown in the figure.
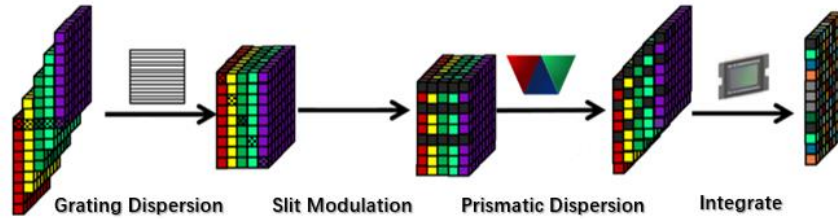

Figure 3. Schematic diagram of simplified forward model

The whole system inputs the spectral data cube of the corresponding size. Each channel only takes the image block of the corresponding size, and the rest is filled with 0. The position of each spectral band image block is moved down by pixels successively. After the raster upward dispersion, the originally slanted parallelepiped becomes a cube, and the image of each channel is the part of the input image that does not contain 0 filling, and the change of position of the pixels in the input image block after dispersion is marked in the diagram using a checkerboard pattern. The data cube after grating dispersion is modulated by a slit and then secondary dispersion is carried out by a splitter prism. The direction of the secondary dispersion is downward dispersion. The size of the data cube obtained after dispersion is: the effective region of each spectral band remains unchanged, and the rest of the data cube is filled with 0.

The two factors that limit the sensitivity are detector system noise and gain. The detector system noise mainly includes circuit reading noise and device intrinsic dark noise. In the photoelectric detection of CCD and CMOS devices, the photoelectric charge has a large reading noise in the quantization reading, and the influence is particularly obvious in the weak light detection. By studying the imaging mode of photon counting, the photoelectron is directly converted into digital pulse for counting imaging, which is fundamentally different from the traditional CCD and CMOS imaging mode, and can completely eliminate the signal reading noise. The upper limit of dynamic range of traditional CCD and CMOS imaging devices is limited by the full well charge, which is determined by the area of the charge storage area, and the dynamic range is limited under a certain device area. By using the photon counting imaging method, there is no limitation of full well charge in a single exposure, only limited by saturation count. The saturation meter value of SPAD device is much higher than the traditional trap charge storage value, and the dynamic range can exceed 150dB.

In order to solve the problem that the structural information of the target is seriously damaged under the condition of extremely low signal-to-noise ratio, the structural decomposition of the target is firstly carried out to form hierarchical semantic information representation, including the modeling and representation of semantic objects and the characterization of semantic relations. Then the semantic object detection is carried out on the input image with extremely low signal-to-noise ratio. The semantic object detection network is constructed by the fusion of semantic knowledge module and semantic network. Then the semantic relationship detection network based on recurrent neural network is used to detect the relationship between semantic objects. For the predicted semantic objects and their relationships, we further use the object detection and recognition network based on data enhancement and graph convolutional neural network to get recognition results. In order to further improve the accuracy, we extract the spectral features of the detected semantic objects and integrate them into the semantic object detection network and semantic relation detection network to highlight the spatial structure relationship of the objects, suppress the background and noise interference, and improve the recognition accuracy of the targets.

3. OPTICAL DESIGN

The system adopts a new weak light spectral imaging system, and its imaging technology principle is as follows: collimated beams with different field angles represent targets with different angles and orientations, which contain the

spectral information of the different targets. At a specific diffraction Angle, the outgoing beam corresponding to each field Angle only has a specific wavelength of light energy, and the beam is combined after incident on the plane transmission grating. The primary image is imaged by the optical focusing system, and a coding plate is placed on the primary image surface. At this time, the light energy contained is the superposition of the light energy of each specific wavelength corresponding to each field Angle. The intensity information of specific wavelength in different field of view can be obtained by placing the dispersion device again behind the coding plate and realizing spectral detection by means of computational optics. Through the on-orbit push sweep of the system motion, the light energy distribution of all wavelengths at all angles of view can be reconstructed, that is, the spectral imaging detection is realized. The specific imaging schematic diagram is shown in the following figure.
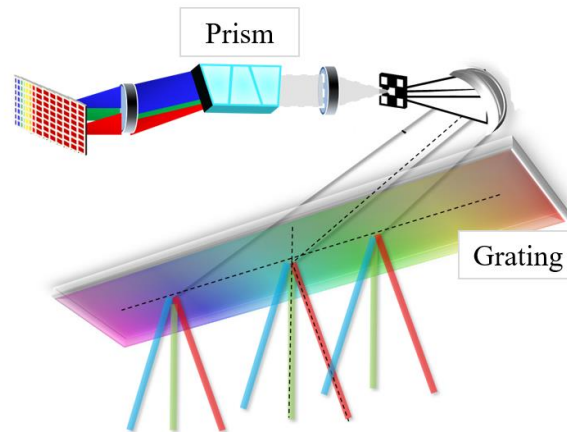

Figure 4. Schematic diagram of the system scheme of the new weak light spectral imaging system

Generally, the grating is mainly used as a spectral component in the optical system, and does not assume the imaging function in the traditional imaging system, so the optimization of the optical system can still be evaluated and optimized by MTF and other evaluation standards.

The grating primary mirror proposed by us takes the grating as the primary mirror of the large-aperture optical system and uses it in reverse. Although its grating expression is consistent with the traditional grating, the grating participates in the imaging of the system. In the process of system design, the grating parameters will have a great influence on the image quality of the system. Moreover, the system is a mixed system of diffraction and reflection, and the reflection and diffraction problems should be considered comprehensively in the design process. The multi-structure joint optimization simulation method is adopted, that is, a set of optical system is established in each band, and the mirror parameters of each system are the same, and the parameters of the primary mirror of the reverse grating are determined by multi-structure joint optimization.
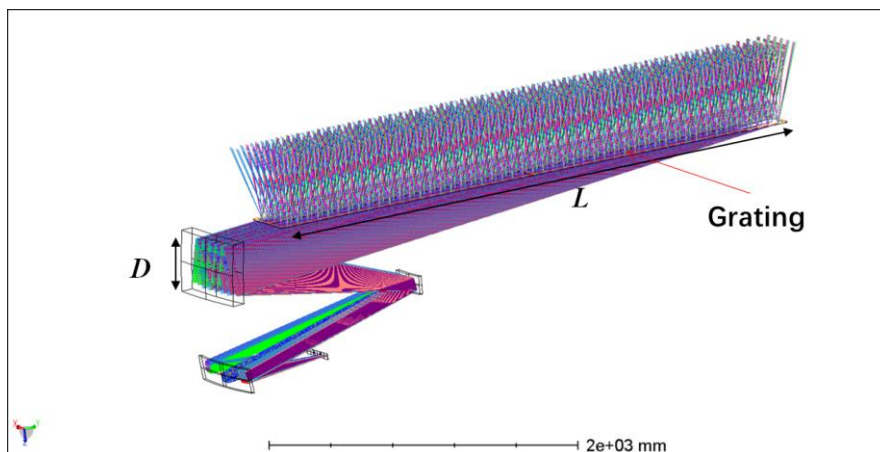

Figure 5. Schematic diagram of reverse grating imaging system

For the expansion error of the primary mirror of the reverse grating, the key link of the expansion error is analyzed first, the key link is modeled, the impact of each key link of the error on the image quality is analyzed separately, and the direct coupling factors of each error link are considered to correct the error model. According to the input conditions, the deformation of the thin film grating is analyzed. The deformation of the thin film grating is brought into the optical simulation system to analyze the effect of the resulting deformation on the image quality. On the basis of perfecting the model, a statistical analysis of the dynamic thermal coupling of the film grating was carried out to calculate the effect of the errors caused by the dynamic thermal on the image quality.

4. EXPERIMENT

A novel spectral imaging device based on the reverse grating primary mirror comprises a reverse grating primary mirror module, a receiving optical module, a dispersion module, a focusing module, a detection module, etc., and a target scene, image acquisition, detection equipment, etc. Among them, the function of the reverse grating main mirror is to collect light, the function of the receiving optical module is to converge light, the function of the dispersion module is to produce dispersion effect, the function of the focusing module is to focus the dispersive light again, and the detection module detects and images.
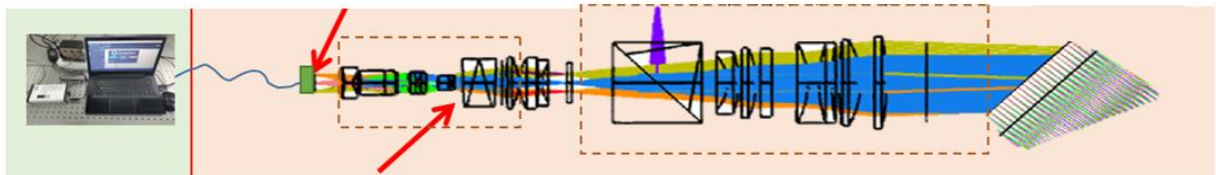
(a) Device composition Diagram
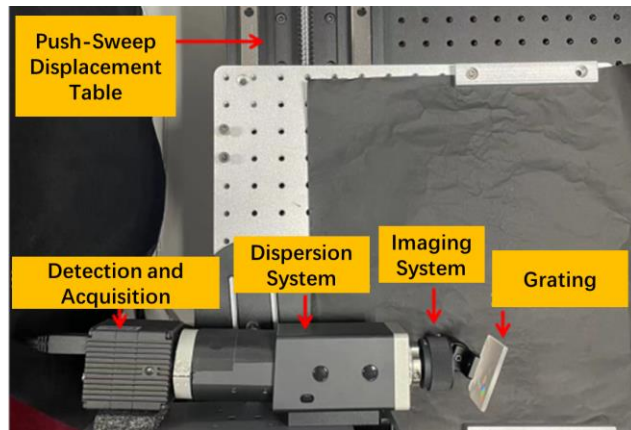
(b) Imaging device diagram

Figure 6. A new spectral imaging device for the primary mirror of the reverse grating

The tabletop experimental device was set up as shown in the figure, the dispersion module, detector and displacement table were fixed, and the lens was installed. During the focusing test, adjust the lens focal length to the best position by observing the output display of the imaging surface; Add the primary mirror grating, adjust the inclination Angle of the grating, adjust the position of the target until the entire target dispersion image enters the back-end optical system and is received by the detector, and complete the spectral dislocation synthesis by data processing.

According to the imaging results, the dispersive single spectral map of the primary image position was corrected by spectral translation, and the spectral images of all spectral segments were obtained as shown in the figure below. The spectrum ranges from 500nm-700nm, the number of spectral segments is 40, and the spectral resolution is 5nm.
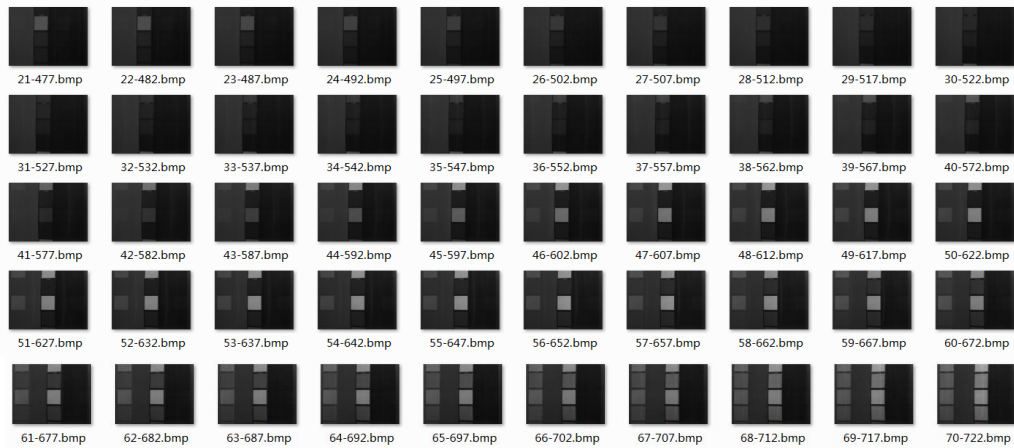
Figure 7. Spectral images of each spectrum segment
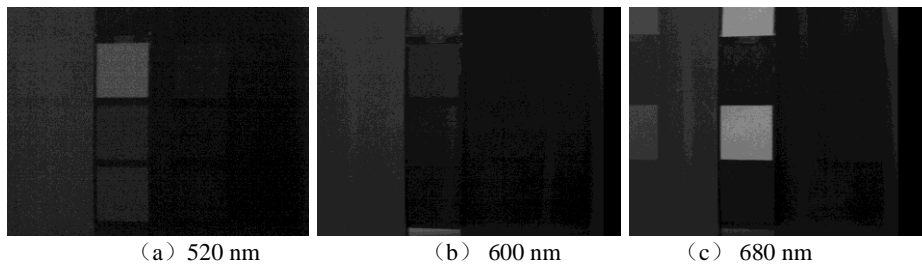

（a）520 nm　　　　　　　（b）600 nm　　　　　　　（c）680 nm

Figure 8. Single spectrum image

Red, green and blue spectrum segments (520nm, 600nm and 680nm) were selected for color synthesis, and the target color image was obtained, with the left and right images as the target original color images, and the spectral color synthesis effect was highly consistent with the target original color.
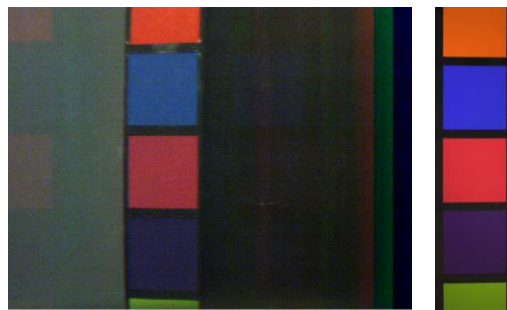

Figure 9. Color composite image of the target (left) (right: original color image of the target)

5. CONCLUSIONS

Aiming at the application of optical remote sensing in the complex environment of low illumination at night, the computational imaging technology of hyperspectral remote sensing under the condition of extremely low illumination and extremely low signal-to-noise ratio is studied. Using the compressed sensing technology of the main mirror of the thin film diffraction grating, the large aperture optical acquisition is carried out, and the large dynamic range perception is obtained by Geiger mode imaging. The system realizes target recognition under the condition of extremely low image signal-to-noise ratio through the design of the new system's main mirror flattening and the aliasing compression and decoupling of spatial information and spectral information. The technology has completed space-based scheme design and desktop principle verification tests, and the spectral resolution reaches 5nm.

REFERENCES

[1] HU Xiaohua, LIU Songtao, Pan Shaoxing, et al. Development of spaceborne detection instruments and their data citation [J]. Chinese Optics,8 (3) : 350-359 (2015).

[2] Lee T E，Miller S D，Turk F J，et al.The NPOESS VIIRS Day/Night Visible Sensor[J]Bulletin of the American Meteorological Society, 87(2):191-199 (2006).

[3] Li X, Li D. Can night-time light images play a role in evaluating the Syrian Crisis[J]. International Journal of Remote Sensing,35(18):6648-6661(2014).

[4] "DMSP-Block 5D-3 Spacecraft Overview" Mar.30,URL: http://spaceflightnow.com/2016/03/30/air-force-ends-effort-to-recover-dmsp-weather-satellitte (2016).

[5] "Visible Infrared Imaging Radiometer Suite(VIIRS)"URL:https://jointmission.gsfc.nasa.gov/viirs.html

[6] Elvidge C D,Zhizhin M,Baugh K,et al. Automatic Boat Identification System for ViIRS Low Light Imaging Data[J].Remote Sensing,7(3):3020-3036 (2015).

[7] Elvidge C D, Cinzano P, Pettit D R,et al. The Nightsat mission concept[J]. International Journal of Remote Sensing,28(12):2645-2670 (2007).

[8] Zhang Q, Seto K C. Can night-time light light data identify typologies of urbanization a global assessment of successes and failures[J] Remote Sensing, 5(7):3476-3494 (2013).

Passive real-time hyperspectral imager based on Fabry-Pérot interferometer filter technology

Alexander Kokka[a], Christer Holmlund[a], Leevi Salonen[a], Olli Ihalainen[a], Heikki Astola[a], Matti Mõttus[a], Rami Mannila[a], Timo Ihatsu[a], and Ingmar Stuns[a]

[a]VTT Technical Research Centre of Finland Ltd, Tietotie 3, 02150 Espoo, Finland

ABSTRACT

This paper presents the system design of a real-time hyperspectral imager based on tunable Fabry-Pérot interferometer (FPI) filter technology. This passive hyperspectral instrument is able to capture spectral data at a rate corresponding to video-like image feed. The instrument is designed to be suitable for handheld operation as well as for missions carried out using uncrewed aerial vehicles.

The frame rate of individual spectral channels of an FPI-based camera, and subsequently the acquisition speed of hyperspectral data, depends on the actuation speed of the FPI filter, exposure time of the sensor, data transfer rate, and all delays between the consecutive operations. In order to minimize the delays when switching between the spectral channels, the large FPI of this instrument is enclosed in a low-pressure housing to reduce air resistance, which would otherwise slow down the mechanical movement of the filter. As various applications require different sets of wavelengths and a variable number of spectral channels to be recorded, the imager enables selecting the desired wavelengths programmatically from within the complete spectral range of the instrument.

FPI-based hyperspectral cameras produce a full two-dimensional image for each spectral channel. The spatial information contained in the images may be used to compensate for any desired or undesired movement of the imager. The spatial information available for individual channels can also be used for data analysis, and it enables employing conventional machine vision algorithms for example to detect and track the objects of interest.

Keywords: hyperspectral, camera, real-time, Fabry-Pérot interferometer, FPI

1. INTRODUCTION

During the past decades, hyperspectral imaging has been increasingly adapted to new fields of applications, as well as more widely employed in its traditional fields.[1–4] Advances in hyperspectral imaging technologies have enabled compact, more affordable, and high-performing cameras with a wide variety of wavelength range options. While the traditional data acquisition techniques are suitable for a multitude of scenarios, some use-cases benefit, or even require, the ability to rapidly record and present hyperspectral data.[5–7]

Hyperspectral technologies can be coarsely divided into three categories based on the data acquisition method employed by the camera. These categories include spatially scanning point or line cameras, spectrally scanning frame cameras, and various snapshot cameras, which record a set of spectral channels, called hyperspectral cubes, with a single exposure of the sensor. Snapshot cameras can rely on a microlens-array optics or on a fixed-filter mosaic deposited on the image sensor.[8,9] In these types of solutions, the pixel resolution of the image sensor is traded off for the spectral resolution.

