Pola4All: survey of polarimetric applications and an open-source toolkit to analyze polarization

Joaquin Rodriguez[a,*] Lew-Fock-Chong Lew-Yan-Voon[a]
Renato Martins[b,c] and Olivier Morel[a]
aUniversité de Bourgogne, ImViA Laboratory, Le Creusot, France
bUniversité de Bourgogne, ICB UMR CNRS 6303, Dijon, France
cUniversité de Lorraine, LORIA, Inria, CNRS, Nancy, France

ABSTRACT. Polarization information of the light can provide rich cues for computer vision and scene understanding tasks, such as the type of material, pose, and shape of the objects. With the advent of new and cheap polarimetric sensors, this imaging modality is becoming accessible to a wider public for solving problems, such as pose estimation, 3D reconstruction, underwater navigation, and depth estimation. However, we observe several limitations regarding the usage of this sensorial modality, as well as a lack of standards and publicly available tools to analyze polarization images. Furthermore, although polarization camera manufacturers usually provide acquisition tools to interface with their cameras, they rarely include processing algorithms that make use of the polarization information. In this work, we review recent advances in applications that involve polarization imaging, including a comprehensive survey of recent advances on polarization for vision and robotics perception tasks. We also introduce a complete software toolkit that provides common standards to communicate with and process information from most of the existing micro-grid polarization cameras on the market. The toolkit also implements several image processing algorithms for this modality, and it is publicly available on GitHub.

1 Introduction

The polarization of the light is present in several real-world physical phenomena. Light coming from rainbows in the sky, reflections from water on highways, and monitors and cellphones based on liquid-crystal display (LCD) screens are typical examples. Light polarization is naturally generated when an unpolarized light source (e.g., light bulbs or the sun) hits a surface and is reflected. The polarized light generated in this way can be of two types: specular, when the reflection of the light is in a single direction; or diffuse, when the reflection is in all directions. The way the reflected wave oscillates depends on the characteristics and the shape of the material. This relationship between the observed light and the object properties is a key feature that vision algorithms can use to improve their accuracy with respect to another one that uses only texture information. These additional features can be leveraged, for instance, to improve object detection and scene segmentation results, detect mirrors and other surfaces that polarize the light, or uniquely identify places in a room that will serve as landmarks in navigation

---

*Address all correspondence to Joaquin Rodriguez, joaquin-jorge.rodriguez@u-bourgogne.fr

algorithms. It is worth noting that polarization cues are the main sources of information used by many biological agents such as insects and bees for their orientation in space.[1]

The introduction of micro-grid polarization sensors, such as Sony Polarsens, boosted the research in the polarization domain since they are capable of capturing the intensity, the color and the linear polarization information in a single snapshot, and they also allow measurements outside laboratory conditions. However, the number of approaches leveraging polarization for performing computer vision and robotics tasks is, unfortunately, still quite limited. For these reasons, to promote the usage of polarization cameras in robotics and computer vision tasks, we provide in this paper a comprehensive review of the latest advances in the field of polarization imaging. The revised papers have been chosen to show the potential of this modality to improve the results given by RGB-only methods, especially in challenging situations, e.g., in the presence of transparent and textureless objects. Furthermore, to push even more the research in this field, we break the barrier of the practical problems by introducing a complete acquisition and processing software toolkit. This toolkit comes with a graphical user interface (GUI) and is capable of processing images coming from commonly available RGB-polarization sensors, such as the Sony Polarsens, regardless the camera manufacturer. The aim of the software is to provide standard acquisition and processing tools for researchers and practitioners in the field to facilitate the visualization and analysis of polarization images. In summary, the main contributions of this paper are:

- A comprehensive review of the most recent advances in the polarization imaging field for computer vision tasks. The selected works include the most representative application cases in which the polarization information contributes outstandingly to an improvement of the results compared to those that would have been obtained with an RGB-only camera.
- We provide a complete acquisition and processing toolkit, including a GUI to fill the gap of a lack of a common software for RGB-polarization cameras. The objective of this tool is twofold. On one hand, it is a toolkit that any researcher can use as a base framework, to analyze and understand the images obtained by any RGB-polarization camera. On the other hand, due to its careful design, it enables easy development and inclusion of additional applications. We include in this paper some different use cases of the developed software with the available features to highlight the potential usages.

In what follows, we start with a brief introduction to the theory of the polarization state of light. We then present the main concepts used by all the reviewed papers, allowing the reader to better understand the contribution of polarization for each approach. Finally, we describe the developed software toolkit made publicly available to the community.

2 Polarization Background

2.1 Mathematical Polarization Model

Light is an electromagnetic wave of high frequency, and when it propagates through space, it can be defined by its amplitude, its frequency, and the way it moves as it travels. The intensity of the wave is equivalent to the brightness of the light, the frequency is equivalent to its color, and the way it moves transversely to the propagation direction defines its polarization state. Let us consider the projection of the oscillation of the electric field vector of the light wave on a plane perpendicular to the propagation direction. The light is said to be linearly polarized if this projection gives a line. It is said to be circularly polarized if the projection describes a circle. If the vector moves in all directions in a random manner, the light is said to be unpolarized.[2] Furthermore, a combination of linearly and circularly polarized light gives an elliptically polarized wave, and a light that has both an unpolarized and a polarized component is a partially polarized light. This last case is the most common type of polarized light that can be found in the nature.

There exist several models to depict mathematically the polarization state, but the most commonly adopted one is the Stokes model.[3] This model defines the light wave as a 4D vector $\mathbf{s} = [S_0, S_1, S_2, S_3]^T$. S_0 represents the total light intensity (polarized and unpolarized). S_1 and S_2 describe the amount of light that is linearly polarized horizontally/vertically, and in the

direction of ± 45 deg, respectively. S_3 represents the amount of light that is circularly polarized. Using this model, it is possible to define two important physical polarization parameters

$$\rho = \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0} \quad \text{and} \quad \phi = \frac{1}{2} \arctan\left(\frac{S_2}{S_1}\right), \tag{1}$$

where ρ is called the degree of polarization (DoP), which represents the portion of the light that is polarized, and ϕ is the angle of polarization (AoP). This angle represents the orientation of the line segment, or of the ellipse, when the light is respectively linearly or elliptically polarized. Circularly polarized light is rare in nature,[4] thus considering only the first three components of the Stokes vector is a good approximation to model the polarization state. Therefore, in most applications, the S_3 component is set to zero, and the Stokes model is represented by a 3D vector of these physical variables as

$$\mathbf{s} = \begin{pmatrix} S_0 \\ S_0\rho \cos(2\phi) \\ S_0\rho \sin(2\phi) \end{pmatrix}. \tag{2}$$

If the linear components of the Stokes vector are used, we refer to the AoP and the DoP as the angle of linear polarization (AoLP), and the degree of linear polarization (DoLP), respectively.

2.2 Polarization Measurement

An advantage of the Stokes model is that the effect produced by an object (either by transmission, by reflection, or by scattering) on the incident wave can be modeled by a Mueller matrix $\mathbf{M}$.[3] More specifically, a Stokes vector $\mathbf{s}_{\text{in}}$ that interacts with an object whose Mueller matrix is $\mathbf{M}$ is converted into a Stokes vector $\mathbf{s}_{\text{out}}$ according to the following equation:

$$\mathbf{s}_{\text{out}} = \mathbf{M}\mathbf{s}_{\text{in}}. \tag{3}$$

If the full-Stokes vector is used, the matrix $\mathbf{M}$ has a shape of 4×4 elements. However, if only the linear part of the Stokes vector is considered, then this matrix has 3×3 components. Until now, the only way to measure the Stokes vector components is by using an indirect measurement method. For the linear components of the light, this process consists in measuring the received intensity when the light passes through a linear polarization filter (LPF) at different orientations. An LPF is an optical device that allows only the waves that have the same orientation as the filter axis to pass through. Any other wave is filtered with a gain that has a sine curve shape. The maximum gain occurs when the filter orientation matches the AoLP of the incident light, and the minimum gain will occur when these angles are separated by $\pi/2$ radians. An optical device as the one described is modeled by the following Mueller matrix:[5]

$$\mathbf{M} = \frac{1}{2} \begin{bmatrix} q+r & (q-r)C_{2\theta} & (q-r)S_{2\theta} \\ (q-r)C_{2\theta} & m_{11} & m_{12} \\ (q-r)S_{2\theta} & m_{21} & m_{22} \end{bmatrix}, \tag{4}$$

where q and r are the major and minor light transmittance of the linear polarizer, respectively, θ is the orientation of the filter, and

$$S_{2\theta} = \sin(2\theta), \ C_{2\theta} = \cos(2\theta),$$
$$m_{11} = (q+r)C_{2\theta}^2 + 2\sqrt{qr}S_{2\theta}^2,$$
$$m_{22} = (q+r)S_{2\theta}^2 + 2\sqrt{qr}C_{2\theta}^2,$$
$$m_{21} = m_{12} = \left(q+r - 2\sqrt{qr}\right)S_{2\theta}C_{2\theta}.$$

If a camera is used to take the measurements of the filtered light, then, only the first component of $\mathbf{s}_{\text{out}}$ can be retrieved, which corresponds to the total intensity of the observed light S_0^{out}. Thus, only the first line of the Mueller matrix $\mathbf{M}$ should be considered

$$S_{0_\theta}^{\text{out}} = \frac{1}{2}[q+r \quad (q-r)C_{2\theta} \quad (q-r)S_{2\theta}]\mathbf{s}_{\text{in}}, \tag{5}$$

where $S_{0_\theta}^{out}$ is the S_0 component of the output Stokes vector, when the filter axis is oriented at an angle of θ radians. If the filter is considered ideal, then $q = 1$ and $r = 0$, and Eq. (5) becomes

$$S_{0_\theta}^{out} = \frac{1}{2}[1 \quad \cos(2\theta) \quad \sin(2\theta)]\mathbf{s_{in}}. \tag{6}$$

In general, $S_{0_\theta}^{out} = I_\theta - d$, where I_θ is the readout intensity given by a pixel, and d is the pixel offset, often called dark current. Most works ignore d since in commercial cameras, this value is negligible compared to the camera measurement.[6] Thus, we have in general that

$$I_\theta = \frac{1}{2}[1 \quad \cos(2\theta) \quad \sin(2\theta)]\mathbf{s_{in}}. \tag{7}$$

To find the vector $\mathbf{s_{in}}$, several measurements at different angles θ are required. This can be done by using a division of focal plane (DoFP) polarization sensor composed of super-pixels. A super-pixel is a matrix of 2×2 pixels on top of which are four linear polarization micro-filters, one on each pixel and oriented at angles of 0 deg, 45 deg, 90 deg, and 135 deg. Moreover, to capture the color information, the super-pixels are organized according to the Bayer pattern where each of its constitutive pixels shares the same color filter, as shown in Fig. 1(a). This method has the advantage of capturing all the required information in a single shot: color and polarization. The choice of the four polarizer orientation values for the super-pixel is demonstrated by Tyo[8], in which it concludes that using equidistant angles in the range [0 deg, 180 deg] optimizes the signal-to-noise ratio (SNR) of the computed Stokes vector.