Spatial scanners record spectral data for one point or one line of points at once, but require movement of either the object or the imager to produce a hyperspectral image with spatial information. For instance, a stationary line scanner can be used to image objects on a moving conveyor belt,[5] or the camera can be installed on an aircraft flying in a linear motion to record the spectral data from the area underneath it.[10] In contrast to this type of spatially scanning imagers, spectral scanners record a full 2D image frame for each spectral channel, but require repeated exposure of the sensor at each wavelenth, which contributes to the acquisition time of the hyperspectral data cube.[11]

Corresponding author: Alexander Kokka
Email: alexander.kokka@vtt.fi

In this study, a passive real-time hyperspectral imager based on Fabry-Pérot interferometer (FPI) filter technology[12] is presented. For this type of spectrally scanning hyperspectral camera, the framerate of spectral channels, and subsequently the acquisition speed of hyperspectral data cubes, is determined by the actuation speed of the tunable FPI filter, exposure time of each channel, data transfer rate, and all delays between the consecutive operations. By minimizing the time consumed by each phase, the instrument can be made to produce a video-like feed of spectral data.

2. MATERIALS AND METHODS

2.1 Tunable Fabry-Pérot interferometer filter

The hyperspectral camera developed in this study is based on a tunable FPI filter. FPI filters consist of two parallel highly reflective mirrors separated by a gap. The constructive interference of the incoming electromagnetic radiation depends on the distance travelled by the radiation between the mirrors. For incoming signal perpendicular to the mirrors, the central wavelength λ of the transmitted band depends on the gap size according to the equation

$$\lambda = \frac{2d}{n}, \tag{1}$$

where d is the distance between the mirrors and n is a positive integer, $n = 1, 2, 3, \ldots$, denoting the order of interference. The spectral bandwidth of the transmitted signal depends on the mirror materials, thickness, and geometry. Any deviation from the parallel geometry of the mirrors increases the bandwidth of the transmitted signal. Figure 1 shows the main components of the tunable piezo-actuated FPI filter.
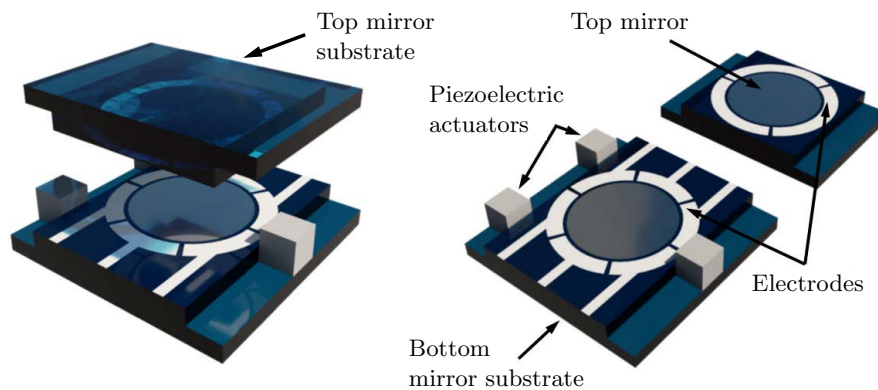

Figure 1. Components and structure of the tunable Fabry-Pérot interferometer filter.

Three piezoelectric actuators are used to set the mirror gap to correspond with the desired wavelength and to keep the mirrors parallel to each other. The tilt and the distance between the mirrors are actively controlled by measuring the capacitance of the electrodes and adjusting the voltage applied to the piezoelectric actuators in a continuous feedback loop.

Using this type of tunable spectral filter, the set of gaps, and thus the wavelength channels, can be set programmatically from within the total wavelength range of the device. This enables employing various application-specific wavelength sets when capturing spectral data. That can be used to reduce the amount of redundant data recorded by the camera, and to increase the data cube acquisition rate of the instrument.

The delay between setting a new wavelength setpoint and settling of the mirrors at the new gap depends on the size of the FPI, the difference Δd between the two gaps, and the direction of the mirror movement. When operating at the atmospheric pressure, the air resists the movement of the mirrors. The impact of air pressure can be particularly significant when decreasing the gap. Decreasing the air pressure around the FPI reduces the damping effect produced by air molecules, and increases the speed at which the filter can be actuated.

2.2 Optics and Mechanics

By installing a tunable FPI filter on the optical path of a camera with a broadband image sensor, the camera is enabled to record spectrally resolved data. Typically, the FPI is installed in front of the focusing optics of the camera. To increase the signal strength, a set of light gathering lenses can be used in front of the FPI. As the constructive interference, and thus the transmitted central wavelength, depends on the size of the gap and on the angle at which the radiation enters the FPI, a set of collimating optics may be used to ensure that the light coming from one point in the scene transmits through the FPI at the same angle throughout the entire area of the mirrors. Moreover, any angular deviation from the perpendicular geometry results in a shift of the transmitted wavelengths. This phenomenon is often referred to as the spectral smile,[13] and can be mitigated using front optics.

In the case of a narrow field-of-view instrument focused to infinity, it is possible to employ an FPI filter directly in front of the camera optics without introducing significant spectral aberrations to the captured data. This enables the use of a commercial off-the-shelf camera objective together with an FPI filter to build a hyperspectral imager, without using any custom optics. In order to prevent out-of-band transmission from outside the active wavelength range of the FPI, or from unwanted orders of interference, n in equation (1), long-pass and short-pass filters are typically installed on the optical path of the instrument.

In addition to the commercial optics, this type of FPI-based hyperspectral camera design allows using an industrial camera module as the detector. With this approach, the image sensor, peripheral electronics, and I/O and data interfaces can be deployed as one integrated package without the need for any modifications. The FPI controller can be programmed to be used together with the I/O lines of the camera module to trigger the data acquisition, and to synchronize the FPI with the sensor exposure. In this type of instrument, the maximum frame rate of the camera module sets the absolute limit for the acquisition speed of hyperspectral data.

3. REAL-TIME HYPERSPECTRAL CAMERA

3.1 Hardware and specifications

Figure 2 shows an overview of the real-time hyperspectral imager developed in this study. The hyperspectral data acquisition rate has been maximized by minimizing the required exposure time for each spectral channel and the delays between the consecutive channels. By using the application-specific subsets of the spectral channels, out of the total wavelength range of the camera, the imager is able to record spectral data at video-like rates. The instrument is designed to be suitable for handheld usage and airborne measurements carried out using uncrewed aerial vehicles.
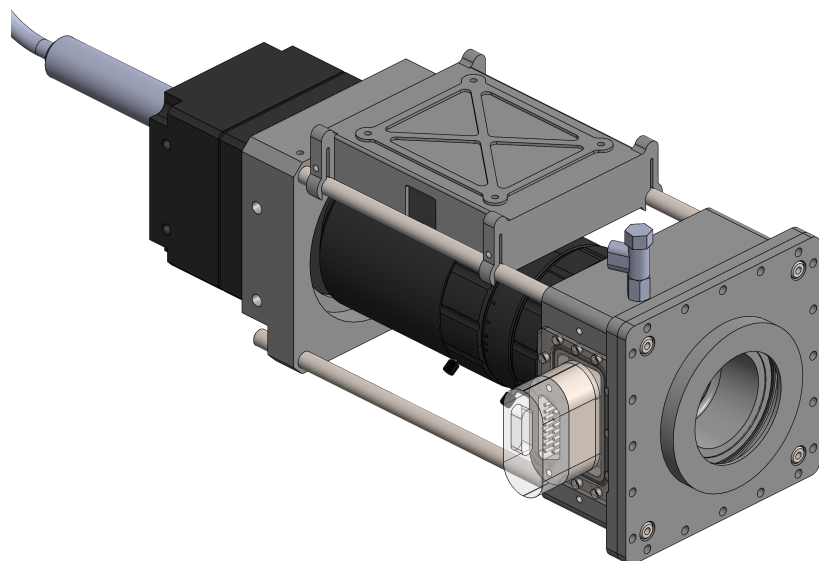

Figure 2. An overview of the real-time hyperspectral imager.

An industrial camera module, with a cooled indium gallium arsenide (InGaAs) sensor, and a commercial 50 mm fixed focal length objective are used as the basis for the camera. The camera module and the objective are both equipped with the standard C-mount interface. The camera module relies on GigE Vision interface for the device control and data transfer. The key technical specifications of the instrument are presented in Table 1.

Table 1. Main hardware specifications of the real-time hyperspectral imager

Parameter	Value
Pixel resolution	1280×1024
Pixel size	5 μm
Field of view	$7.4° \times 5.9°$
Wavelength range	1100–1450 nm
Spectral bandwidth	20–25 nm
FPI diameter	23 mm
Focal length	50 mm
f/N	2.8
Mass	2.1 kg

A large FPI with a pair of 23 mm dielectric mirrors is used as the spectral filter. This enables large-aperture optics improving the light-gathering ability of the optical system, which consequently reduces the exposure time required by each frame. Increasing the size of the FPI mirrors increases the impact of the air pressure when the gap between the mirrors is changed. The damping effect created by air molecules can be reduced by decreasing the air pressure around the FPI. For that, the FPI is enclosed in an airtight housing, which is installed in front of the camera objective using an assembly consisting of four steel rods. Figure 3 shows a cross section of the FPI housing. The case has an air valve, area-limiting apertures for the FPI, and two vacuum-compatible D-subminiature connectors for the electrical signals and power. Band-limiting filters, one 1100 nm long-pass filter and one 1450 nm short-pass filter, act as the windows for the housing. A pressure sensor is integrated inside the case, and all the mechanical interfaces of the assembly are sealed using nitrile rubber (NBR) O-rings.
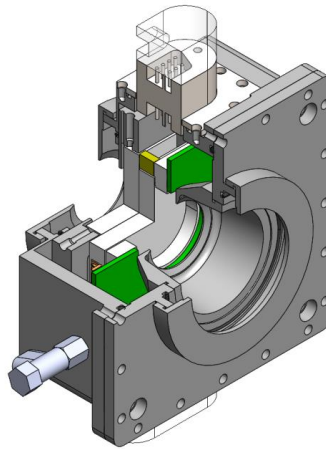

Figure 3. The low-pressure FPI housing rotated 90° about the optical axis, and with a section cut out.

As the instrument is designed for long-range remote sensing, the infinity focus and the narrow $7.4° \times 5.9°$ field-of-view of the optics allows installing the FPI directly in front of the camera objective without introducing significant spectral aberrations. Combined with the field-of-view of the instrument, the 5 μm pixel size of the image sensor would produce the ground sampling distance of approximately 10 cm from 1 km altitude.

3.2 Data acquisition scheme

The enclosure for the FPI control electronics is on the top of the instrument in Figure 2. The cables of the FPI controller are not displayed in the figure. The electronics monitor the capacitance of the electrodes on the mirror substrates, and apply the control voltage to the piezoelectric actuators to tune the FPI. The FPI controller also triggers the exposure of the image sensor once the mirrors have settled at the desired gap, and autonomously sets the new setpoint for the FPI once the signal from the camera module indicates that the exposure is ready.

Hyperspectral cameras based on the tunable FPI filter technology allow setting the desired wavelength channel using software. For instance, this enables first measuring a target using the full wavelength range of the camera, and then determining a subset of the most significant wavelengths for a given application. This subset of wavelengths can then be used to carry out the real-time analysis of the new samples.

In real-time data acquisition mode, the subset of setpoints, corresponding to the wavelengths of interest, is stored in the flash memory of the FPI controller. Once triggered, the controller autonomously goes through this list of setpoints. At each new setpoint, the controller sets the FPI actuation indicator output and starts to monitor the FPI gap by comparing the measured capacitance with that of the new setpoint. Once the difference between the two is within a given threshold for the set number of consecutive samples, the FPI controller resets the actuation indicator output. The falling edge of the indicator line triggers the camera module to expose one frame. While the sensor exposure is active, the output line of the camera module is set high, and once the exposure is ready, the falling edge of the line triggers the FPI to move to the next setpoint. This is continued until all the channels of the subset are captured. Figure 4 visualizes the acquisition scheme for the data cubes with seven evenly spaced FPI-mirror gaps. The figure shows the control voltage of one piezoelectric actuator, the FPI actuation indicator output, and the sensor exposure indicator output measured for the FPI in the low-pressure conditions of less than 1 mbar. The exposure time was set to 5 ms for every channel, which is sufficient in full sunlight conditions. The acquisition rate in this configuration is 9.4 data cubes per second.
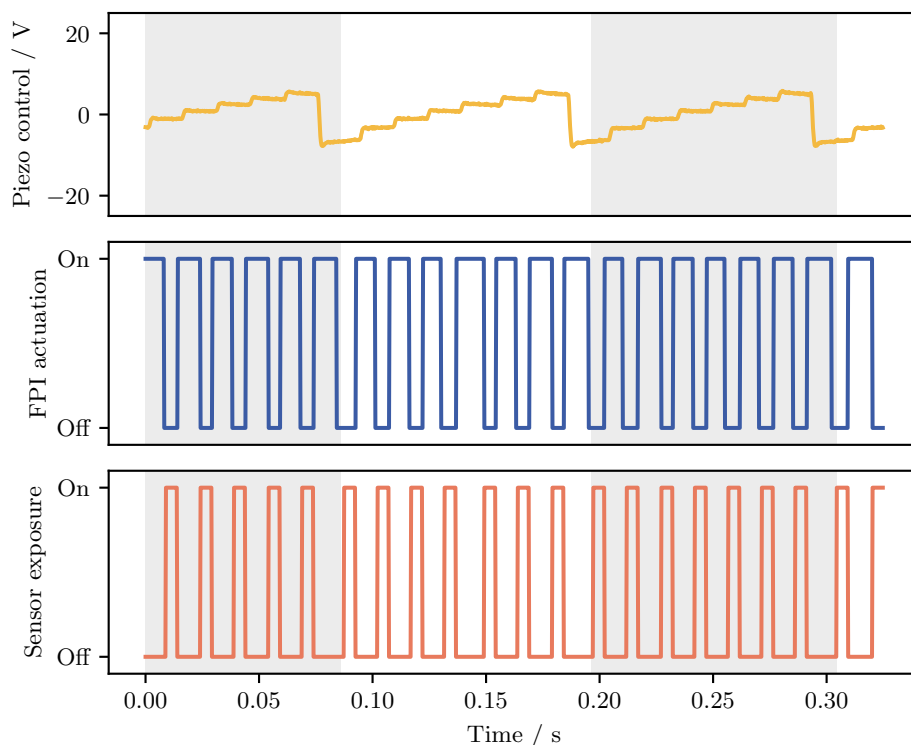

Figure 4. Data acquisition scheme of the real-time hyperspectral imager with the FPI in low air pressure. Background colors indicate individual data cubes.

The captured frames for each spectral channel are stored in the pre-allocated memory buffer of the camera module, and are transferred to the computer as soon as each channel is ready. In the case of a moving camera, the temporal offset between the channels can cause spatial distortion between the channels in form of translation and rotation. The offset needs to be corrected before any further analysis. Transferring the data for each channel as soon as it is available allows starting the image registration process for the first channels while rest of the data cube is being recorded. Once the complete data cube has been transferred, it is ready to be processed and analyzed while the next cube is being captured.

The number of required spectral channels per data cube depends on the application. For instance, calculating most spectral indices typically requires no more than two suitable spectral bands. In the case of the FPI-based hyperspectral camera, this means that only two spectral channels, out of the whole wavelength range of the instrument, would need to be recorded for each data cube, leading to the acquisition rates of 20–30 cubes per second. The exact rate depends on the exposure time and the gap size difference between the setpoints.

3.3 Impact of air pressure

Figure 5 shows the same data acquisition scheme as in Figure 4, but this time measured in atmospheric pressure. In addition to the increased time spent on tuning the FPI, the control voltage of the piezoelectric actuators increases substantially compared to the low-pressure conditions. The increased control voltage indicates the increased load on the piezoelectric actuators when setting the gap associated with each setpoint. The data acquisition rate in this configuration is 8.6 data cubes per second.
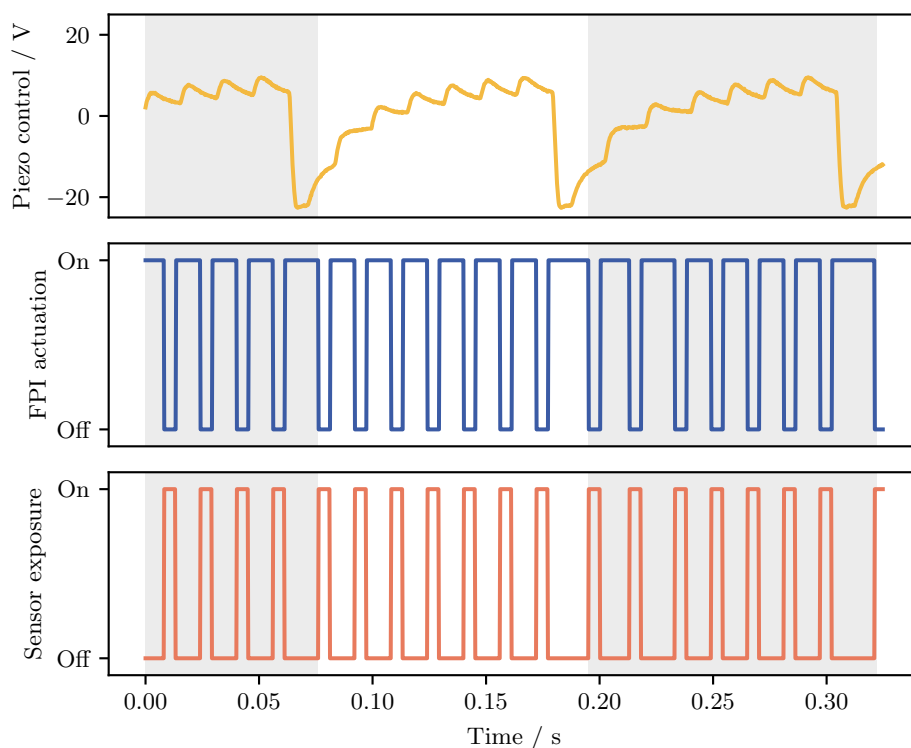

Figure 5. Data acquisition scheme of the real-time hyperspectral imager with the FPI in atmospheric pressure. Background colors indicate individual data cubes.

In this configuration of the exposure time and the number of channels per cube, the acquisition rate is approximately 10% faster in the low-pressure conditions compared to the results in atmospheric pressure. In addition to the slower FPI actuation speed in atmospheric pressure, it can be seen from the control voltage plot of Figure 5 that the sensor starts to expose the first frame of each data cube while the control voltage is still

stabilizing. This is likely due to the edges of the mirror substrates getting pulled in before the center region, as the air inside the gap is being gradually displaced, which temporarily deforms the mirrors. As the electrodes reside at the edges of the mirror substrates, the measured capacitance reaches that of the target setpoint, and the frame is triggered prematurely, leading to errors in spectral data.