For a DoFP sensor, four different orientations are used $(\theta_1, \theta_2, \theta_3, \theta_4) = (0 \text{ deg}, 45 \text{ deg}, 90 \text{ deg}, 135 \text{ deg})$, thus four different intensity measurements are obtained $(I_{\theta_1}, I_{\theta_2}, I_{\theta_3}, I_{\theta_4}) = (I_{0 \text{ deg}}, I_{45 \text{ deg}}, I_{90 \text{ deg}}, I_{135 \text{ deg}})$. Then, four expressions of Eq. (7) are obtained, which when stacked together gives a linear system in the form $\mathbf{i} = \mathbf{As_{in}}$ with

$$\begin{bmatrix} I_{\theta_1} \\ I_{\theta_2} \\ I_{\theta_3} \\ I_{\theta_4} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & \cos(2\theta_1) & \sin(2\theta_1) \\ 1 & \cos(2\theta_2) & \sin(2\theta_2) \\ 1 & \cos(2\theta_3) & \sin(2\theta_3) \\ 1 & \cos(2\theta_4) & \sin(2\theta_4) \end{bmatrix} \mathbf{s_{in}}, \tag{8}$$

where $\mathbf{i}$ is the intensity measurements vector, $\mathbf{A}$ is called the pixel matrix, and $\mathbf{s_{in}}$ is the Stokes vector we want to estimate. Then, it is possible to find the Stokes vector $\mathbf{s_{in}}$ by computing the pseudo-inverse of $\mathbf{A}$, which results in the following analytical form:
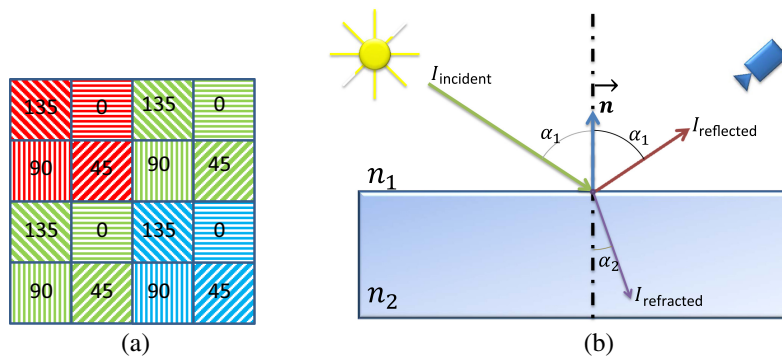

(a)   (b)

Fig. 1 (a) Micro-grid sensor scheme with four super-pixels with polarizer orientations of 0 deg, 45 deg, 90 deg, and 135 deg arranged following a Bayer pattern of a typical RGB-polarization sensor.[7] (b) Sketch of the interaction between a light ray coming from the air with refractive index n_1 and an object with refractive index n_2. When the incident light ray hits a point on the surface of the object, a portion of the light is reflected and a portion of it is refracted or transmitted in the second medium. The law of reflection states that the angle of reflection is equal to the angle of incidence α_1. Regarding the refracted light, its direction of travel will change, and it will be equal to α_2.

$$\mathbf{s}_{\text{in}} = \begin{pmatrix} \frac{I_{0\ \text{deg}} + I_{45\ \text{deg}} + I_{90\ \text{deg}} + I_{135\ \text{deg}}}{2} \\ I_{0\ \text{deg}} - I_{90\ \text{deg}} \\ I_{45\ \text{deg}} - I_{135\ \text{deg}} \end{pmatrix}. \tag{9}$$

2.3 Naturally Generated Polarization

An important property of the polarization state is that it conveys information about the shape and the composition of the objects when they are made of dielectric materials. After hitting a surface, an incident wave will create two new waves:[3] a reflected and a refracted wave. In general, when a camera observes an object, the captured light is the result of the reflection.

Furthermore, for insulator materials that produce specular light, there is a single angle in which the reflected light will be 100% linearly polarized.[3] Any other direction will produce partially polarized light. The angle at which this type of reflection occurs depends on the index of refraction, which is a value related to the material type and the wavelength of the incident light. The interaction of the light, the material, and the observed intensity is depicted in Fig. 1(b). In this image, n_1 and n_2 are the indexes of refraction of the top and bottom medium respectively, α_1 the angle of the reflected light with respect to the normal vector $\mathbf{n}$ to the surface, and α_2 the angle of the refracted light ray with respect to the opposite to the normal vector. Another important property regarding the reflected light is that the polarization state is related to the surface orientation. It is for this reason that, through Fresnel theory,[9] if the ratio $\lambda = n_2/n_1$ is known, by measuring the AoP and the DoP, it is possible to retrieve the normal to the surface at each point of the object.

In summary, the polarization state of the light can provide rich geometric cues about the shape, pose, and material of objects, notably in challenging conditions to classic imaging approaches, such as in the presence of highly reflective objects' surfaces (mirrors, windows, water,...) or transparent/translucent objects. We let the reader refer to these Refs. [9–12].

3 Deep-Learning Background

In this section, we introduce the basic concepts of deep learning algorithms, with the aim of facilitating the understanding of the subsequent sections of the paper.

3.1 Overview of Deep-Learning Networks

In the fields of computer vision and robotics, data-driven algorithms are the most performant ones, more particularly, deep-learning algorithms. These methods are based on a network model developed to solve a task through a training process where the network would learn to interpret the input data. Once the network is trained, it can be used on new and unseen data to perform the intended task.

The training process is an optimization routine in which a set of input data, known as the training data, is passed through a large equation with coefficients that can be tuned and the corresponding outputs determined. Next, an error function, called the loss function is computed based on the input data, the output data, and the constraints with which the output must comply. Finally, the equation coefficients, or network weights, are updated to try to reduce the loss function value. This process of updating the weights is called the back-propagation algorithm.[13] Once the weights have been updated, the process is repeated for a certain number of iterations, or epochs.

The theory of deep neural networks has existed for several years, but it is only after Krizhevsky et al.[14] that these algorithms have become popular. The authors have shown that convolutional neural networks (CNNs) are capable of outperforming hand-crafted theoretical algorithms. A CNN is composed of several layers, and each of them includes several convolutional, optimization, and normalization operations. At the output of each layer, we obtain the different images called feature maps. The deeper the layer in the network, the higher the level of information it represents. In other words, the output of the layers close to the network input represents low-level features (edges, lines, and circles), and the output of the layers close to the output of the network represents high-level features such as object labels (chair, building, car,

dog, etc), depth of the objects in the scene with respect to the camera coordinate frame, a text description of what the scene represents, or the pose of an object with respect to the camera frame. In a CNN, a large number of convolutions are done to the input image, and the training process is in charge of finding the convolution kernels coefficients to convert the input image into the expected output image.

3.2 Popular Developed Models

In the last decade, several deep neural network algorithms have been developed. They significantly improved the performance of their predecessors in various aspects, such as achieving higher accuracies, maintaining the same precision with fewer parameters, and faster training times. Examples of CNNs are the VGGNet,[15] MobileNet,[16] EfficientNet,[17] DeepLabV3,[18] and ResNet.[19] Even though there is a vast variety of models, in general, the models can be divided into two main blocks: an encoder block and a decoder block. The encoder is composed of layers designed to extract valuable information from the input data (also called features). The features are then passed to the decoder for interpretation and determination of the output data. Many authors use a known encoder block and tailor a specialized decoder module to meet their specific needs. Additionally, when dealing with color data, the encoder blocks of the most popular networks are already included in the deep learning frameworks, and they have already been pre-trained with the largest datasets available. This way, the optimization routine will converge faster than training the same model with a tiny dataset.

Very recently, a new type of neural network architecture has been developed that challenges the existing CNNs. It is known as transformers, and it has been presented by Vaswani et al.[20] This new type of network does not use the convolution operation, and it has been shown to outperform all the known networks in the field of natural language processing. Furthermore, Dosovitskiy et al.[21] showed that transformers can also provide more accurate results in the computer vision field. The network is based on what is called the attention mechanism, and more precisely, on the self-attention mechanism. This method computes the dot product between the different entries, and the output is an indication of the importance of the data at a given position for the task. Therefore, the attention mechanism can be used to ignore the areas where no important data is present and to keep the regions where the data is valuable to accomplish the final task. The disadvantage of this type of architecture is that it requires a large amount of data to outperform CNN. If the dataset used for training is not large enough, the performance of transformers is poor.

3.3 Polarization and Multi-Modal Networks

In data-driven approaches, it is not common to use polarization data exclusively, either because they are very noisy due to the acquisition process, or because there are not enough constraints based purely on the polarization theory for the network to converge. Therefore, polarization data are usually mixed with other modalities in a multi-modal network. This type of network takes several sources of data (at least color and polarization), and try to mix them efficiently. The process of mixing data in this way is called data fusion in the literature, and in general, one of the following three configurations is used: early, middle, or late fusion.

In the early fusion architecture, the data are mixed before entering the encoder network for feature extraction. This fusion mechanism is performed in general through operations that contain learnable parameters as the network, and these parameters are adjusted during the training process. In late fusion, the data from each modality passes through different encoders. The weights for each modality can be shared between the encoders, but it is not a common practice since each modality extracts different types of patterns from the inputs. Once the input modalities have passed through the encoders, the high-level feature maps are fused into a single one that is then passed to the decoder module. Finally, in the middle fusion architecture, as for the late fusion model, the data from each modality is inputted to an independent encoder, but in this case, the feature maps in the encoder at the different hierarchical levels are fused. In other words, at each network layer of the encoder, the corresponding feature maps are fused. Middle fusion is the heaviest fusion architecture in terms of parameters and forward pass time. Nonetheless, it ensures a complete mixture of the data at all levels in the network, avoiding information loss while the data passes from one layer to another.

In what follows, we will review the different approaches that make use of the polarization information in different ways, using distinct approaches such as optimization algorithms, hand-crafted features, and data-driven approaches.

4 Related Work

In this section, we discuss recent advances in the field of polarization imaging within both model-based and data-driven strategies. We have reviewed the latest applications in the computer vision and robotics field that utilize polarimetry between the years 2016 and 2022. Older applications have not been considered because the most significant advancements began after the release of the first micro-grid DoFP sensor around 2014. Since then, DoFP sensors have been adopted and widely used as it allows real-time imaging. In 2018, Sony introduced a DoFP sensor, named PolarSens, which became the core device of many commercial polarization cameras. It offers much better quality data and facilitates the development of more performant algorithms and real-time applications beyond laboratory conditions. The works have been grouped into four representative categories and complementary tasks: image enhancement, segmentation, surface depth and normal estimation, and pose estimation. Table 1 shows a summary of the papers included in each group. Most reviewed works focus on the application fields of computer vision and robotic vision. Particularly, we focus our study on the context of scene understanding in both ground and underwater environments. Polarization has also been extensively used in the field of remote sensing, notably in combination with synthetic aperture radar data. Applications in remote sensing are not the focus of this survey and a good review of the recent advances of applications of polarization in this field using data-driven algorithms can be found in Ref. 56.