The relative increase of capturing rate in low-pressure conditions, when compared to the atmospheric pressure, depends on the number of spectral channels in the data cubes, the employed exposure time, and the gap difference between the consecutive setpoints. In the case of a large number of channels with long exposure times and small gap steps, the impact of the FPI actuation time on the data acquisition rate is outweighed by these factors.

4. CONCLUSION

In this study, a real-time hyperspectral imager was developed. The instrument is based on the tunable Fabry-Pérot interferometer filter technology. In addition to capturing continuous hyperspectral data, the instrument is able to record user-set spectral channel subsets, producing video-like feed of spectral data. The wavelength subsets can be set programmatically from within the wavelength range of the instrument. By employing the tunable FPI filter, this type of spectral scanner can be built using commercial off-the-shelf imaging optics and an industrial camera module.

The data acquisition rate of an FPI-based hyperspectral camera depends on the actuation speed of the FPI filter, exposure time of the sensor, data transfer rate, and the delays between the operations. To reduce the exposure time required by individual frames, an FPI filter with large 23-mm mirrors is employed in the instrument. Tuning the spectral transmission band of the FPI filter requires changing the distance between the mirrors. When adjusting this gap, the air pressure resists the movement of the mirrors, which slows down the data acquisition rate of the imager. The damping effect created by air molecules can be reduced by decreasing the air pressure in which the filter is operated. For that purpose, a low-pressure enclosure was designed for the FPI of the instrument.

To reduce the delays associated with tuning the FPI filter, the control electronics and firmware were developed to incorporate image acquisition control. In this approach, the electronics of the FPI actively measure the deviation of the gap from the target setpoint, and trigger the exposure of the image sensor as soon as the FPI has settled on the desired wavelength. The camera module, in turn, is programmed to signal the FPI electronics to switch the setpoint of the FPI to the next wavelength, once the frame exposure is ready.

REFERENCES

[1] Dale, L. M., Thewis, A., Boudry, C., Rotar, I., Dardenne, P., Baeten, V., and Pierna, J. A. F., "Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: A review," *Applied Spectroscopy Reviews* **48**(2), 142–159 (2013).

[2] Lu, G. and Fei, B., "Medical hyperspectral imaging: a review," *Journal of biomedical optics* **19**(1), 010901–010901 (2014).

[3] Khan, M. J., Khan, H. S., Yousaf, A., Khurshid, K., and Abbas, A., "Modern trends in hyperspectral image analysis: A review," *Ieee Access* **6**, 14118–14129 (2018).

[4] Lu, B., Dao, P. D., Liu, J., He, Y., and Shang, J., "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sensing* **12**(16), 2659 (2020).

[5] Ni, C., Li, Z., Zhang, X., Sun, X., Huang, Y., Zhao, L., Zhu, T., and Wang, D., "Online sorting of the film on cotton based on deep learning and hyperspectral imaging," *Ieee Access* **8**, 93028–93038 (2020).

[6] Picón, A., Ghita, O., Bereciartua, A., Echazarra, J., Whelan, P. F., and Iriondo, P. M., "Real-time hyperspectral processing for automatic nonferrous material sorting," *Journal of Electronic Imaging* **21**(1), 013018–013018 (2012).

[7] Khan, A., Munir, M., Yu, W., and Young, B., "A review towards hyperspectral imaging for real-time quality control of food products with an illustrative case study of milk powder production," *Food and Bioprocess Technology* **13**, 739–752 (2020).

[8] Yu, C., Yang, J., Song, N., Sun, C., Wang, M., and Feng, S., "Microlens array snapshot hyperspectral microscopy system for the biomedical domain," *Applied Optics* **60**(7), 1896–1902 (2021).

[9] Hahn, R., Hämmerling, F.-E., Haist, T., Fleischle, D., Schwanke, O., Hauler, O., Rebner, K., Brecht, M., and Osten, W., "Detailed characterization of a mosaic based hyperspectral snapshot imager," *Optical Engineering* **59**(12), 125102–125102 (2020).

[10] Jurado, J. M., Pádua, L., Hruška, J., Feito, F. R., and Sousa, J. J., "An efficient method for generating UAV-based hyperspectral mosaics using push-broom sensors," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **14**, 6515–6531 (2021).

[11] Saari, H., Aallos, V.-V., Holmlund, C., Mäkynen, J., Delauré, B., Nackaerts, K., and Michiels, B., "Novel hyperspectral imager for lightweight UAVs," in [*Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications VII*], **7668**, 44–52, SPIE (2010).

[12] Antila, J., Miranto, A., Mäkynen, J., Laamanen, M., Rissanen, A., Blomberg, M., Saari, H., and Malinen, J., "MEMS and piezo actuator-based fabry-perot interferometer technologies and applications at VTT," in [*Next-Generation Spectroscopic Technologies III*], **7680**, 192–203, SPIE (2010).

[13] Yokoya, N., Miyamura, N., and Iwasaki, A., "Preprocessing of hyperspectral imagery with consideration of smile and keystone properties," in [*Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques, and Applications III*], **7857**, 73–81, SPIE (2010).

Spectropolarimetry for space object identification

Ralph Snel[a], Bogdan Vasilescu[b], Eugenio Di Iorio[a], Pierre Piron[b],
Jérôme Loicq[b], Ivan Ferrario[a], Fabrizio Silvestri[a]

[a]TNO Netherlands Organisation for Applied Scientific Research, Stieltjesweg 1, 2628 CK Delft, the Netherlands; [b]Delft University of Technology, Kluyverweg 1, 2629 HS Delft, the Netherlands

ABSTRACT

A novel method for space object identification is proposed, based on full Stokes spectropolarimetry in the visible and near-infrared wavelength range. Space objects that have been previously detected and are illuminated by the sun can be observed with a telescope to simultaneously obtain intensity, spectra, and polarimetry, and compose light curves of these parameters as function of time. The intention is to thus assign a unique identification, or at least a classification to these objects.

Single, double, and multiple reflections of sunlight off the space object (natural or artificial objects, including debris) will introduce spectrally dependent polarisation into the scattered light, the spectral signature of which is affected by the complex refractive index of the scattering materials and the geometry. The simultaneous measurement of the full Stokes vector allows separation of the light source unpolarised spectral signatures on the one hand from the polarisation spectral features on the other hand.

To illustrate the concept, we have performed a number of simulations for double scattering off a small selection of materials, for a large range of scattering geometries. Examples of individual scattering geometries and statistical summaries of all geometries are presented.

A demonstrator spectropolarimeter is being built, we present an overview of the design and the high level planning, as well as some predicted performance parameters.

Keywords: Spectropolarimetry, polarimetry, space object identification, UV-vis-NIR, scattered sunlight

1. INTRODUCTION

With the ever-increasing number of artificial satellites in orbit around the Earth, space objects may need some kind of identification more than just their orbital elements [1][2]. Spectroscopy [3] and polarimetry [4] can be used to increase the knowledge about previously detected space objects. We propose to combine both approaches and use full Stokes vector spectropolarimetry for space object identification.

This paper present the rationale behind this choice (Section 2), the polarisation modulator chosen (Section 3), the design of a breadboard spectropolarimeter to demonstrate the principle and learn about spectropolarimetric behaviour of materials and satellites (Section 4), the design of a test bench to characterise the polarisation behaviour of the components in the breadboard (Section 5), the plans for data acquisition and analysis (Section 6), and concludes with a summary and discussion (Section 7).

2. SPECTROPOLARIMETRY FOR SPACE OBJECT SIGNATURE DETECTION

Starting with the spectral complex index of refraction of dielectric media and using the Fresnel equations and a scattering or reflecting geometry, the full Stokes vector of the reflected light can be calculated for any wavelength for which the index of refraction is specified, assuming single reflection off the surface, thus excluding volume scattering effects. Also, multiple reflections can be calculated this way. Figure 2-1 shows the result of the calculation of the spectrally resolved normalised Stokes vector for double reflection off two materials, aluminium and silicon, under a scattering geometry of 50 degrees angle of incidence on both the first and second surface, with a 45-degree rotation of the coordinate frame to produce out-of-plane scattering and promote the generation of circular polarisation. Unpolarised illumination was used in this calculation. Note how the Q/I curves overlap or nearly overlap for linear polarisation (Q/I and U/I) and how the differences become more distinct for the circular polarisation component (V/I).
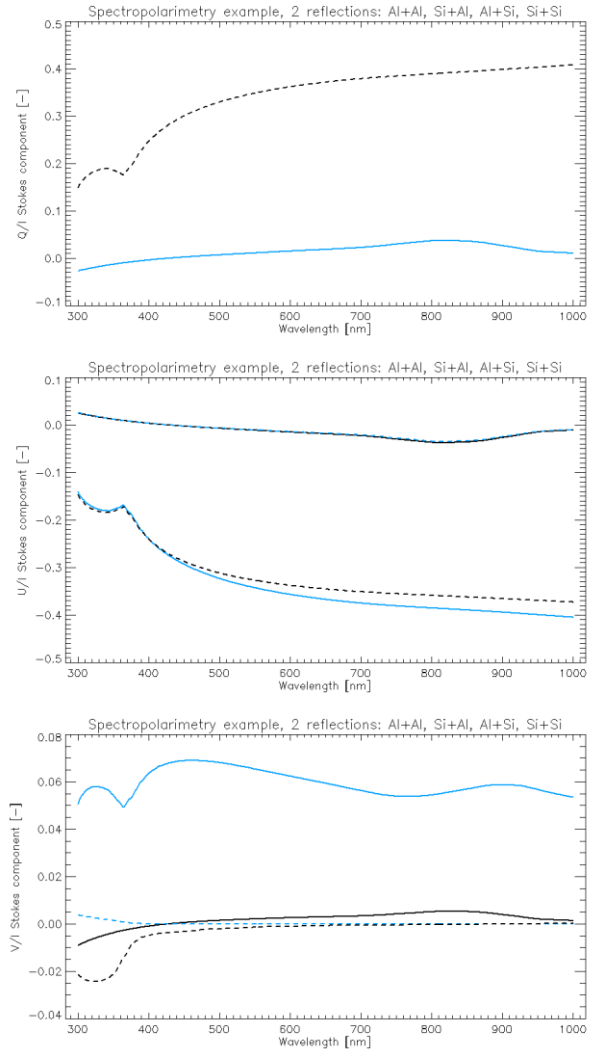
Figure 2-1. Wavelength dependent Stokes vector parameters Q/I (top), U/I (middle), and V/I (bottom) for the geometry and materials described in the text. The full black curve corresponds to two reflections off aluminium, the full blue curve for reflection off aluminium and silicon in that order, the dashed blue curve for silicon and aluminium, and the dashed black curve for two reflections off silicon.

This simple simulation shows that the wavelength dependence of the complex indices of refraction, as well as the order of reflection of the two surfaces, can affect the spectrally resolved Stokes parameters, in particular the circular polarisation component.

In more realistic cases, reflection of sunlight off a satellite will be composed of a mix of single- and multiple scattering, weighed with the relative scattering area and reflectivity for the various light paths from the sun to the observer. Figure 2-2 shows an example of possible light paths in a rendered image of a satellite, note the reflections of the gold-coloured multi-layer insulation (MLI) in the dark solar panels, and the reflections of the solar panels in the MLI. As the illumination and viewing geometry changes over the orbit of a satellite, flares of single- or multiple reflection may arise, carrying potentially distinctive information of the surface materials and configuration of the satellite.

Figure 2-2. ESA Sentinel 5 Precursor satellite, rendered image [5].

The normalised Stokes vector components Q/I, U/I, and V/I can also be expressed in Poincaré sphere angles and degree of polarisation. This allows to potentially highlight differences between different space objects and orientations which may be less prominent in the spectrally resolved normalised Stokes vector.
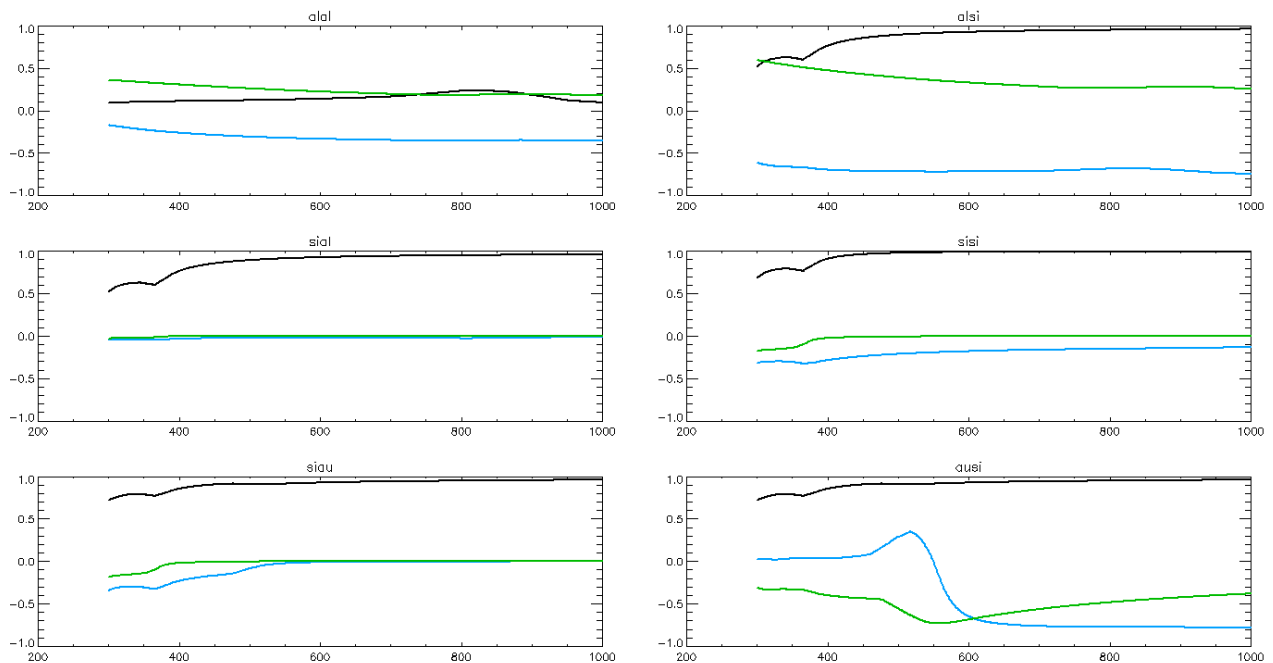

Figure 2-3. RMS variation of the Poincaré sphere angles in radians (blue and green curves) and the degree of polarisation (black curves) as function of wavelength in nm for 6 different material combinations: double scattering off aluminium and aluminium (alal), aluminium and silicon (alsi), silicon and aluminium (sial), silicon and silicon (sisi), silicon and gold (siau), and gold and silicon (ausi). Note the distinctly different spectral behaviour between the different material combinations but also order of scattering for the blue and green curves, representing the Poincaré sphere angles. There is much less distinction visible in the degree of polarisation (black curves)

Figure 2-1 is calculated for a scattering geometry that is beneficial for circular polarisation, in order to get an idea of the "typical" variation to expect in the Poincaré sphere angles assessed over all possible illumination geometries, the Stokes

vectors and Poincaré sphere angles were calculated for multiple material combinations and scattering geometries, which were subsequently used to determine the root mean square variation (RMS) over all geometries. Figure 2-3 shows the results of this calculation. It is clear from these plots that the RMS variation as function of wavelength of in particular the Poincaré sphere angles (green and blue curves) is very scattering material specific, allowing to distinguish between different materials just on the basis of spectrally resolved RMS variation of its polarisation features. This also implies that space objects composed of different materials or mixtures of materials may be distinguished.

Note that for a single observation point in the spectropolarimetric light curve of a space object the complete spectrally resolved and polarisation resolved information is available, and that as the space object progresses through its orbit, the scattering geometry and thus the spectropolarimetric signature changes. Figure 2-3 only gives a statistical summary of all possible scattering geometries, in reality a space object will pass through only a limited range of geometries and its individual spectropolarimetric light curve may be used for identification purposes of the object.

3. POLARISATION MODULATOR

The main part of the instrument is the modulator. This is composed of three prisms in Magnesium Fluoride, optically glued together (see Figure 3-1)[6][7][8]. In addition, the fast axis is differently oriented in each prism. Therefore, as can be observed in Figure 3-1, the prism (1), of apex angle ξ, has the fast axis oriented along the x-axis. The middle component of the modulator, (2), has the fast axis oriented along the z-axis. In the third wedge of apex angle ψ, the fast axis makes an angle of 45° with the x-axis, in the (xy)-plane.

The prismatic shape of the elements (1) and (3) ensures a continuous variation of the phase difference between the orthogonal components of light in the y direction.

Therefore, given the geometry and orientation of the fast axis, the first prism can convert any incoming state of polarisation, except the Q state, into a continuous variation of polarisation in the vertical direction.

The second prism, (2), plays only the role to keep everything together, into a single part. It helps also mitigate the deviation of rays, as it is built from the same material as the other two.

The last prism, (3), is complementary to the first one: it serves to modulate the Q state, left unaffected previously.

Using this geometry, the polarised light arriving collimated on the left side of the modulator exit the prisms exhibiting a continuous variation of the polarisation in the vertical direction. Therefore, by placing an analyser after this modulator, this variation can be converted into a variation in intensity, that can be measured with a detector, along the vertical direction.
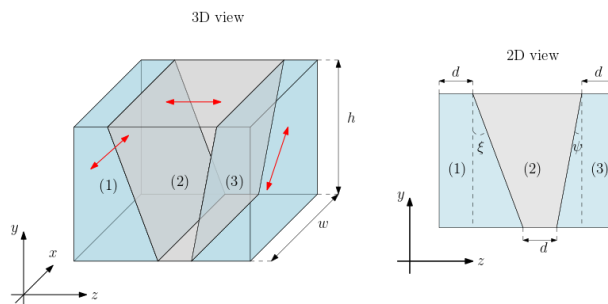

Figure 3-1. 3D and 2D view of the modulator. Each prism of the modulator has a particular orientation of the fast axis: along the x-axis in prism (1), along z in prism (2), and at 45° with respect to x-axis in the prism (3). The height of the modulator, h, must cover at least the largest period of the signal described by the Eq.(1). Because there is no modulation of the signal in the x direction, the width can be chosen according to the constraints of the optical system. Concerning the length, it is dictated by the height h, the angles ξ and ψ, and by the additional thickness, d.

To avoid the split of the ordinary and extraordinary rays and to decrease the frequencies of the modulation, the angles ξ and ψ must be kept as small as possible. Also, to aid manufacturing and handling of the optical component, an additional thickness, d, of constant value can be added to each prism.