4.1 Image Enhancement

In real-world applications, changes in viewing conditions can strongly impact the performance of computer vision algorithms. Thus, enhancing the quality of the visual information is often a required step to keep the accuracy of the developed computer vision system. The type of image quality improvement depends on the application itself. In some cases, this implies having high-quality measurements independently of the camera used, which can be achieved through camera calibration.[5–7,57] In others, the improvement can be related to removing highly bright, specular reflections, requiring a separation between this type of reflection from the diffuse ones.[22,58,59] In more complex cases, the background structure needs to be recovered when an atmospheric phenomenon is present such as mist or fog, as shown for some examples in Fig. 2. In most cases, the

Table 1 Summary of the reviewed works (2016 to 2022) of this polarization survey.

| Survey section	Application	Papers
Image enhancement	White balance and demosaicing	Refs. 22–25
	Dehazing, reflection removal, and image restoration	Refs. 26–34
	Specular-diffuse separation	Ref. 35
Image segmentation	Material-based segmentation	Ref. 36
	Glass segmentation	Ref. 37 and 38
	Urban scenes segmentation	Refs. 39–41
Surface normal and depth estimation	Normal map estimation	Refs. 42–46
	Depth estimation	Refs. 10, 11, 47, and 48
	3D reconstruction/scene rendering	Refs. 49–51
Pose estimation	Object's pose	Refs. 12, 52, and 53
	Optical flow	Ref. 54
	Underwater solar-tracking	Ref. 55
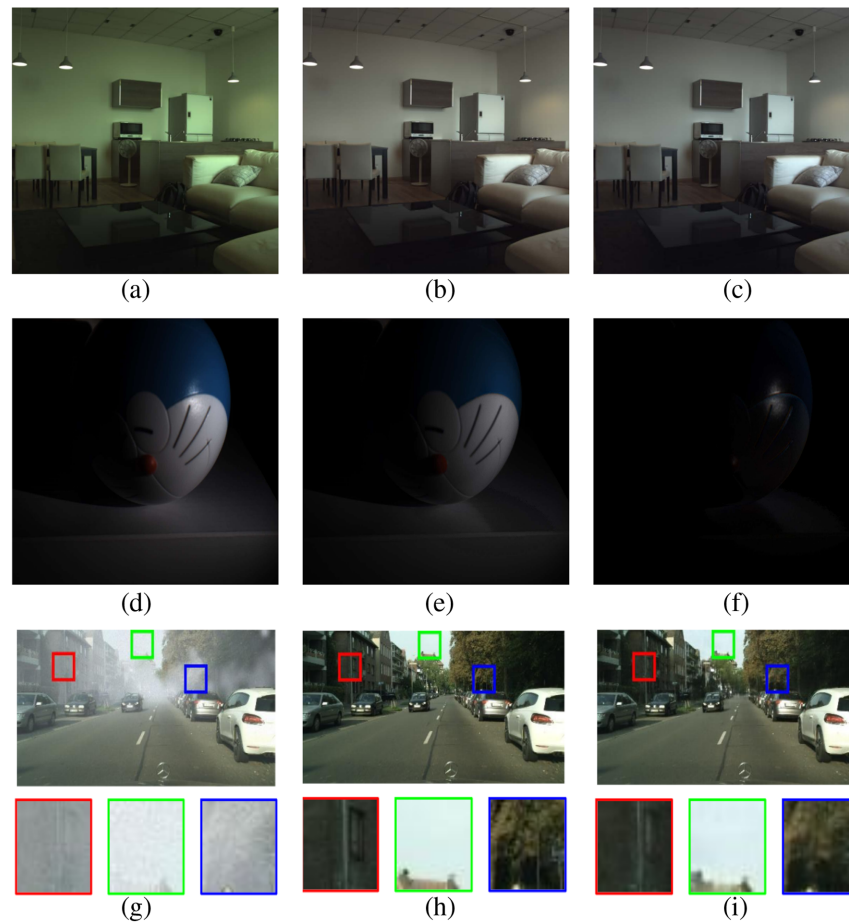
Fig. 2 Results of some of the reviewed algorithms for image enhancement. (a)–(c) Input image, ground truth, and algorithm results of DoLP-based color constancy.[22] (d)–(f) Input image, obtained diffuse image, and obtained specular image of polarization-guided specular reflection separation.[35] (g)–(i) Input image, ground truth, and resulting images of the polarization dehazing method.[27] Images courtesy of the respective works.

physical constraints defined by the polarization state of the light can be used to improve the results obtained by conventional cameras.

In this context, Ono et al.[22] presented a white balance algorithm for RGB-polarization sensors based on the achromaticity of the Stokes vector in the visible spectrum. Rodriguez et al.[7] relaxed the experiment setup to calibrate micro-grid color-polarization sensors to achieve a flat-field response in all the polarization parameter images. Wen et al.[35] solved the separation of specular from diffuse reflections with a model-based optimization strategy.[60] Their pipeline is independent of the illumination source by exploiting polarization and chromaticity images. Wen et al.[24] proposed to jointly demosaic RGB and polarization information to obtain high-quality, 12-channeled RGB-polarization images using a sparse representation model. The model is obtained through an optimization algorithm that uses the ADMM scheme. Similarly, Morimatsu et al.[23] obtained high-quality polarization images by extrapolating the residual interpolation for RGB images[61] to the monochrome and color polarization sensors. They achieve their results by changing the guidance image so that it is edge-aware, and by making use of the raw polarization intensity measurements. Tanaka et al.[32] achieved better quality images by improving the condition number of the transport matrix in comparison with conventional, passive non-line-of-sight system. This is done using the polarization leakage effect model produced by the oblique reflection over a filter oriented at the Brewster angle[9] of a wall.

Using hand-crafted theories provides good results when the scene and the effect to analyze are not complex since a high-precision mathematical model of the problem can be established. When this is not possible, data-driven algorithms can be used, as they have the capability to

learn complex theories during training. For example, they can handle scenes with several objects at the same time, or model effects for which no known mathematical model exists. Data-driven approaches as described in Lei et al.[29] have been designed to remove the reflections produced by different types of glasses using polarization theory. The input to their network architecture, composed of pre-trained U-net and VGG-19 networks, is an image that is a combination of the raw measurements, split by polarization channel and polarization parameters (I, ρ, ϕ), as presented in Eq. (1). Zhou et al.[27] used a single polarization image and a deep learning network composed of two U-net models and two autoencoders to dehaze urban scenes. Hu et al.[26] developed a data-driven approach and a dataset to increase the brightness and quality of images under low-illumination conditions. They created a CNN that works in two steps based on the raw measurements of the camera: first, an enhancement in the intensity domain for all the color channels, and then each color channel is treated by a separate network. Liu et al.[28] proposed a generative adversarial network architecture to fuse the DoLP and the intensity images into a single intensity image. By dividing the image into background and foreground, the network fuses these two polarization images into a single image that has an increased and better contrast than the original intensity image. The results produced by this network can be used to train other networks, i.e., to perform an improved data augmentation, in order to obtain models with increased generalization capabilities than the ones obtained when trained only with the original intensity images. Despite the outstanding results of data-driven algorithms with respect to the optimization-based approaches, the quality of these results depends on the data and the type of model used. Particularly for the data, if not all the cases have been considered in the images provided during training, the missing cases might produce in less accurate results during testing.