Based on this architecture, the intensity of light detected after the passage through the modulator and the analyser follows the relation

$$I(\theta, y, \lambda) = \frac{1}{2}[S_1 + S_2 \cdot m(\theta, y, \lambda) + S_3 \cdot n(\theta, y, \lambda) + S_4 \cdot p(\theta, y, \lambda)] \tag{1}$$

where θ is the angle of the analyser with respect to the x-axis, y is the position in the vertical direction, λ is the wavelength, and $\vec{S} = [S_1, S_2, S_3, S_4]^T$ is the Stokes vector of the incoming light. The functions m, n, p, computed with the help of the Mueller calculus, are given by:

$$m(\theta, y, \lambda) = cos(2\theta)cos(\Delta\phi_3)$$
$$n(\theta, y, \lambda) = sin(2\theta)cos(\Delta\phi_1) + cos(2\theta)sin(\Delta\phi_1)sin(\Delta\phi_3) \tag{2}$$
$$p(\theta, y, \lambda) = sin(2\theta)sin(\Delta\phi_1) - cos(2\theta)cos(\Delta\phi_1)sin(\Delta\phi_3)$$

where $\Delta\phi_1$ and $\Delta\phi_3$ are the phase differences induced by the prisms (1) and (3):

$$\begin{cases} \Delta\phi_1 = \dfrac{2\pi}{\lambda}\Delta n(\lambda)[d + (h - y)tan(\xi)] \\ \Delta\phi_3 = \dfrac{2\pi}{\lambda}\Delta n(\lambda)[d + (h - y)tan(\psi)] \end{cases} \tag{3}$$

where $\Delta n(\lambda) = |n_o(\lambda) - n_e(\lambda)|$ is the absolute value of the difference between the ordinary and the extraordinary indices of refraction, also called the birefringence of the medium, while h is the height of the modulator.

From the Eq. (1) we infer that the intensity detected by the pixels in the y direction, when the wavelength is known and the orientation of the analyser θ well established, can be expressed by the following system:

$$\begin{pmatrix} I(y_1) \\ I(y_2) \\ \vdots \\ I(y_N) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & m(y_1) & n(y_1) & p(y_1) \\ 1 & m(y_2) & n(y_2) & p(y_2) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & m(y_N) & n(y_N) & p(y_N) \end{pmatrix}\begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \end{pmatrix} \tag{4}$$

Solving this system of equations for S_1, S_2, S_3, S_4 gives us access to the incoming state of polarisation.

The spectral limitation of this modulator is dictated by the transmission band and the birefringence of the material. With the help of Magnesium Fluoride, we can access the entire spectrum between 0.12 and 7 μm [6][7].
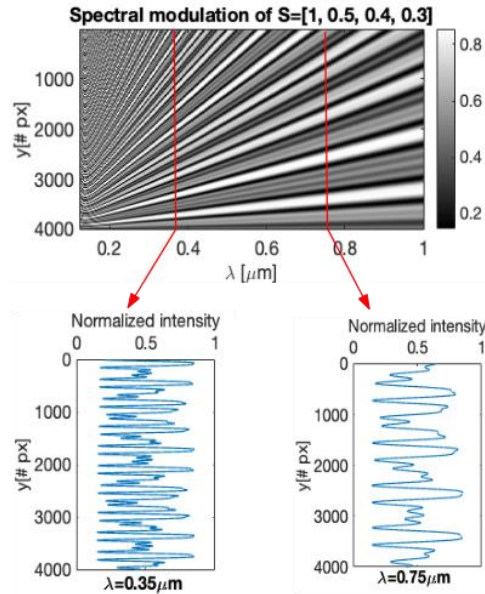
Figure 3-2. Spectral modulation of an incoming state of polarisation S=[1,0.5,0.4,0.3] obtained on a column of 4000 pixels, with a size of 5μm. In this simulation we supposed that the height of modulator is h=2cm, the apex angles are ξ=2.6°,ψ=1.8°, and the additional thickness is d=1mm. The orientation of the analyser was θ=72°.

Given the possibility to work at different wavelengths and the fact that the modulator is redundant in the x direction, then the x-axis can be used to spectrally disperse the light on the detector. Therefore, on the detector plane, we can have access simultaneously to the polarimetric information in the vertical direction, and to the spectral information in the horizontal.

For the manufacturing of the modulator, we have used the apex angles $\xi = 2.6°$ and $\psi = 1.8°$. Together with an orientation of the analyser at $\theta = 72°$, these values can ensure a hight efficiency of the modulation scheme of the polarimeter [9] and an almost equal precision on the measurement of the Stokes parameters. The additional thickness, d, is 1mm, ensuring easy and safe handling of the modulator.

4. BREADBOARD DESIGN

The breadboard optical design starts with the interface to the telescope focal plane. Input is considered to be a point source, limited in angular extent by the quality of the telescope and the seeing conditions of the atmosphere. An appropriately sized pinhole is placed around this point source. From this, a collimated beam is produced, followed by the polarisation modulator and analyser, and subsequently an equilateral prism to create the desired dispersion in the plane perpendicular to the polarisation modulation. A cylindrical lens images the pinhole in the spectral dimension onto the detector, while maintaining the polarisation modulation in the pupil plane of the instrument plus telescope. Figure 4-1 (left) shows an impression of the optical design, without the telescope which would be placed on the right.
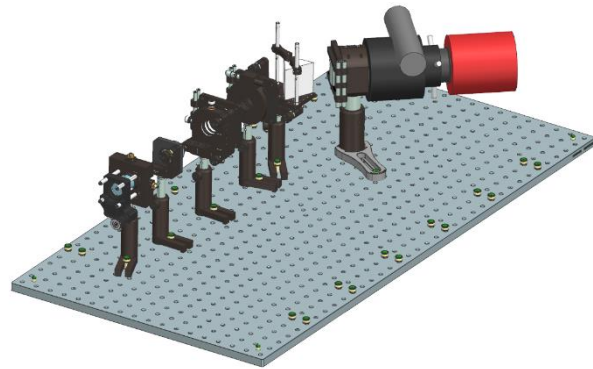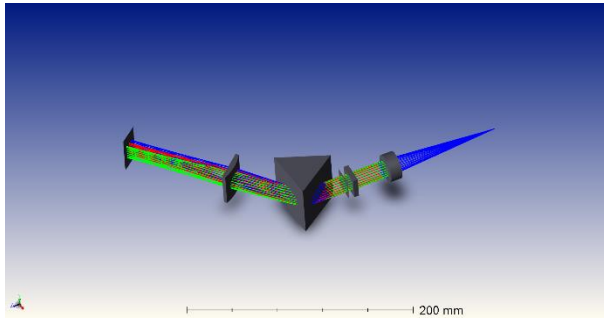
Figure 4-1. Left: Impression of the optical design of the spectropolarimeter. The incoming light is on the right, the detector on the left. Right: Rendered image of the optical and mechanical parts of the breadboard spectropolarimeter. Light enters on the left; the detector is mounted in the red cylinder on the right.

For laboratory use it is possible to place a microscope objective instead of a telescope in front of the pinhole. This allows characterisation of satellite mock-ups or flight spare models.

The optical and mechanical parts are all commercially off the shelf components, apart from the polarisation modulator. A rendered image of the breadboard spectropolarimeter (without the light tight enclosure) is shown in Figure 4-1 (right).

5. TEST BENCH DESIGN

The characterisation of the birefringent properties of the modulator is inspired by the one described by [10], and its implementation will be similar to the one presented in [11]. The purpose is to measure the phase birefringence ($\Delta\phi_1, \Delta\phi_3$) of the prisms and the fast axis orientation of the prisms constituting the modulator (α_1, α_3). The optical setup is presented in Figure 5-1.
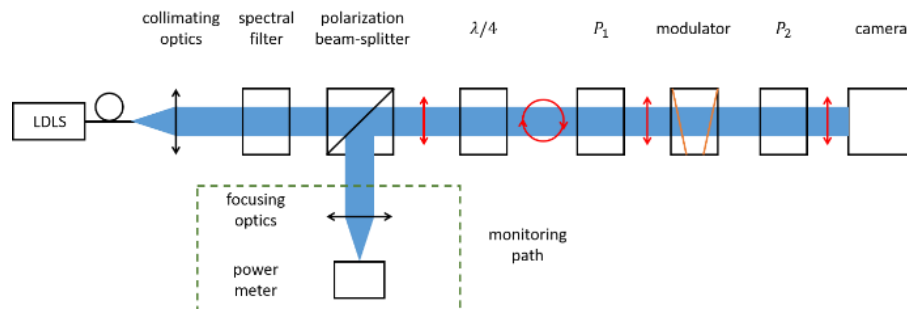

Figure 5-1. Scheme of the polarisation characterisation measurement setup, the beam is pictured in blue, the red arrows represent the polarisation at different places of the beam.

The setup is composed of:

- a fibered laser driven light source (LDLS), emitting light from 300 to 1500 nm,
- a spectral filter to define a small bandwidth for the measurement,
- collimating optics to convert the diverging beam at the end of the fibre into a collimated one,
- a polarisation beam-splitter, to provide a monitoring path and to define a reference polarisation
- a power meter for monitoring the power of the beam to correct for intensity variation during the acquisition
- a quarter wave-plate ($\lambda/4$) to create a circular polarisation,
- 2 linear polarisers on automated rotation stages (P_1 & P_2),
- the prism modulator,
- a CMOS camera.

The measurement principle is the following one. For a specific bandwidth, the transmission axis of P_1 will perform a full rotation by increments of $10°$. For each orientation of P_1 (θ), two maps of intensities will be recorded: $I_\parallel(x, y, \theta)$ & $I_\perp(x, y, \theta)$. $I_\parallel(x, y, \theta)$ corresponds to the intensity map recorded with the transmission axis of P_2 is parallel to the one of P_1 and $I_\perp(x, y, \theta)$ stands for the intensity when P_2 axis is perpendicular to P_1. Afterwards, the ratio of intensity Γ is computed as equation 1.

$$\Gamma(x, y, \theta) = \frac{I_\perp(x, y, \theta)}{I_\parallel(x, y, \theta)} \tag{5}$$

Using Mueller matrices for the polarisers and prisms, Γ can be expressed as a function of $\theta, \Delta\phi_1, \Delta\phi_3, \alpha_1$ & α_3. Then, by fitting this equation pixel by pixel, using θ as a variable, the values of the phase retardances and fast axis orientation can be computed (see Figure 5-2 for some curves of Γ for the wavelength of 633 nm for several heights in the modulator). Finally, those experimental values of the phase retardances and the fast axis orientations will be used as parameters for the measurement of the polarisation of incident beams during the real measurements.
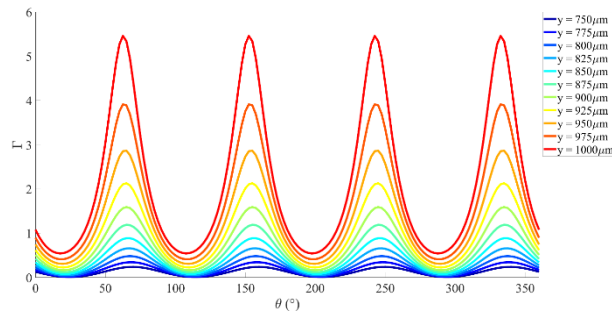

Figure 5-2. Representation of the intensities ratio Γ for pixels on the same column and at different height inside the modulator. The representation step was selected as 25μm to provide more variation with fewer curves. The real system will be on the 5μm range for the pixel size.

6. DATA ACQUISITION AND ANALYSIS PLANS

Once the spectropolarimeter breadboard is complete and the polarisation modulator has been characterised, the first step is to characterise the end-to-end performance of the breadboard with known polarisation sources. After that, the instrument will be equipped with a microscope objective in order to image a mock-up of a satellite onto the spectropolarimeter pinhole and simulate a point source observation. This will allow the first representative measurements with full control over the materials and the illumination-, scattering-, and viewing geometry.

The final test will be to mount the breadboard on a telescope with 800 mm diameter provided with a fast and accurate mount that was specified to track Low Earth Orbit satellites (Figure 6-1). Observations of sufficiently bright satellites (e.g. magnitude 9 and brighter) should yield sufficient quality measurements that allows in principle to reach normalised Stokes vector values to be assessed with about 0.001 precision or better, at the equivalent of 50 nm spectral resolution.
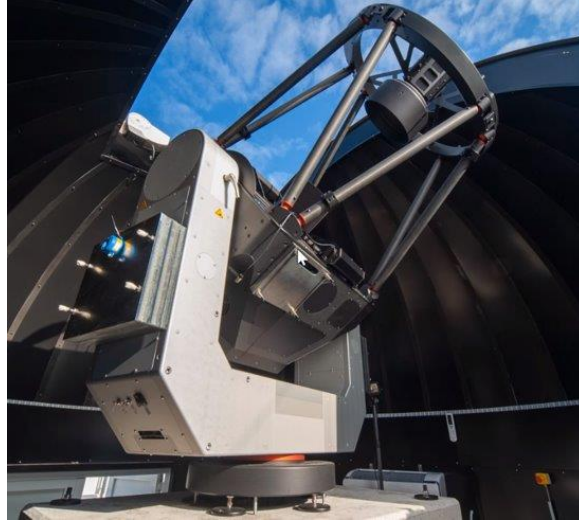
Figure 6-1. The 800 mm diameter telescope at TNO the Hague, to be used with the spectropolarimeter.

Data analysis will initially focus on understanding the measurement results, in particular the impact of the telescope and environmental parameters (temperature). As we expect from the telescope design that it will introduce complex instrumental polarisation [12], the data analysis will consider these instrument effects as well. A likely approach will be to analyse the measurements and express them not only in Stokes parameters $[S_1, S_2, S_3, S_4]^T$, but also express them in degree of total polarisation $p = \frac{1}{S_1}\sqrt{(S_2^2 + S_3^2 + S_4^2)}$, and angles χ and ψ on the Poincaré sphere, all as a function of wavelength.

As the different materials involved in scattering of sunlight at the satellite will affect the polarisation of the light, so will the mirrors in the telescope and the temperature effects of the polarisation modulator. The differences between the satellite and the instrument are that the reflection geometry and materials at the instrument side are well known, so any instrumental spectropolarimetric signatures can in principle be identified. This allows either correction of the signatures, or inclusion in an instrument model as known parameters that can thus separate the effects from the spectropolarimetric signatures of the incoming light.

A next layer of analysis could be interpretation of the spectropolarimetric results as visible fractions of materials at the satellite surface, similar to being done in [13], but in this case also using the polarimetric signature to help distinguish between different materials and even get an indication of multiple scattering.

The intention of these measurements and analyses is to investigate the potential of spectropolarimetry for space object identification and create a roadmap for further development.

7. SUMMARY AND DISCUSSION

We present our concept and breadboard design for a full Stokes spectropolarimeter intended for space object identification. The instrument concept allows measurement of point sources and separates the polarisation and spectral properties into two perpendicular dimensions on a 2-dimensional detector. Initial calculations suggest that geostationary satellites (typically with a brightness of stellar magnitude 9) could be measured with sufficient precision to allow useful investigations into means to further identify space objects. Brighter satellites can be measured with more ease, while fainter satellites could also be measured at the cost of temporal, spectral and polarisation resolution. Instrumental polarisation aspects are considered a challenge, but a methodology for a way forward to deal with this has been identified.

Currently, all hardware for the spectropolarimeter breadboard is in house and integration of the breadboard is foreseen to start this year, with first light intended to be obtained also this year.

The breadboard measurements will indicate whether the approach is sound, and in that case will allow detailing of a roadmap to arrive at a mature instrument and processing chain that can be used for Space Situational Awareness.

REFERENCES

[1] Ivan Ferrario, Eugenio Di Iorio, Fabrizio Silvestri, Ralph Snel, Quentin Chavet, Optical Technologies for Space Situational Awareness, 2nd ESA NEO and Debris Detection Conference, 2023

[2] R. Snel, B. Vasilescu, I. Ferrario, J. Loicq, P. Piron, E. Di Iorio, T. Hooftman, Q. Chavet, Full Stokes Spectropolarimetry for Space Object Identification, 2nd ESA NEO and Debris Detection Conference, 2023

[3] Endo, T., Ono, H., Hosokawa, M., et al. "Spectroscopic characterization of GEO satellites with Gunma LOW resolution Spectrograph" (2017), Advanced Maui Optical and Space Surveillance (AMOS) Technologies Conference, 2017amos.confE..81E

[4] Speicher, A., "Identification of Geostationary Satellites Using Polarization Data from Unresolved Images" (2015). Electronic Theses and Dissertations. 1050. https://digitalcommons.du.edu/etd/1050

[5] SkywalkerPL - Own work, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=45634546

[6] B. Vasilescu, Y. Nazè, and J. Loicq, "Solution uniqueness and noise impact in a static spectropolarimeter based on birefringent prisms for full Stokes parameter retrieval," J. Astron. Telesc. Instrum. Syst. 6, 1 (2020).

[7] M.J. Dodge, "Refractive properties of magnesium fluoride",Appl.Opt.23,1980–1985(1984).

[8] H. H. Li, "Refractive index of alkaline earth halides and its wavelength and temperature derivatives", J. Phys. Chem. Ref. Data 9, 161–290 (1980).

[9] J. C. del Toro Iniesta and M. Collados, "Optimum modulation and demodulation matrices for solar polarimetry", Appl. Opt. 39, 1637 (2000)

[10] Emoto, A., Nishi, M., Okada, M., Manabe, S., Matsui, S., Kawatsuki, N., & Ono, H. (2010). Form birefringence in intrinsic birefringent media possessing a subwavelength structure. Applied optics, 49(23), 4355-4361.

[11] Piron, P., Catalan, E. V., Absil, O., & Karlsson, M. (2018). Birefringence measurements of diamond space-variant subwavelength gratings. Applied Optics, 57(17), 4909-4917.

[12] de Juan Ovelar, M., Snik, F., Keller, C. U. Venema, L., "Instrumental polarisation at the Nasmyth focus of the E-ELT", A&A, 562 (2014) A8, DOI: https://doi.org/10.1051/0004-6361/201321717

[13] Vasile, M. & Walker, L. & Dunphy, D. & Zabalza, J. & Murray, P. & Marshall, S. & Savitski, V. (2022). Intelligent Characterisation of Space Objects with Hyperspectral Imaging, Acta Astronautica, Volume 203, 2023, Pages 510-534, ISSN 0094-5765, https://doi.org/10.1016/j.actaastro.2022.11.039

Performance studies of two active SWIR instrument designs for vision enhancement in indoor applications

Enno Peters[a], Max Anders[a], Tristan Preis[a], Laura Kontschak[a],
Jendrik Schmidt[a], and Maurice Stephan[a]

[a]Institute for the Protection of Maritime Infrastructures, German Aerospace Center (DLR),
Fischkai 1, Bremerhaven, Germany

ABSTRACT

Two compact and portable SWIR active imaging instrument configurations aiming at vision enhancement in indoor applications are tested and compared, working at 1300 nm and 1550 nm, respectively. Both configurations are in-house developments, but based on a limited number of standard and commercially available components (cameras, LEDs). The instruments provide images (640×512, resp. 1280×1024 pixels) at a rate of ca. 17 Hz (live stream) that can be displayed either directly on an integrated display or send via (wireless) network. Key specifications (optical power, field of view, heat development) have been characterized in laboratory tests. The performance of the two system configurations in terms of vision enhancement is compared both practically (field tests) and theoretically (Mie scattering theory). The 1300 nm illuminator has almost double power compared to the 1550 nm illuminator. However, Mie calculations predict more backscatter and less transmission through fog and smoke, which is highly depending on the particle size. Field tests using artificial fog and an in-house developed transmissiometer have been performed to validate the findings from modeling and found a vision enhancement in the order of one magnitude due to use of SWIR (instead of the Visible) for use in typical environments for which the instruments are designed for. A substantial additional improvement in terms of vision enhancement could be achieved by using polarized light and polarization optics to reduce the backscatter signal. In contrast to other research studies, this vision enhancement is not based on polarization difference imaging but on reducing the backscatter component only, enabling a robust and simple system design.