Several relevant image enhancing approaches have also been proposed to deal with the challenging scenario of underwater imaging. Li et al.[30] aimed to improve the contrast of underwater images due to turbidity using polarization and an optimization strategy. They propose to split the Stokes vector into three contributions (diffuse, specular, and scattered light) since they claim that the scattering reflection underwater cannot be neglected. In the same direction, Hu et al.[31] present a novel CNN based on residual blocks that fuse the polarization features to restore the contrast of underwater images. Using the raw measurements of three polarization channels of a monochrome polarization camera, they are able to see through turbid water and obtain a clear image of the hidden objects. Amer et al.[33] proposed a static pipeline to increase the image quality for underwater applications. Based on the active cross-polarization technique and an optimized version of the dark channel prior, they achieve contrast improvement for underwater imaging, with a single snapshot in real-time. Shen and Zhao[34] developed an iterative pipeline to jointly improve the image contrast and denoise the image. With two polarization images taken with a rotative filter at 0 deg and 90 deg, they compute the transmittance and the irradiance maps in underwater conditions for each color channel. Then, they establish an iterative process to refine these results using an adaptive bilateral filter and an adaptive color correction routine.

Despite the remarkable improvement brought by polarization to image enhancement, as compared to similar applications for RGB-only cameras, several challenges still remain. For example, Ono et al.[22] outperformed different baselines in many scenes, but it is left as future work to improve the results obtained when the sky occupies a large portion of the image. Similarly, Zhou et al.[27] retrieved the hidden structure of objects behind the haze with good accuracy in real-world situations, after training the network with computer generated images. Despite this, the authors claim that the model does not produce adequate reconstructions for fog and mist due to the physical phenomena produced by these perturbations that are not the same as for the haze. It is important to note that one of the barriers in the polarization image enhancement field is the lack of standard benchmarks. Indeed, most works had to create a dataset to demonstrate their contributions. Some of the created datasets, which often required a huge amount of work, can be reused as in Lei et al.,[29] where the authors realized acquisitions in a large variety of environments, and with different types of glasses. On the other hand, other applications have been demonstrated using small non-available datasets, which may not necessarily cover all the required cases,[24] or they have created a polarization dataset based on RGB ones and a mathematical model of the polarization effect.[27]

4.2 Image Segmentation

The polarization state of the light is directly linked to the object's material and shape as introduced in Sec. 2.3. This property can provide insightful and complementary information to guide object segmentation approaches in scenarios where only the surface color is not discriminant. Indeed, the index of refraction depends on the internal structure of the objects, and on the wavelength of the incident light as defined by the Fresnel equations.[2] It is for this reason that material classification is one of the most fruitful application of polarization theory. Previously, this task was accomplished using hand-crafted features, in controlled acquisition conditions, using rotative filters, and considering a single object to be analyzed at a time.[62–65] Nowadays, with the advances in sensors and the advent of data-driven algorithms, object characterization with polarization cues has been ported to more complex, constraint-relaxed scenarios.

In the domain of infra-red imaging, Li et al.[39] succeeded in efficiently detecting the road area in urban scenes using the zero-distribution prior in the AoLP and the difference in the DoLP of the objects to increase the accuracy of the segmentation. This information is further used in a visual tracking algorithm to continuously track the road online. In an extension of their previous work, Li et al.[40] used the zero-distribution prior of the AoLP to create a coarse map of the road. Then, they developed a deep-learning network to refine the coarse road map. Their network consists of two branches that analyze different aspects of the scene. The main branch receives the information captured by an infra-red camera, already converted into a fake color image, i.e., a three-channel image result of stacking the AoLP, the DoLP, and the total intensity together, and then converted from the Hue-Saturation-Value (HSV) to the RGB color space. The objective of this branch is to extract multi-modal features of the scene. The other branch or polarization-guided branch also receives the AoLP and the DoLP of the scene, but it does not receive the intensity image. Instead, the coarse map obtained from the zero-distribution prior of the AoLP is provided. By doing so, the authors aim to guide the network based on the polarization properties of the road and not of the entire scene. In the visible spectrum, Xiang et al.[41] developed a fusion network to combine color and polarization data to better segment objects of urban scenes. They tested several combinations of polarization information with attention mechanisms and concluded that using only color and the AoLP is the best combination to improve the results. Kalra et al.[38] improved the instance-semantic segmentation network Mask R-CNN[66] to handle transparent objects (as shown in Fig. 3) by adding monochrome polarization cues to the original mid-fusion pipeline. Each polarization parameter image (intensity, AoLP, and DoLP) is fed into a different backbone encoder, and the fusion of the feature maps is performed at each encoder level. Mei et al.[37] extended the work presented in Ref. 38 using RGB polarization cues, instead of monochrome, with the aim of segmenting glasses in urban scenes. Since they use an RGB-Polarization camera, they can measure RGB intensity, RGB DoLP, and RGB AoLP. Each of the two RGB polarization images (DoLP and AoLP) is balanced using an attention mechanism. Then, these two results, and the intensity RGB image are fed into three independent conformer encoders[67] and fused using local and global guidance. In a more complex scenario, Liang et al.[36] built a network to fuse RGB, infra-red, and polarization cues to produce an outdoor scene segmentation based on the object material type. The proposed pipeline is composed of two core elements: a network that will classify the objects present in one class of a subset of the segmentation labels from the CityScapes dataset[68] and a region-based filter selection module that chooses the modality that provides the most relevant information for determining the type of material of the constitutive elements of each detected object. The full network is composed of four encoders: one for the RGB intensity, one for the AoLP image, one for the DoLP image, and one for the infra-red image.