Keywords: Short-wave infrared (SWIR), active imaging, sensor development, Mie scattering, vision enhancement, linear polarization, circular polarization

1. INTRODUCTION

Limited visibility in indoor environments, e.g. caused by smoke, is a severe challenge for operations of emergency and rescue services such as fire fighters or the police. Active optical systems working in the short-wave-infrared (SWIR) spectral range are a powerful tool for vision enhancement in those conditions because of the increased transmission through fog and smoke in the SWIR range compared to the visible spectral range. At the same time, photons in the SWIR range are invisible for the human eye and therefore enable the operator in darkness to see without being seen, which is an added value for police operations.

In a previous study,[1] a cost-efficient and portable SWIR sensor design using 1550 nm illumination was introduced. The whole instrument housing plus interior (mounts for camera, illuminator etc.) is 3D printed. A flat touch screen computer controls camera, illumination, and displays the images. Internal communication and power supply are mainly via USB. The focus was on simple, lightweigth, compact and robust design. In this study, the original instrument is updated in terms of camera, illumination unit, instrument case and thermal management. In order to optimize the performance, two kinds of SWIR LEDs (center wavelength 1300 nm and 1550 nm) as well as two SWIR cameras were compared during a field test.

The following Sect. 2 gives a recap of the original 1550 nm instrument design and introduces the updated one, including laboratory measurements of the instrument characteristics. Theoretical considerations on the vision enhancement of the instrument based on Mie calculations are presented in Sect. 3. The vision enhancement is

---

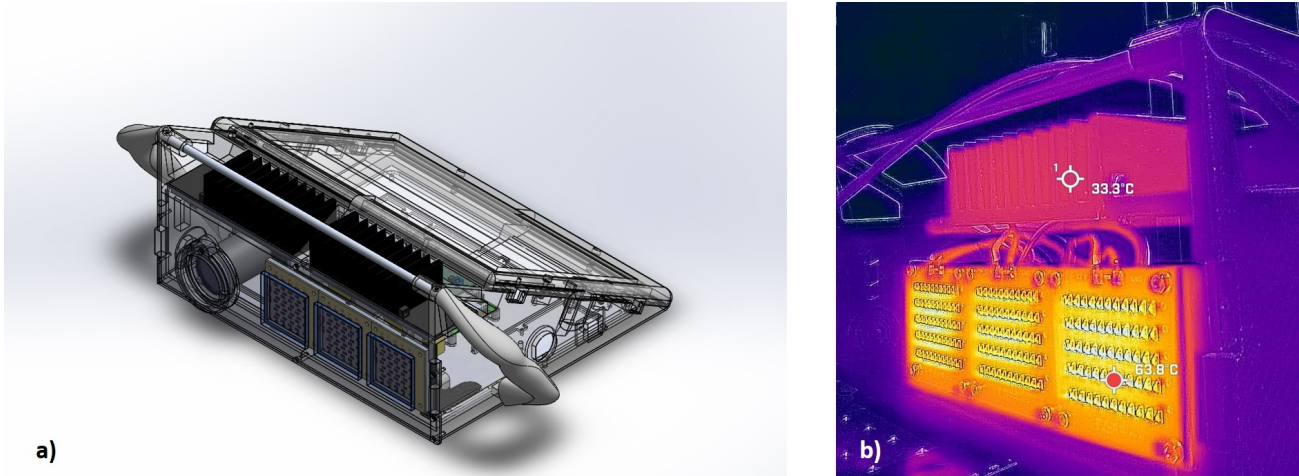Correspondence to: E. Peters (E-mail: Enno.Peters@dlr.de)

Figure 1. a) 3D drawing of the updated instrument using a SenS 1280 SWIR camera and 3 LED sources for active imaging. b) Thermal image of the instrument after several minutes of operation. The main heat sources are the SWIR LEDs. Clearly visible are the heatpipes connecting the backside of the LED panels to heatsinks mounted on top of the instrument. The heatsinks reach temperatures of $\sim 35°$ C, keeping the LEDs $< 85°$C and avoid overheating of the instrument's electronics.

then practically evaluated and compared during a field test in Sect. 4 using a commercial fog machine. During the field test, the extinction was measured using a self-made transmission sensor in the Vis and SWIR spectral range. Additionally, the improvement in terms of vision enhancement brought by polarization optics was investigated.

2. INSTRUMENTS

The original 1550 nm system is explained in detail in Mischung[1] et al. (2022). To summarize, it is based on a WiDy SenS 640 SWIR camera from New Imaging Technologies having a resolution of 640×512 pixels (15 μm pixel width) and a 12 mm F1.4 SWIR lens from EHD imaging GmbH giving an angular field of view (FOV) of 44.6° horizontally. Active illumination is based on 55 single SWIR LEDs (model MTSM5015-844-IR from Marktech Optoelectronics) emitting at approx. 1550 nm. The emitted optical power is $\sim$0.28 W. The angular field of view of the illumination is $33.5° \pm 4.8°$, which is by purpose smaller than the camera's FOV as scattering in smoke is expected to effectively broadening the angular FOV of the emission. An integrated processing unit displays the image on a 10 inch waterproof display.

For updating the instrument, a different camera and illumination unit was tested. These components, and any combination of them, have been evaluated during a field test (Sect. 4). In particular:

- To increase the illumination, three boards of 1550 nm LEDs (as described above) are used, resulting in a total optical power of 3×0.28 W $= 0.84$ W.

- Alternatively, a 1300 nm illumination unit was developed having three boards of 55 individual SWIR LEDs each (model MTSM0013-844-IR from Marktech Optoelectronics). The emitted optical power was measured in laboratory tests and is $\sim$0.49 W per LED panel, i.e. $\sim$1.47 W in total and therefore substantially larger than the emission of the 1550 nm illumination unit due to better efficiency. The angular FOV of the 1300 nm illumination has been measured to $46.8° \pm 2.0°$, which is larger than the 1550 nm illuminator and thus enabling combination with a wider camera FOV.

- A SenS 1280 SWIR camera from New Imaging Technologies having 1280×1024 pixels (10 μm pixel width) increases the resolution and/or FOV. When used with a 16 mm F1.4 SWIR lens (from EHD imaging GmbH), the same angular FOV of the original instrument is obtained while the resolution is doubled. However, in combination with the wider 1300 nm illumination unit, a lens with smaller focal length could
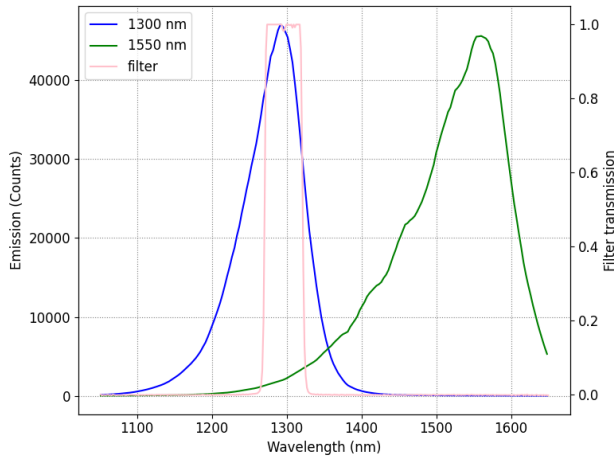
Figure 2. Emission spectra of the 1300 nm and the 1550 nm illumination units and spectral transmission of the optionally tested 1300 nm spectral filter.
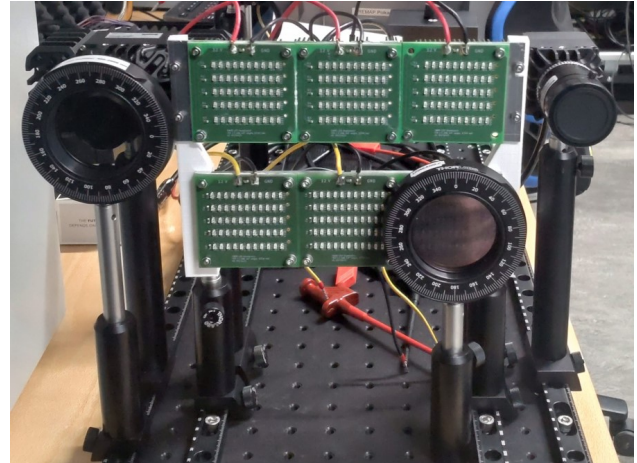
Figure 3. Test setup on a breadboard used for the field test. The SenS 1280 camera is on the left (here with optional polarization optics), the Widy Sens 640 on the right. The three top LED units emit at 1300 nm, the bottom units at 1550 nm.

be used. Nevertheless, throughout this study, the 16 mm lens giving the same FOV has been used, to simplify the comparison.

In addition, the instrument's case was updated and is more ergonomically shaped for increased user comfort (Fig. 1a). It is equipped with a larger 13 inch display with Full HD+ resolution in order to display the larger image of the SenS 1280 camera. The framerate is mainly limited by the computational power of the integrated computer and typically in the range of 16-17 Hz, which is sufficient for addressed applications as providing a smooth live stream. Alternatively to displaying the image on the integrated screen, data can be send via (wireless) network, enabling to operate the instrument remotely. To overcome internal heating, predominantly caused by the LEDs, the instrument is updated with heat pipes connecting the LEDs with heatsinks mounted on top of the instrument's case (Fig. 1b).

The spectral emission of both illumination units is plotted in Fig. 2. When operated with the 1300 nm unit, the instrument can be optionally equipped with a bandpass filter, which is installed in front of the camera, to block ambient light. At the same time the rather broad spectral emission is reduced predominantly at smaller wavelengths which are considered to be preferentially backscattered.

For the field test in Sect. 4, the illumination units and cameras have been operated without the instrument case. Instead, they were mounted on a breadboard allowing for same distances between illumination units and cameras (Fig. 3). Also, this enables to test any combination of camera and illumination unit (SenS 1280, Widy SenS 640, 1300 nm and 1550 nm illumination). Some of the measurement sequences during the field test were performed with polarized light. For these measurements, polarization filters and quarter-waveplates were placed in front of the SenS 1280 camera. Due to their spatial extent, polarization optics could be placed in front of a single LED board only (consisting of 55 individual LEDs). The two LED boards without polarization optics were switched off for these tests, so that all light in the respective experiments was polarized.

3. THEORETICAL CONSIDERATIONS

3.1 Vision enhancement due to SWIR

In the field test discussed in Sect. 4, a commercial fog machine using a propylene glycol, triethylene glycol and water mixture was used. The produced droplets have a slightly higher refractive index than pure water (due to the glycol) resulting in more efficient light scattering,[2] and they are smaller in size than droplets in natural
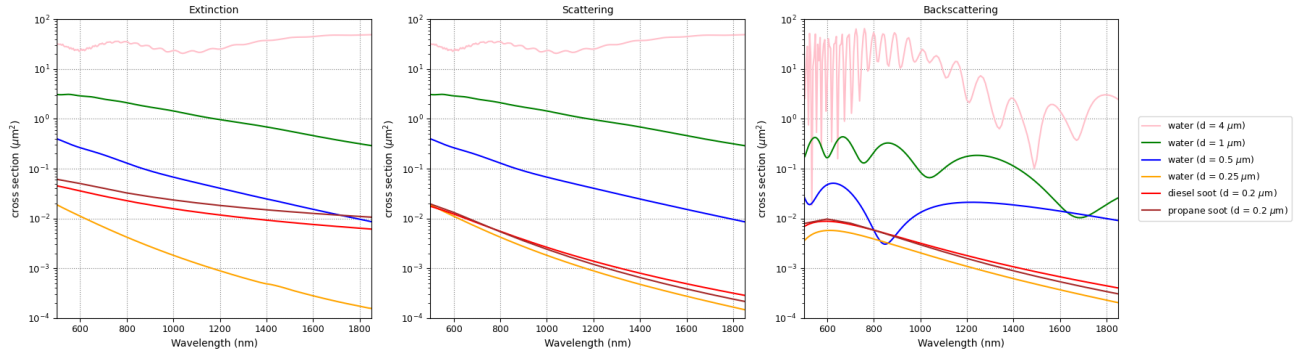
Figure 4. Cross sections (extinction, scattering and backscattering) from Mie calculations for different aerosol types and particle diameters.

fog. While droplets in real fog are typically in the range of 2 μm (continental fog) to 10 μm (maritime fog),[2] we could not detect any fog signal with an *Alphasense OPC N3* particle detector having a particle range from 0.35-40 μm. As a consequence, we estimate the particle size to be < 0.35 μm. To evaluate how the findings derived from the field test in Sect. 4 (small droplets) relate to real-world scenarios (larger droplets), Mie calculations were performed using the *Python* module *miepython* (https://miepython.readthedocs.io/en/latest/). In particular, the extinction, scattering and backscattering cross sections for water droplets have been calculated as a function of wavelength for different droplet sizes ranging from 0.25 μm (which is assumed to be representative for the droplet size during the field test) to 4 μm (Fig. 4). The complex refractive index for water used as input for Mie calculations was taken from literature.[3,4] As the SWIR instrument is intended for applications in different environments, Mie calculations were performed for diesel[5] and propane[6] soot as well, which are dry particles and smaller in size. In Fig. 4, 0.2 μm diameter was assumed for soot calculations, which is a typical value for dry condensation nuclei that we obtained from long-term measurements using a Palas Fidas 200 S aerosol spectrometer at our instrument rooftop, where we expect diesel soot from (marine) traffic.

As can be seen from Fig. 4, extinction, scattering and backscattering generally increase by magnitudes with increasing droplet size, making the conditions using artificial fog during the field test in Sect. 4 somewhat an ideal case. However, every real fog will consist of a distribution of particle sizes including small particles.

The basic finding from Fig. 4 is that all quantities (extinction, scattering, backscattering) decrease with wavelength, but the slope is larger for small droplets and decreases with increasing particle size until no slope is present anymore for droplets of 4 μm. Consequently, vision enhancement can be achieved using optical instruments in the SWIR instead of the Visible (Vis), but this vision enhancement is stronger when particles are small. Therefore, SWIR systems will have less advantage over Vis instruments in maritime environments where droplets are typically large, while the advantage of using SWIR instead of Vis in continental environments is larger.

An interesting feature in Fig. 4 is that for all water droplets, the extinction and scattering graphs look very similar, while they differ for soot. This is because scattering is the dominant contribution to extinction in case of water particles (photons are scattered out of the line of sight), while soot has much more loss due to absorption. As a consequence, the decrease in extinction as a function of wavelength is less steep for soot compared to small water droplets. However, the extinction of soot at 1550 nm is almost one order of magnitude smaller than at 550 nm, and more than one order of magnitude for backscattering, which clearly demonstrates the vision enhancement due to SWIR instruments in continental or indoor environments where smaller water aerosols or soot will be the dominant aerosol type.

Because of the general decrease of extinction, scattering and backscattering, illumination at 1550 nm can be expected to perform slightly better in terms of vision enhancement compared to 1300 nm. However, the 1300 nm LED illumination has almost 2 times more optical power as explained in Sect. 2, which is beneficial to keep exposure times short in environments having large extinctions.
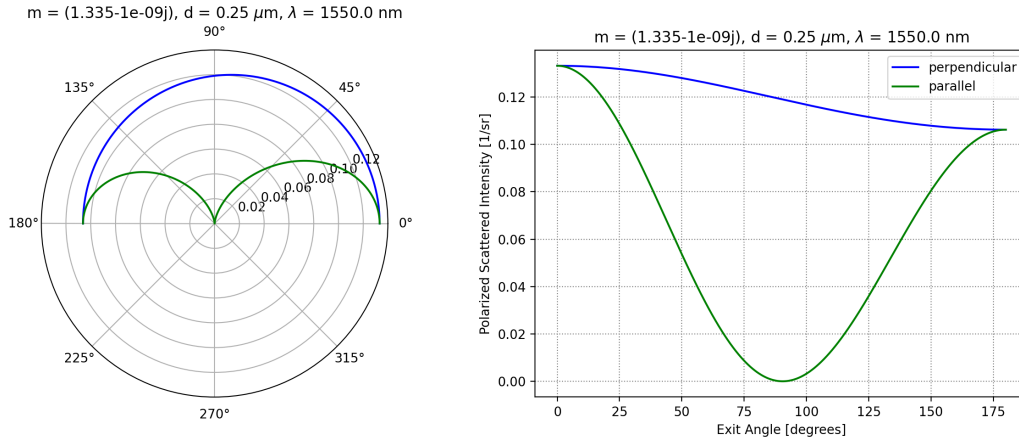
Figure 5. Simulated phase function for the parallel and perpendicular polarized component of scattered light of 1550 nm on water droplets of 0.25 μm diameter, which is the assumed size of droplets during the field test in Sect. 4.

3.2 Vision enhancement due to polarization

Use of polarized light can support vision enhancement applications by reduction of the backscatter signal. The idea is, if backscatter events preserve the polarization, then this component can be blocked. The (ideal) depolarization ratio δ of water droplets is $\delta = 0$ for direct backscatter (i.e. scattering in the angle $\vartheta = 180°$), which can be seen from Fig. 5 showing the phase function of the perpendicular (to the scattering plane) and parallel components of scattered light. Again, these calculations have been performed with *miepython*. As the wavelength (here 1550 nm) is larger than the particle size (here 0.25 μm), the phase function is rather Rayleigh-like with some increasing forward scattering (forward scattering increases with increasing particle size). While the perpendicular and parallel components show strong differences for most scattering angles (the physical explanation is that both components are produced by induced dipoles that emit dipole radiation, i.e. no emission in the direction of oscillation), the scattering probabilities in 180° are identical, meaning that backscattering will not change the polarization of incoming light. However, some depolarization occurs in angles close to 180°, which is in agreement to more detailed studies[7] reporting that depolarization still occurs in angles close to 180°. Nevertheless, using linear polarized light and a polarization filter in crossed direction in front of the receiver will effectively block the backscattered light, while diffuse reflection on targets leads to depolarization making them still visible. While water droplets are somewhat an ideal case, small dust-like aerosols and ice-crystals have small depolarization ratios in backward directions as well,[8] thus the technique will work for a variety of difficult visibility environments to different extents.