All these works outperform RGB systems when the polarization information is added to each developed pipeline. A higher gain in performance is also often obtained when the network is adapted to correctly process the AoLP and the DoLP, and not when an RGB network is trained with polarization images. This is why most of these works propose carefully designed fusion schemes. However, the lack of datasets including polarization information in the field of image segmentation poses limitations on the development of polarization-based approaches. It is important to highlight that all the previously discussed works have presented their own dataset to show that polarization is a path to consider in image segmentation. For instance, Kalra et al.[38] used a private dataset acquired in a very specific environment, focused on the particular application with a pick-and-place robotic arm. Xiang et al.[41] provided a small-scale dataset of RGB-polarization
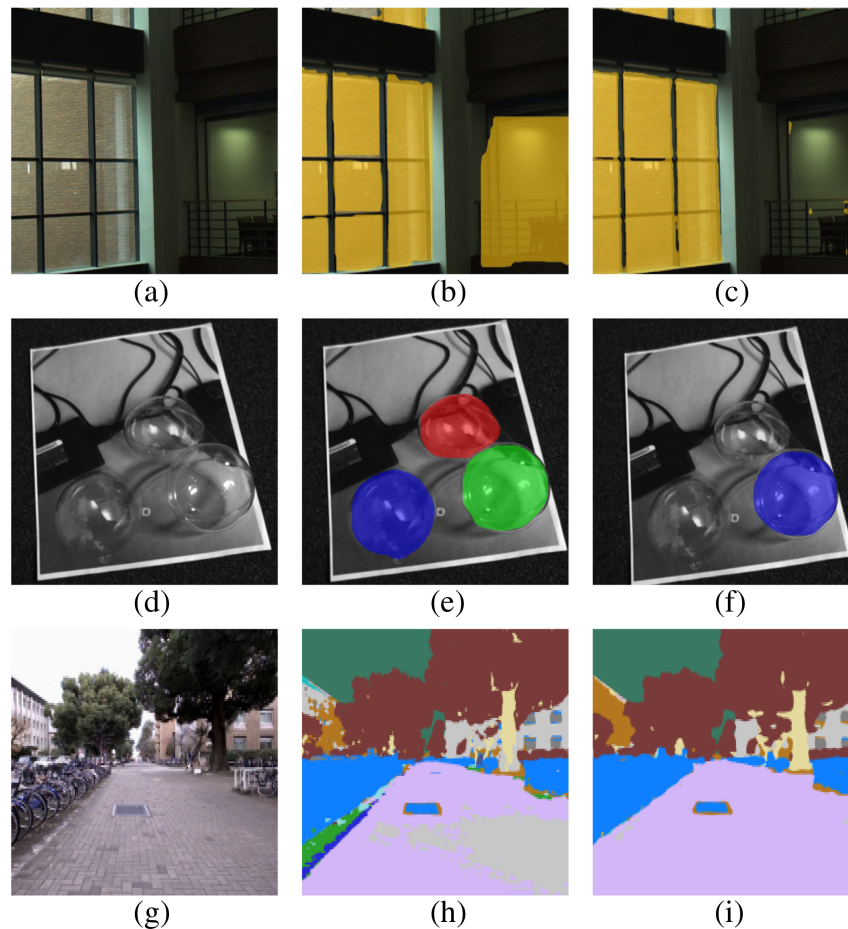
Fig. 3 Segmentation examples with polarization cues. From left to right: (a)–(c) Input image, results obtained with an RGB-only method, and the result of glass segmentation using intensity and spectral polarization information.[37] (d)–(f) Input image with transparent objects, RGB-only method object segmentation results, and results obtained with the polarization data-driven method.[38] (g)–(i) Input image, RGB-only method results, and the results of the multimodal material segmentation algorithm.[36] Images courtesy of the respective works.

images captured in various urban scenes. Although informative, the dataset contains only 394 annotated images segmented into 9 different classes. Mei et al.[37] introduced a medium-scale dataset, with 4511 images annotated only for the labels glass and no-glass. Similarly, Liang et al.[36] made publicly available a dataset of semantic segmentation of urban scenes, with multi-modal sensors, but it only includes 500 labeled images, and Li et al.[39] did it for road segmentation, and with their personalized infra-red polarization camera. Thus, there is a need for a common large-scale benchmark to evaluate the performance of these different segmentation algorithms to trace the direction toward a generalization of the polarization modality.

4.3 Surface Normal and Depth Estimation

Polarization is well known to encode shape information of the different objects being observed. Most existing approaches consider an orthographic projection of the incoming light, and if the refractive index of the object is known, then the normal vectors to the surface can be estimated, and through integration, the depth map can be retrieved. These approaches are again based on the Fresnel's theory, which has been briefly introduced in Sec. 2.3, which establishes the relationship between the degree of linear polarization with the zenith angle of the normal, and the AoLP with the azimuth angle. Even though effective in several works and scenarios,[69–72] the fact that both the object's refraction index and the light direction must be known makes this approach limited to laboratory strict conditions. Additionally, the relationship between polarization and the normal vector has geometric ambiguities. Therefore, one important research direction is to reduce the constraints and priors required for the acquisition while maintaining low reconstruction errors.