Other research studies[9,10] have shown that circular polarization has a better persistence when transmitting smoke and fog than linear polarization, but the performance depends on particle size distribution, fog density and wavelength. Furthermore, it was recently experimentally demonstrated[11] that circular polarized light enhances range and contrast in fog (aiming to enhance road safety and support autonomous vehicles/systems) using polarization imaging techniques that are based on measuring the difference between two polarization states. In contrast to these studies, our instrument is not doing polarization imaging. The objective here is to use a conventional camera in combination with polarized illumination and optics in order to block backscattered light as explained above.

However, circular polarization has some further advantages: (1) It is not common in nature,[12] while linear polarization is produced by reflection (Fresnel coefficients) on surfaces; (2) relative changes in orientation between emitter and receiver (i.e. tilt) will affect techniques using linear polarization, but not circular polarization. On the other hand, the main drawback is that usually quarter-waveplates are used, which is an additional optical component and thus increasing costs and complexity (although cost-efficient and ready-to-use combinations of linear filter and quarter-waveplate exist). Furthermore, quarter-waveplates are designed for a limited wavelength range and thus circular polarization is produced for these specific wavelengths only. This limits the efficiency in terms of vision enhancements when active illumination has a broader emission spectrum. In the field test

Figure 6. Example images without smoke of the observed scene during the field test acquired with the Widy SenS 640 (a) and the SenS 1280 (b) using unpolarized illumination. Subplot (c) is the same as (b), but using the SenS 1280 with linear polarized illumination and a crossed polarization filter in front of the camera, i.e. observing diffuse illumination only.

(Sect. 4), three measurement sequences using circular polarization were performed to test the vision enhancement that can be achieved by combining the SWIR LED illumination with circular polarization.

4. FIELD TEST

4.1 Measurements

A field test was carried out on 24 May 2023 in Bremerhaven, Germany, using the breadboard-setup of the instruments as described in Sect. 2 and shown in Fig. 3. The whole test was performed inside a black tent to achieve (1) darkness and (2) prevent the smoke produced during the test to dilute or be blown away by wind. A rescue dummy wearing a life jacket with retroreflectors was placed next to a mount with reflective targets (95%, 50%, 20% and 5%), in a distance of approximately 5 m from the SWIR instruments. These targets were used to calculate the Michelson contrast in acquired images under different fog concentrations. Another dummy with helmet and safety vest equipped with retroreflectors was placed on the other side of the reflectivity panels. Fig. 6 shows images of the test setup without artificial fog (prior to the measurements) acquired by the Widy SenS 640 (a) and the SenS 1280 (b) using unpolarized illumination, as well as the SenS 1280 (c) equipped with a linear polarization filter and using polarized illumination.

Without polarization, retroreflectors lead to strong blooming effects, which disappeared completely when using polarization because the retroreflectors used here obviously preserve polarization. Specular reflection on metallic surfaces (e.g. the tripod) disappeared as well, and only diffuse reflection leading to depolarization is present in image (c). Interestingly, while blooming effects are present in both images (a) and (b), the effect is much stronger in the SenS 1280 image (b). This is most likely because of the smaller pixel size (10 μm) of the SenS 1280 leading to a smaller full well capacity (10 000 e$^-$ for the high gain mode used here) compared to the Widy SenS 640 (15 μm and 17 000 e$^-$).

During the field test, 6 measurement sequences were performed each consisting of 10 individual images acquired at different fog concentrations produced by a commercial theater fog machine (Eurolite Dynamic Faze 1000). Between measurement sequences, the tent was ventilated to remove all fog. Within each sequence, the fog concentration was constantly increased from image to image. The extinction as a function of wavelength in the Visible and SWIR spectral range was measured using a self-build transmission instrument explained in detail in a previous publication.[1] To summarize, it consists of two collimated beams each having a $d = 30$ cm long free optical path exposed to fog, and two spectrometers for the Visible and SWIR. An initial spectrum I_0 was acquired at clean air before starting the fog machine in each sequence. Another spectrum I was then recorded for each camera measurement. Using Beer-Lambert's law

$$I = I_0 \cdot e^{-\sigma(\lambda) \cdot d} \tag{1}$$

the extinction $\sigma(\lambda)$ in units m^{-1} was obtained as a function of wavelength λ. Fig. 7 shows exemplarily the obtained extinctions for measurement sequence 1. In this case, maximum extinctions of $\sigma(\lambda) \approx 3$ m^{-1} at
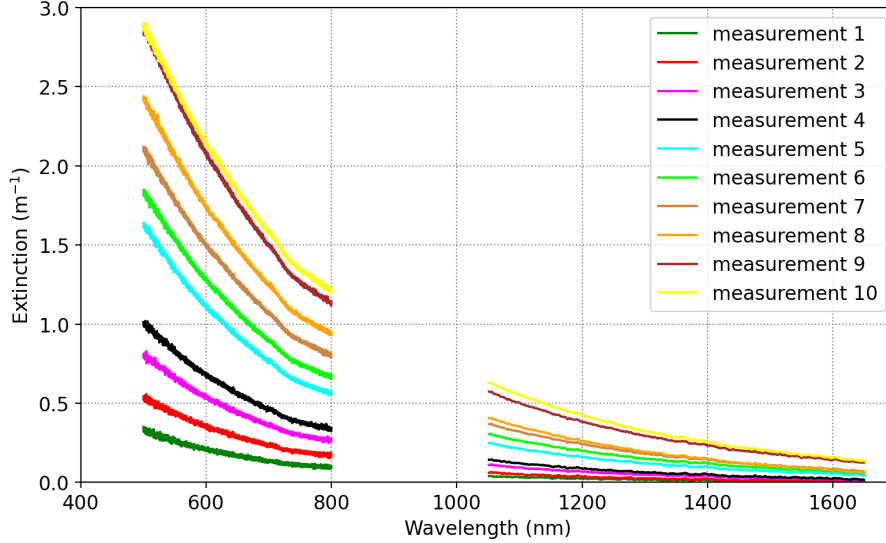
Figure 7. Extinction as a function of wavelength during measurement sequence 1. Two spectrometers were used to cover the spectral range from 550-800 nm and 1050-1650 nm, respectively. The gap between 800-1050 nm is not covered by any spectrometer. The extinction is increasing from measurement to measurement as fog was produced during the sequence. The advantage of using SWIR illumination (here around 1300 nm and 1550 nm) instead of Visible light is obvious.

$\lambda = 500$ nm were produced. The extinction strongly decreases with wavelength, which is in agreement to Mie calculations in Sect. 3. The ratio of the measured extinctions at 550 nm and 1550 nm is on average ≈ 18, which agrees roughly to Mie calculations of water droplets of 0.5 μm, which is a bit larger than the expected particle size, either due to the glycol content or measurement uncertainties of the transmission sensor.

In measurement sequence 1, images of both cameras were acquired using the 1300 nm illumination unit. The same test was repeated in the second measurement sequence, but using 1550 nm illumination. Sequence 3 was again performed using 1300 nm illumination, but only images from the SenS 1280 were recorded, which was equipped with the 1300 nm spectral filter shown in Fig. 2. The additional vision enhancement brought by linear polarization was investigated in sequence 4, while sequences 5 and 6 were investigating circular polarization. Tab. 1 summarizes the different measurement sequences performed during the field test.

4.2 Results

Example images of measurement sequence 1 acquired by the Widy SenS 640 are shown in Fig. 8. Backscattering is increasing with increasing extinction (fog concentration), most dominant at the right edge of the images, which

Table 1. Measurement sequences performed during the field test.

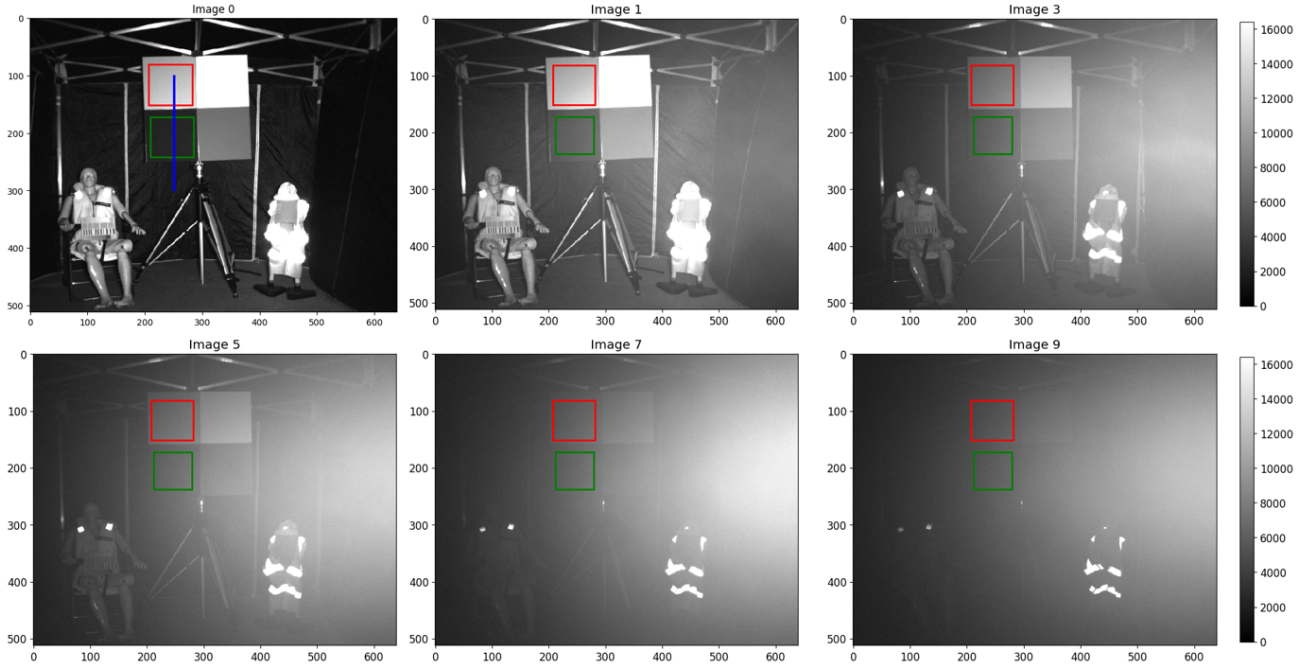| Sequence	Illumination	Camera	Polarization
1	1300 nm	both	unpolarized
2	1550 nm	both	unpolarized
3	1300 nm with spectral filter	SenS 1280	unpolarized
4	1300 nm	SenS 1280	linear
5	1300 nm	SenS 1280	circular (left-handed)
5a	1300 nm	SenS 1280	circular (right-handed)
6	1550 nm	SenS 1280	circular (left-handed)

Figure 8. Example images taken with 1300 nm illumination and the Widy SenS 640 in measurement sequence 1 for increasing particle concentration. The 50% reflectivity panel as well as the 5% panel used for calculation of the Michelson contrast are indicated by a red and green box, respectively. In addition, a blue line indicates cross sections plotted in Fig. 9.

is where the LED panels were installed. Red and green boxes indicate regions of interest (ROI) across the 50% and 5% reflectivity panel. Gray values within these boxes have been averaged to calculate the Michelson contrast

$$C_{Michelson} = \frac{I_{\text{mean, 50\%}} - I_{\text{mean, 5\%}}}{I_{\text{mean, 50\%}} + I_{\text{mean, 5\%}}} \tag{2}$$

The intrinsic contrast between these panels is $\frac{50-5}{50+5} \approx 0.81$. A blue line in the first image of Fig. 8 indicates the position of cross sections shown in Fig. 9. A sharp decrease is visible in cross-sections of images without fog (or little fog) at the transition from 50% to 5% panel, and even the black wall can be distinguished from 5% panel in the cross-section without any fog. When the fog concentration increases, (a) gray values across the dark 5% panel increase as a result of backscattering (that is also clearly visible in the example images in Fig. 8), and (b) the gray values across the brighter 50% panel decrease as a result of extinction. Therefore the height of the step between 50% and 5% panels in Fig. 9 decreases until no transition is visible anymore for the last images of sequence 1 having the largest fog particle concentration.

As a consequence of this behavior, the Michelson contrast plotted in Fig. 10 is decreasing as a function of extinction. Note, as the Beer-Lambert law (eq. 1) used to calculate extinctions (on the x-axis) requires two input spectra and the first spectrum (I_0) was recorded together with the first image in each sequence for clean air conditions, no extinction could be calculated for the first image. Instead, the meteorological extinction of the day of the field test was used as extinction (x-axis value) of the first data point.

In Fig. 10, even negative contrasts are obtained for large extinctions. These are unphysical and an artifact of the inhomogeneous illumination, which is gauss-shaped with maximum intensity in the center of the image. As the 50% panel is in the upper half of the image, it is exposed to less light than the 5% panel. This inhomogeneous illumination can be seen also directly in cross-sections (Fig. 9) or example images (Fig. 8) when the backscattering is large. As a consequence, the obtained Michelson contrast in image 0 (clean air) is smaller than the intrinsic
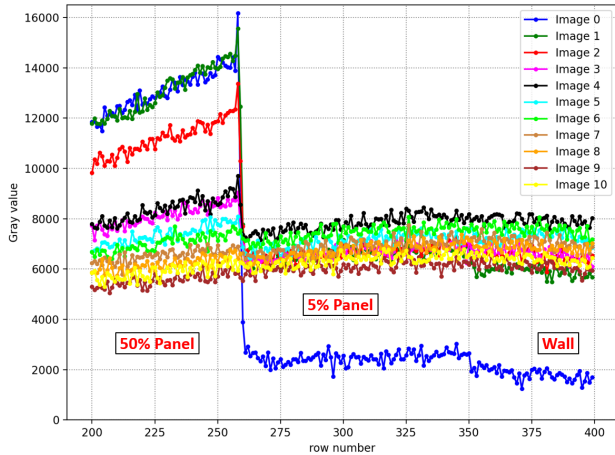
Figure 9. Cross sections of gray values along the blue line indicated in Fig. 8 during measurement sequence 1. Regions of the 50% panel, the 5% panel as well as the wall of the (black) tent are labeled.
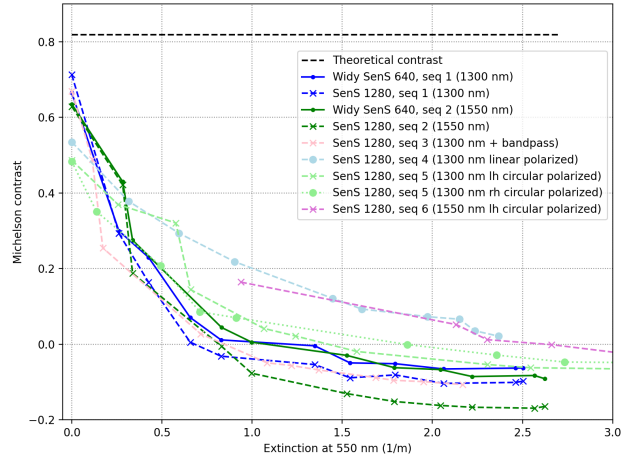

Figure 10. Michelson contrast calculated from 50% and 5% reflectivity panels as a function of the measured extinction at 550 nm (which roughly corresponds to the naked eye). Note, the extinctions in sequences 5 and 6 exceed the plotted range of the x-axis, which is restricted to $3 m^{-1}$ for better comparison.

contrast calculated by the reflectivity of the panels. When backscattering becomes dominant, it is brighter across the 5% panel compared to the 50% panel, resulting in the observed negative contrasts for large extinctions.

Fig. 10 also shows that the contrast obtained from SenS 1280 images are systematically slightly smaller than contrasts derived from Widy Sens 640 images. This is again an artifact of the illumination: The Sens 1280 is on the right side of the illumination unit and, as a consequence, largest backscattering occurs on the left edge in Sens 1280 images. In contrast, it occurs at the right edge of Widy SenS 640 images as seen in Fig. 8. As the 50% and 5% panel used to calculate the contrast are located closer to the left edge of the images, the Sens 1280 is affected more by backscattering compared to the Widy Sens 640 and consequently obtained Michelson contrasts are systematically smaller. In reality however, both cameras are expected to behave similarly in fog conditions. In cases of inhomogeneous fog, the SenS 1280 is even expected to perform better because of its better resolution potentially enabling it to look through areas of smaller particle concentration in fog pitches. On the other hand, the better full well capacity of the Widy Sens 640 avoids blooming effects, which potentially do not only occur with retroreflectors (as seen in Fig. 6) but at intense backscattering of highly reflective particles and large concentrations as well. Thus, both cameras have advantages.

The contrast obtained in measurement sequence 2 (1550 nm, green lines in Fig. 10) generally resembles the values of sequence 1, both for the Widy Sens 640 and the SenS 1280. For contrast at extinctions < 1 m^{-1}, sequence 2 mostly exhibits slightly larger contrasts than sequence 1, which is in agreement with Mie calculations (see Sect. 3) suggesting slightly better performance of 1550 nm in the case of scattering at very small particles. For extinctions > 1 m^{-1}, sequence 1 (1300 nm) has larger contrasts, but values are negative and unphysical (as explained above), so no conclusions should be drawn from this.

Measurement sequence 3 (1300 nm + bandpass filter) performs similarly to sequence 1, and therefore no vision enhancement (as speculated) is obtained by use of the spectral filter. Another drawback of the filter is a reduction of intensity and thus longer exposure times are needed. However, in real applications an advantage is blocking of ambient light sources, and therefore optional use of the filter appears still reasonable.

The largest improvement in terms of contrast is achieved by using linear polarization (light blue line in Fig. 10). Although still exponentially decreasing, the slope is much smaller compared to sequences without use of polarization, and it is the only line without the contrast dropping below 0. The reason is that backscattering is eliminated almost entirely, which is demonstrated in Fig. 11 showing example images with and without use
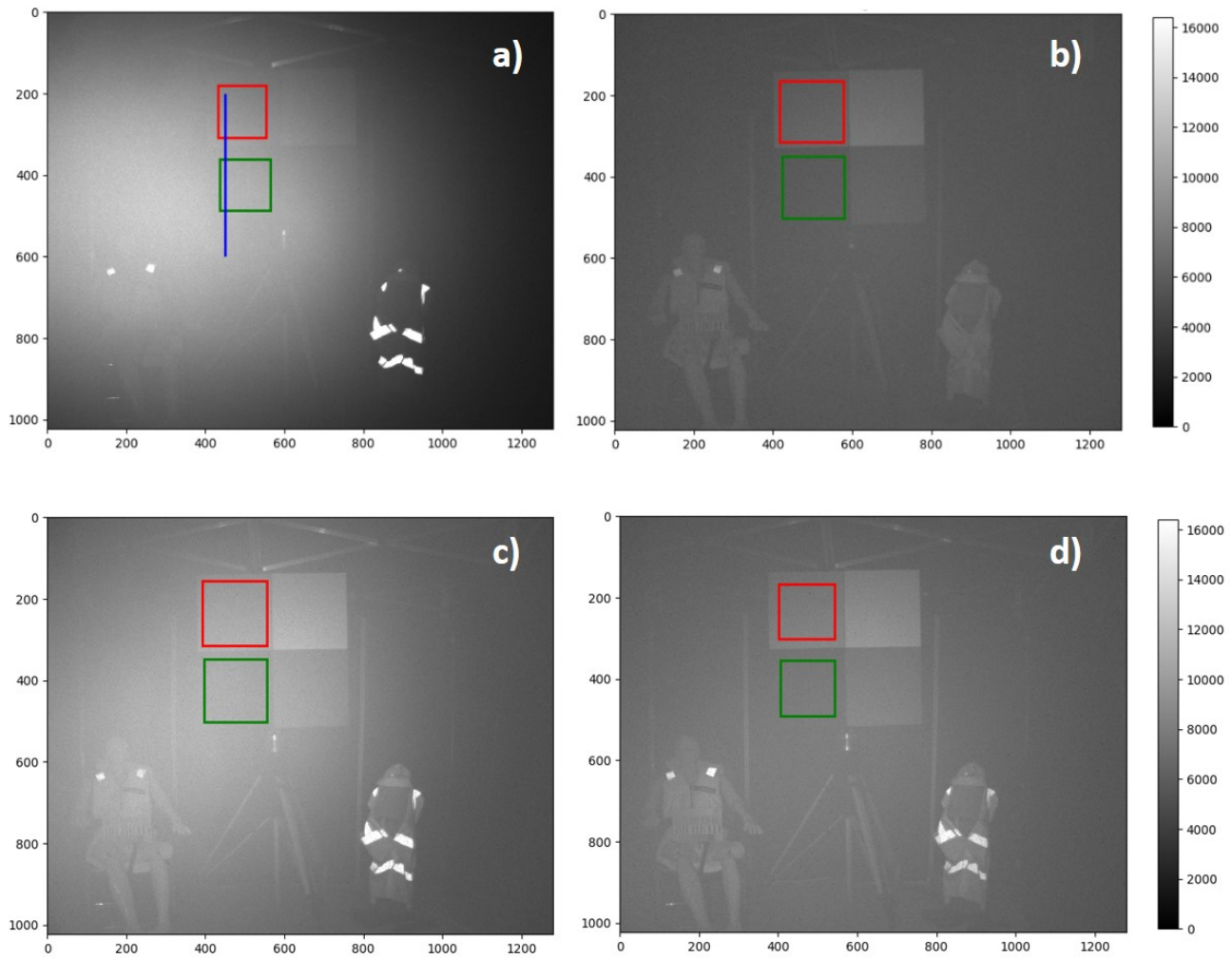
Figure 11. Example images acquired at similar fog conditions (extinction $\approx 2.4\ m^{-1}$) during **a)** measurement sequence 1, i.e. without polarization showing strong backscattering; **b)** sequence 4 using linear polarization; **c)** circular polarization using 1300 nm illumination in sequence 5; **d)** circular polarization using 1550 nm in sequence 6. While backscattering is completely removed using linear polarization (b), it is partially present in the circular polarization images c and d. This is caused by the broad spectral emission and the quarter-waveplate, which is designed for 1550 nm only (i.e. incomplete circular polarization is produced). All images were recorded with the SenS 1280 camera.

of polarized light (compare subimages a and b). This is in agreement with theoretical considerations in Sect. 3 predicting no depolarization for scattering in 180°. However, loss of photons by absorption and scattering out of the beam (i.e. extinction) is still present. A disadvantage is that not only the backscatter signal but also retroreflections and specular reflections are eliminated. Specular reflections give rise to contours and improve interpretation of the image in case of a human observer (this can be also easily seen from Fig. 6), and omitting retroreflections entirely is on the one hand successfully avoiding blooming effects (again see Fig. 6), but also results in a loss of (potentially important) details.

Sequences 5 and 6 (light green and magenta lines in Fig. 10) test the vision enhancement caused by use of circular polarization. In sequence 5, no difference in using left-handed (lh) or right-handed (rh) circular polarization could be observed. Furthermore, the performance of using circular polarized light in our field test was found to be better than using unpolarized light, but worse than linear polarization, which appears to be

in contrast to studies suggesting circular polarization to be preserved better in fog.[9–11] The reason is that the illumination unit produces circular polarization by a combination of linear polarization filter and quarter-waveplate for 1550 nm (i.e. having a retardance $\neq 0.25$ for $\lambda \neq 1550$ nm), i.e. *true* circular polarization is produced for 1550 nm only. The emission spectrum on the other hand is rather broad (see Fig. 2). Thus, blocking of the backscattered light by the polarization optics used is less efficient. This can be easily seen in Fig. 11: Especially Fig. 11c, which is emission of the 1300 nm unit and thus far away from the quarter-waveplate's 1550 nm, shows remaining backscattering at the left edge of the image. Fig. 11d, which is (broad) emission around 1550 nm, shows almost no backscattering because the emitted spectrum is much closer to the quarter-waveplate's 1550 nm. This is in agreement to the quantitative comparison of Michelson contrasts in Fig. 10 showing much better performance of sequence 6 (1550 nm with circular polarization) than sequence 5 (1300 nm with circular polarization). Sequence 6 performs almost as good as sequence 4, which blocks the backscatter signal entirely.

5. CONCLUSIONS AND OUTLOOK

An updated version of an optical and hand-held vision enhancement sensor working in the SWIR range was presented. Therefore, configurations based on different cameras and illumination units at 1300 nm and 1550 nm were compared during a field test using artificial fog.

Mie calculations show that for smaller water droplets and soot the extinction, scattering and backscattering in the SWIR are about one order of magnitude smaller than in the Vis (the exact value highly depends on the aerosol scenario), suggesting strong potential for vision enhancement in continental environments or indoor use. For the case of small water droplets, theoretical findings were confirmed by spectral measurements of a self-build transmissiometer during the field test. However, in maritime environments, SWIR will have only limited advantage in terms of vision enhancement as the dominant aerosols are mainly large water droplets.

Six measurement sequences were conducted to compare any combination of camera and illumination. During each sequence the extinction was increased and the Michelson contrast was obtained for a quantitative comparison. Within the limits of this study (in particular systematic differences due to inhomogeneous illumination) no configuration, i.e. combination of camera and illuminator, could clearly outperform the competing configurations. However, strong vision enhancement was achieved by additional combination of emitting and receiving components with polarization optics. Despite systematic differences from illumination, detectable (positive) contrasts in measurements without polarization were found until extinctions (at 550 nm) of 1 m^{-1} (resp. 3 m visibility) and for measurements with polarization until extinctions of 3 m^{-1} (resp. 1 m visibility). Besides vision enhancement, polarization was found to avoid blooming effects from retroreflectors.

Based on the results from the field test, our new instrument will be updated with polarization optics for 1300 nm as almost two times more intensity is produced compared to 1550 nm LEDs and use of polarization absorbs photons at the filters.

REFERENCES

[1] M. Mischung, J. Schmidt, E. Peters, M. W. Berger, M. Anders, and M. Stephan, "Development and characterisation of a portable, active short-wave infrared camera system for vision enhancement through smoke and fog," in *Electro-optical and Infrared Systems: Technology and Applications XIX*, Proc. SPIE, 2022.

[2] H. Ivanov, L. Mustafa, E. Cerncic, E. Leitgeb, and T. Plank, "Estimation of mie scattering influence for the fso channel under artificially simulated fog conditions," in *Environmental Effects on Light Propagation and Adaptive Systems*, Proc. SPIE, 2018.

[3] G. M. Hale and M. R. Querry, "Optical constants of water in the 200-nm to 200-microm wavelength region," *Applied optics* **12**(3), pp. 555–563, 1973.

[4] D. J. Segelstein, *The complex refractive index of water*. Master thesis, University of Missouri-Kansas City, 1981.

[5] M. R. Querry, *Optical Constants of Minerals and Other Materials from the Millimeter to the Ultraviolet*, CRDC-CR, Chemical Research, Development & Engineering Center, US Army Armament, Munitions, Chemical Command, 1987.

[6] W. H. Dalzell and A. F. Sarofim, "Optical constants of soot and their application to heat-flux calculations," *Journal of Heat Transfer* **91**(1), pp. 100–104, 1969.

[7] R. Bernier, X. Cao, G. Tremblay, and G. Roy, "Polarimetric active imaging in dense fog," in *Optics in Atmospheric Propagation and Adaptive Systems XVIII*, Proc. SPIE, 2015.

[8] J. Zenker, K. N. Collier, G. Xu, P. Yang, E. J. T. Levin, K. J. Suski, P. J. DeMott, and S. D. Brooks, "Using depolarization to quantify ice nucleating particle concentrations: a new method," *Atmospheric Measurement Techniques* **10**(12), pp. 4639–4657, 2017.

[9] J. D. van der Laan, D. A. Scrymgeour, S. A. Kemme, and E. L. Dereniak, "Detection range enhancement using circularly polarized light in scattering environments for infrared wavelengths," *Applied optics* **54**(9), pp. 2266–2274, 2015.

[10] J. D. van der Laan, J. B. Wright, S. A. Kemme, and D. A. Scrymgeour, "Superior signal persistence of circularly polarized light in polydisperse, real-world fog environments," *Applied optics* **57**(19), pp. 5464–5473, 2018.

[11] J. D. van der Laan, B. J. Redman, J. W. Segal, K. Westlake, J. B. Wright, and B. Z. Bentz, "Increased range and contrast in fog with circularly polarized imaging," *Applied optics* **62**(10), pp. 2577–2586, 2023.

[12] G. A. Atkinson and J. D. Ernst, "High-sensitivity analysis of polarization by surface reflection," *Machine Vision and Applications* **29**(7), pp. 1171–1189, 2018.

Optical choppers: novel approaches for higher chop frequencies

Eduard-Sebastian Csukas[1,2], Dorin Demian[1], Virgil-Florin Duma[1,2, *]

[1]3OM Optomechatronics Group, Faculty of Engineering, *Aurel Vlaicu* University of Arad,
77 Revolutiei Ave., 310130 Arad, Romania
[2]Doctoral School, Polytechnic University of Timisoara, 2 Victoriei Square, 300006 Timisoara,
Romania

ABSTRACT

We present an overview of our on-going work in the domain of optical choppers with rotational shafts, which we proposed and patented. The different constructive solutions are presented and discussed referring to the shape of the shafts and of the slits. Our Finite Element Analyses (FEA), developed as a multi-parameter approach considering all the material, constructive and functional characteristics of the devices is briefly presented – regarding both the structural integrity and the level of deformation of fast rotational shafts with multiple holes. The potential of this novel type of optomechanical choppers is pointed out, in comparison to classical solutions of choppers with rotational disks.

Keywords: optomechanics, optical choppers, choppers with shafts, Finite Element Analysis (FEA), modelling, rotational velocity, multi-parametric analysis.

1 INTRODUCTION

Choppers with rotational elements are one of the most utilized optomechanical modulators [1], together with optical scanners of different types [2-6]. Most common choppers are built with rotational disks [7-24], but we have introduced, patented [25], and developed [26-30] a novel type of choppers, with rotational shafts of different shapes (for example cylindrical, conical, or spherical), with slits of different profiles (i.e., with linear margins or as cylindrical holes). Such devices have the potential of surpassing the well-known 10 kHz limit of the chop frequency f_c of choppers with disks. For applications such as laser impulses generating, f_c is equal to the frequency of these impulses. Such macro-choppers with shafts are a simple and cost-effective alternative to Micro-Electro-Mechanical Systems (MEMS)-based choppers [31,32], which have higher costs (unless mass production is in place) and for which the technology may not be mature, as well. Also, macro-choppers are a more robust solution from a mechanical point of view. Many of their applications (for example for radiometers and pyrometers [7-9], telescopes [10,11], laser manufacturing [12,13], polarimetry [14], spectroscopy [15], light attenuation [24] (as an alternative to other devices [33]) and biomedical systems [16]) are in fixed laboratory-based systems, therefore their (higher) weight and dimensions are not major issues.

The aim of this presentation is to briefly point out some of our directions of research and contributions in the field, as well as advantages and drawbacks in comparison to other types of choppers.

2 CHOPPERS WITH SHAFTS VS. CHOPPERS WITH DISKS

Choppers with disks may reach their maximum f_c of 10 kHz for disks with 100 windows, which are currently commercially available [34]. However, in practice issues such as vibrations already appear around 3 kHz, therefore, MEMS with oscillatory elements may be a solution, while choppers with shafts can be a more robust and reliable one.

In addition, disks choppers with windows with linear margins (what we called "classical" choppers) have a limited possibility of generating different profiles of laser impulses, although we have demonstrated that, using certain relationships between its geometrical parameters and those of the laser beam, one can obtain a range of such profiles [14,15], even for the simpler case (from an analytical point of view) of top-hat incident laser beams.

---

*duma.virgil@osamember.org; phone: +40-751-511451; sites: https://www.researchgate.net/lab/3OM-Optomechatronics-Group-Virgil-Florin-Duma; http://3om-group-optomechatronics.ro/

Therefore, we have first proposed and patented [17] a configuration of chopper disk with non-linear margins, outward or inward oriented. We have demonstrated that such devices, in their simpler configuration, with semi-circular margins of the windows do allow for a more flexible design of the profile of the generated laser impulses [16,18], as a supplemental degree-of-freedom (DOF) is added to the analysis and design process, i.e. the radius ρ of the semi-circular margins. It is worth mentioning that the transmission functions deduced for such devices are actually the general case of classical choppers. Therefore, the transmission functions of the latter can be obtained as particular cases, when $\rho \rightarrow \infty$ (thus, the windows margins would become linear). We have named these novel devices "eclipse" choppers, because of the way the obscuration of the section of the beam in the plane of the disk occurs, similarly to a planetary eclipse [15-17], and somehow similar to the way vignetting is produced in optical systems [35].

3 CHOPPERS WITH SHAFTS

Several examples of designs of such choppers are presented in Fig. 1. Proof-of-concept devices were manufactured for the initial testing in the patenting process, while mechanically reliable variants have been further developed on our Bridge Grant (BG) project [36].
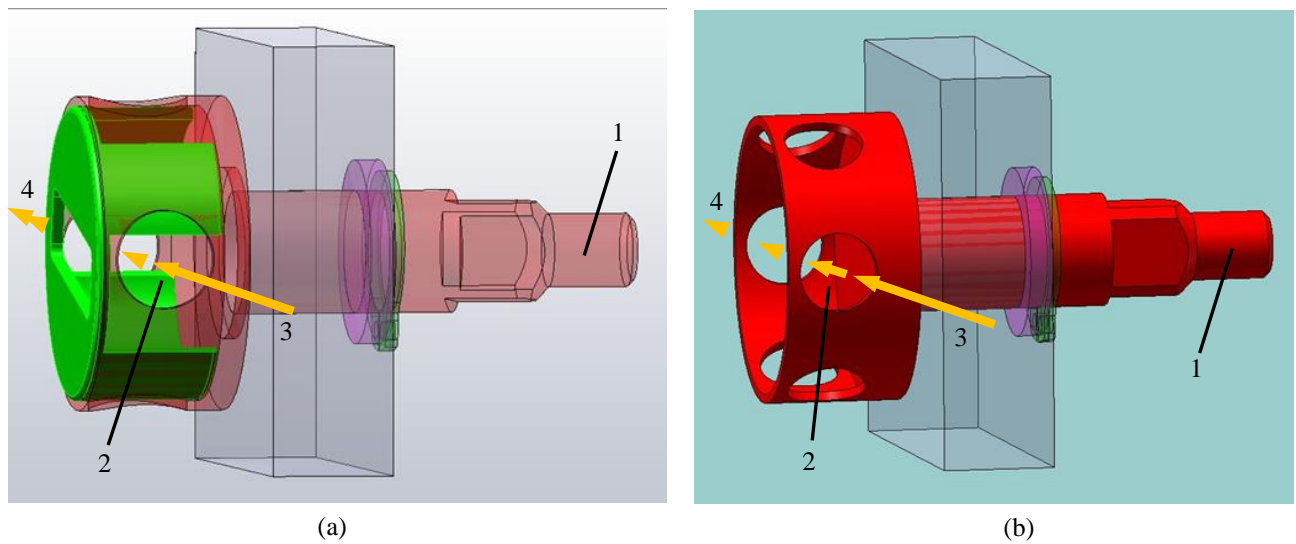

Figure 1. Optomechanical choppers with cylindrical rotational shafts: (a) with longitudinal slits; (b) with circular holes [25]. Notations: 1. shaft; 2. slit; 3. incident laser beam; 4. emergent laser beam/laser impulses.

A first analysis of the transmission function of the cylindrical shaft configuration of such choppers was performed in [26], for top-hat laser beams. We demonstrated that the chopping process of these devices is similar to that of choppers with disks (although with several peculiarities). The characteristic mathematical functions were deduced and analysed with regard to the parameters of the device.

However, the most challenging analysis of such choppers proved to be the mechanical one. Thus, the chop frequency f_c (therefore the frequency of the generated laser impulses) increases both with the rotational velocity ω of the shaft and with the number n of slits through the shaft, with the simple relationship

$$f_c = n \cdot \omega \,/\, 2\pi. \tag{1}$$

The necessity to increase both n and ω arises structural integrity and deformation issues. In this respect one must point out that ω can be similar for such devices with the high values reached for example for polygonal mirror (PM)-based scanners [5,31], which can have a common ω of 54 krpm, and up to 70 krpm. Appropriate sensors and control structures have been developed (in general by the industry) for such levels of ω.

The above issues must be addressed through a Finite Element Analysis (FEA) of each chopper configuration, similar to the efforts made in such respect for PMs [1-5]. Figure 2 presents an example of a FEA performed for one of the considered shapes of shafts for these choppers.

A: Static Structural
Equivalent Stress
Type: Equivalent (von-Mises) Stress
Unit: MPa
Time: 1

212.12 Max
156.06
100
85.72
71.441
57.161
42.881
28.602
14.322
0.042202 Min

(a)

A: Static Structural
Equivalent Plastic Strain
Type: Equivalent Plastic Strain
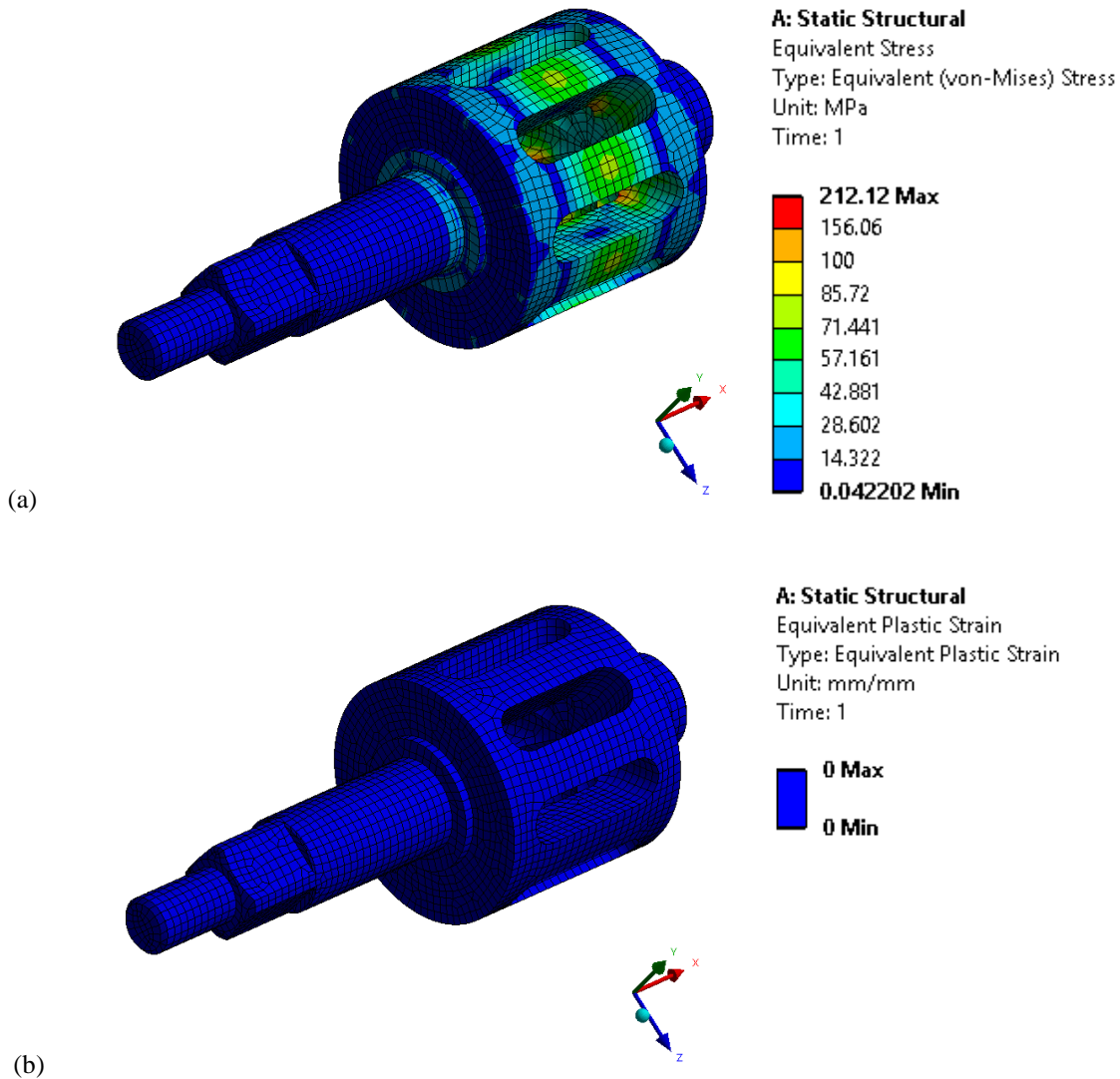Unit: mm/mm
Time: 1

0 Max

0 Min

(b)

Figure 2. Optomechanical choppers with cylindrical rotational shafts: (a) with longitudinal slits; (b) with circular holes [25]. Notations: 1. shaft; 2. slit; 3. incident laser beam; 4. emergent laser beam/laser impulses.

As presented in detail in [27-30], the performed analyses included deducing through ANSYS simulations the Von Misses stresses and the level of deformations for all three types of shafts, for both longitudinal (i.e., with linear margins) slits and for all constructive and functional parameters of these choppers: diameter D of the cylinder, dimensions of the slits, number n of slits (from 2 to 4 and 8 – in that first analysis), as well as rotational velocity ω of the shaft.

In all the performed FEA, 25 configurations of choppers were considered for each type of shaft, with diameters of the shafts of 20, 40, 55, 80, and 110 mm and with slits with widths considered with a step of 1 mm, from 1 to 5 mm.

Also, two different materials have been considered: (i) steel, with the density equal to 7850 kg/m^3, Poisson coefficient of 0.3, elastic modulus of 200000 MPa, and yield stress of 250 MPa; (ii) Al alloy, with the density equal to 2770 kg/m^3, Poisson coefficient of 0.33, elastic modulus of 71000 MPa, and yield stress of 280 MPa [27-30].

Stress vs. Shaft Velocity - Slits Configuration 2x2 mm

(a)

Stress vs. Velocity - Slits Configuration 3x3 mm
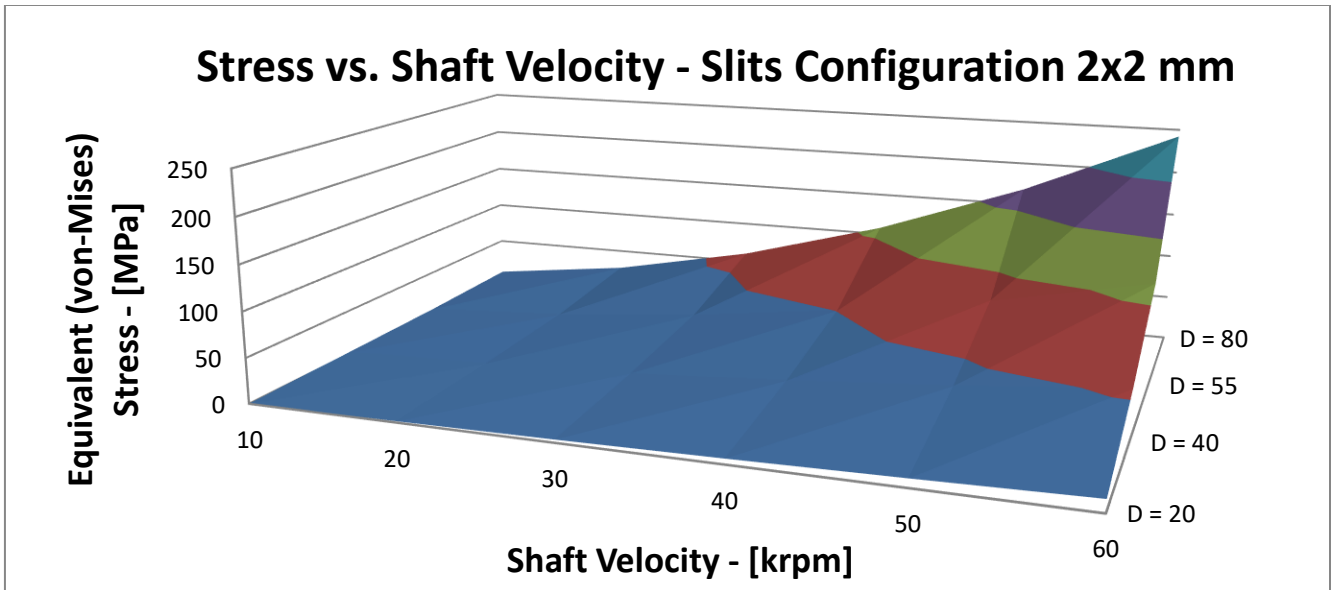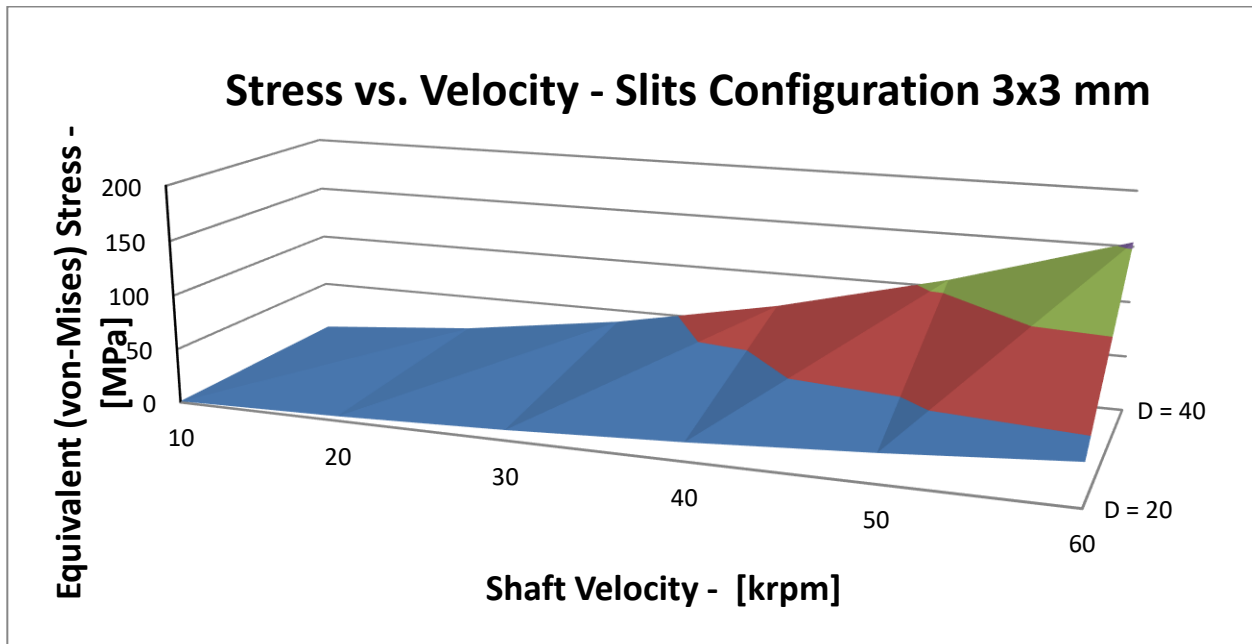
(b)

Figure 3. Result of the FEA for a chopper with a cylindrical shaft with two dimensions of longitudinal slits, (a) with [2 x 2 mm] and (b) with [3 x 3 mm], as well as with two different materials, (a) structural steel and (b) aluminium alloy: equivalent Von Misses stress [MPa] vs. shaft velocity ω [krpm] and diameter D.

The results, such as these presented as an example in Fig. 3, allow for determining the limits for a certain parameter when all other parameters (and the shaft material) are set or can be set within a certain interval. The imposed conditions are to mantain the mechanical stress lower than the threshold of the maximum admissible stress, as well as the deformations in the elastic (with no trespassing into the plastic) domain. Thus, the values of equivalent (von Mises) stresses must be lower than the yield stress of the considered material: 250 MPa for structural steel and 280 MPa for aluminum alloy.

One must highlight that, besides the equivalent Von Misses stress in the rotational shaft (as presented in Fig. 3), the considered configurations can reach plastic deformations as well. Such configurations are also considered failures and the design process must avoid them and reach configurations that do not reach the above limits.

The cross sections displacements that are produced by the high levels of rotational velocities have been investigated, as well, because these sections change shape, as the entire cylindrical shaft may tend to become barrel-like. This may impact the process of generating certain profiles of laser impulses, therefore it must be kept within certain limits.

4 CONCLUSIONS

The novel device, of choppers with shafts have been showing potential in reaching much higher chop frequency in comparison to the existing choppers with disks. Further studies are necessary to reach the limits of today's motors, i.e. 120 krpm in rotational velocity. Also, the trade-off between shafts diameter and number of slits must be studied considering the level of ω – and for specific materials. Applications are planned for such choppers, in radiometry and laser manufacturing, but also in biomedical imaging in dual systems that correlate optical coherence tomography and confocal microscopy in low coherence interferometry systems. Other applications of interest include to obscure the secondary mirrors of telescopes, as well as to generate laser impulses of controlled frequencies and profiles.

ACKNOWLEDGEMENTS

REFERENCES

[1] Bass, M., Ed., [Handbook of optics], Mc. Graw-Hill Inc., New York, (2010).
[2] Marshall, G. F., [Handbook of Optical and Laser Scanning], CRC Press – Taylor and Francis, New York (2011).
[3] Montagu, J., Scanners - galvanometric and resonant, in: C. Hoffman, R. Driggers (Eds.) Encyclopedia of Optical and Photonic Engineering, 2nd ed. CRC Press (2015).
[4] Duma, V.-F., "Laser scanners with oscillatory elements: Design and optimization of 1D and 2D scanning functions," Appl. Math. Modelling 67, 456-476 (2019).
[5] Duma, V.-F., Duma, M.-A., "Optomechanical Analysis and Design of Polygon Mirror-Based Laser Scanners," Appl. Sci. 12, 5592 (2022).
[6] Duma, V.-F., Dimb, A-L., "Exact Scan Patterns of Rotational Risley Prisms Obtained with a Graphical Method: Multi-Parameter Analysis and Design," Appl. Sci. 11(18), 8451 (2021).
[7] Leonov, V.N., Butler, D.P., "Two-color Thermal Detector with Thermal Chopping for Infrared Focal-Plane Arrays," Appl. Opt. 40, 2601-2610 (2001).
[8] He, Y., Jin, W., Liu, G., Gao, Z., Wang, X., Wang, L., "Modulate chopper technique used in pyroelectric uncooled focal plane thermal imager," Proc. SPIE 4919, 283-288 (2002).
[9] Giraldo, D., Correa, H., Peña Lara, D., "Implementation of a programmable electromechanical chopper with adjustable frequency and duty cycle for specific heat measurements," Measurement 110, 60-64 (2017).
[10] Bittner, H., et al., "Optical system of the SOFIA Telescope," Proc. SPIE 3356, 512-521 (1998).
[11] McDermid, I.S., Beyerle, G., Haner, D.A., Leblanc, T., "Redesign and improved performance of the tropospheric ozonelidar at the jet propulsion laboratory table mountain facility," Appl. Opt. 41, 7550-7555 (2002).
[12] Wan, D.-P., Liu, H.-B., Wang, Y.-M., Hu, D.-J., Gui, Z.-X., "CO_2 laser beam modulating for surface texturing machining," Opt. & Laser Tech. 40, 309–314 (2008).
[13] Swapna, M.S., et al, "Ultralow duty cycle chopper instigated low power continuous wave laser assisted synthesis of silver nanoparticles: A novel approach," *J. Laser Appl.* 32, 042017 (2020).
[14] Matchko, R.M., Gerhart, G.R., "High-speed imaging chopper polarimetry," Opt. Eng. 47(1) 016001 (2008).
[15] Song, S., Li, G., Hou, X., Zhang, S., Yu, Y., Lin, L., "Principal frequency component analysis based on modulate chopper technique used in diffuse reflectance spectroscopy measurement," Appl. Opt. 57, 1043-1049 (2018).

[16] Podoleanu, A., Dobre, G.M., Cucu, R.G., "Sequential optical coherence tomography and confocal imaging," Opt. Letters 29, 364-366 (2004).

[17] Udommai, P., Harvey, M., Murray, A.J., "A low jitter, low cost optical chopper," Measurement Science and Technology 31(12), 125903 (2020).

[18] Lekkas, I., Frogley, M.D., Achtnich, T., Cinque, G., "Rapidly frequency-tuneable, in-vacuum, and magnetic levitation chopper for fast modulation of infrared light," Review of Scientific Instruments 93, 085105 (2022).

[19] Duma, V.-F., "Theoretical approach on optical choppers for top-hat light beam distributions," J. of Opt. A: Pure and Appl. Opt. 10, 064008 (2008).

[20] Duma, V.-F., "Prototypes and modulation functions of classical and novel configurations of optical chopper wheels," Latin American Journal of Solids and Structures 10(1), 5-18 (2013).

[21] Duma, V.-F., "Optical choppers with circular-shaped windows: Modulation functions," Communications in Nonlinear Science and Numerical Simulation 16, 2218-2224 (2011).

[22] Duma, V.-F., et al, "Optical modulator with rotation movement," Patent RO 126505, (2016).

[23] Cira, O., Duma, V.-F., "Transmission functions of optical choppers for Gaussian beam distributions: Modeling and simulations," Proc. SPIE 8789, 87890E (2013).

[24] Duma, V.-F. "Optical attenuators with Risley prisms and disk choppers," Proc. of the Romanian Academy Series A 20(2), 141-150 (2019).

[25] Duma, V.-F., Demian, D., "Optical modulator, has solid rotating shaft with some through slots of well-defined profiles," Patent RO129610-A0, (2014).

[26] Duma, V.-F., et al, "Optical choppers with high speed rotating elements," Proc. SPIE 9131, 9131OE (2014).

[27] Duma, V.-F., Demian, D., Csukas, E.-S., Pop, N., Cira, O., "Non-conventional optomechanical choppers: analysis and design of novel prototypes," Proc. SPIE 10448, 104481W (2017).

[28] Csukas, E.-S., Duma, V.-F., "Optical choppers with rotational shafts for laser applications: Finite Element Analysis," Proc. SPIE 10785 (2018).

[29] Csukas, E.-S., et al, "Optomechanical Choppers with Conical Shafts," Proc. SPIE 11151, 1115122 (2019).

[30] Csukas, E.-S., Duma, V.-F., "Optical choppers with spherical shafts: An optomechanical analysis," Proc. SPIE 11028, 1102805 (2019).

[31] Ueda, M., Shiono, T., Ito, T., Yokoyama, K., "High-efficiency diffractive micromachined chopper for infrared wavelength and its application to a pyroelectric infrared sensor," Appl. Opt. 37, 1165-1170 (1998).

[32] Syms, R.R.A., et al, "Sliding-blade MEMS iris and variable optical attenuator," J. Micromech. Microeng. 14, 1700 (2004).

[33] Duma, V.-F., Nicolov, M., "Neutral density filters with Risley prisms: analysis and design." Appl. Opt. 48, 2678-2685 (2009).

[34] Optical Chopper System and Chopper Wheels (thorlabs.com), accessed on 10 August 2023.

[35] Duma, V.-F., "Radiometric versus geometric, linear and non-linear vignetting coefficient," Appl. Opt. 48(32), 6355-6364 (2009).

[36] Bridge Grant BG297 (2016-2018) - 3OM Optomechatronics Group, accessed on 10 August 2023.

[37] Oh, W.Y., Yun, S.H., Tearney, G.J., Bouma, B.E., "115 kHz tuning repetition rate ultrahigh-speed wavelength-swept semiconductor laser," Opt. Lett. 30, 3159-3161 (2005).