Ba et al.[45] proposed a learning-based approach to estimate the normal map of objects as shown in Fig. 4. The ambiguous normal maps from Fresnel's theory are used as priors given directly to a deep neural network as inputs. Fukao et al.[44] present a shape from polarization algorithm that uses a stereo pair of polarization cameras. The coarse map from the stereo vision is refined by filtering the normal maps with a belief propagation scheme. They exploit Fresnel's theory, and an improved modeling of the micro-facet reflection effect by considering that it is a linear combination of the diffuse and the specular lobe reflection. Similarly, Ichikawa et al.[42] relaxed the constraints for shape from polarization by using the Rayleigh and the Perez models to estimate the sun polarization state and direction on a clear day. Then, through mathematical optimization, the normal and shading maps are obtained. The additional cues about the incident Stokes vector serve to determine how the object modulates the incident light (Mueller calculus) and jointly with the shading constraints, the normal map can be estimated. Deschaintre et al.[46] proposed a 3D object shape estimation, jointly with a spatially variable bidirectional reflectance distribution function (BRDF) model estimation, by using a single-view polarization image fed to a U-Net based network architecture. The full input of the model is the intensity image, the normalized Stokes map, and the normalized diffuse color, which encodes the object reflectance information. Lei et al.[43] proposed a deep-learning network to estimate the normal map of complex scenes. Their aim is to improve the accuracy limits by incorporating viewing encoding as input to the network, which accounts for the non-orthographic projection. This input is an image where each pixel represents the direction of the incident light. When estimating the normal
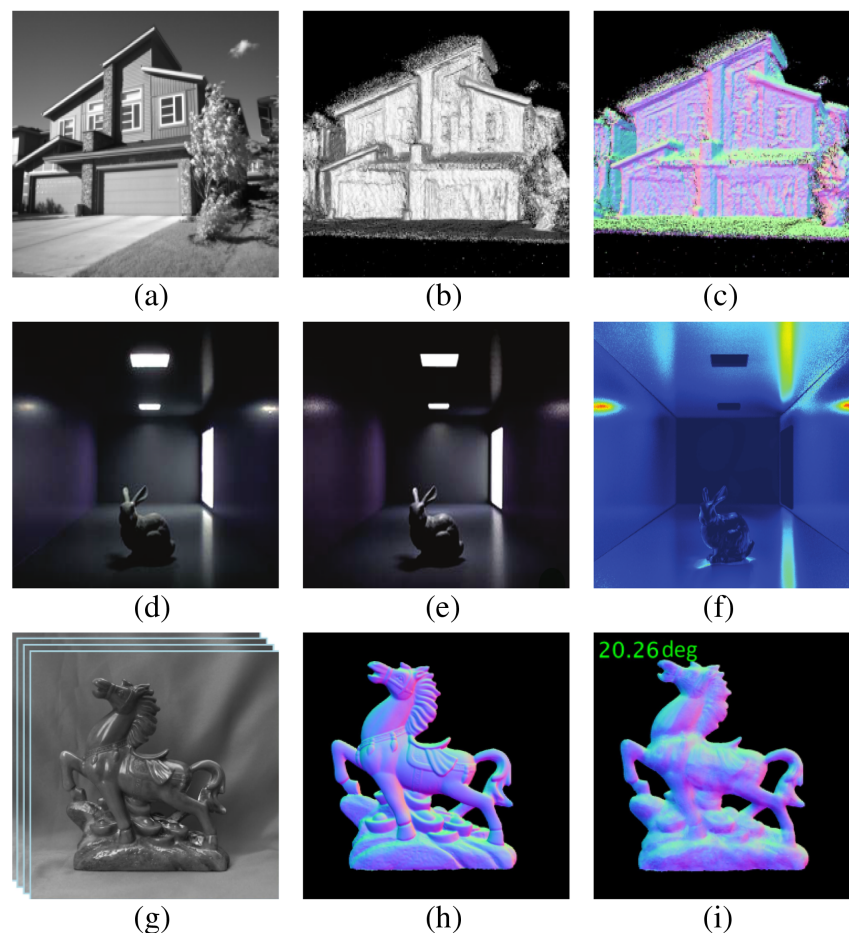

Fig. 4 Results of some of the reviewed algorithms for surface normal and depth estimation. From left to right: (a)–(c) intensity input image, reconstructed mesh, and estimated normal map by the method introduced by Ref. 49. (d)–(f) Real RGB image, corresponding rendered image, and the corresponding rendered degree of linear polarization. This result corresponds to the scene rendering technique implemented in Ref. 51. (g)–(i) Input polarization images, ground truth, and estimated normal maps. These results are courtesy of Ref. 45.

vectors from polarization under the orthographic assumption, all the incident light rays are supposed to be colinear to the Z axis of the camera coordinate frame. Thus, all the zenith angles are measured with respect to a common coordinate frame. When using a perspective lens, the zenith angle given by the polarization theory is measured with respect to the direction of propagation of the light, which in this case, will be different for each pixel. By providing the viewing encoding, the authors claim the network will understand the viewing direction of the polarization state and use this information to improve the results of a network that works under the orthographic assumption with a perspective lens. The other inputs to the model are the raw measurements of the camera separated by polarization channel, the AoLP, the DoLP, and the total intensity. Their network is also grounded by an architecture similar to a U-Net model, with a multi-head self-attention module in the bottleneck. Smith et al.[10] defined the shape from polarization problem as a large linear system of equations. They combine the physics theory of polarization with the geometry of the problem to formulate the depth equations directly, without passing through the computation of the normals. Berger et al.[47] present a depth estimation algorithm that uses the polarization cues in a stereo vision system. They improve the correspondence matching by adding the AoLP-normal constraint to the intensity similarity function. In the same direction, Zhu and Smith[11] propose a hybrid RGB-polarization acquisition system to obtain a dense depth reconstruction. By classifying the pixels into specular or diffuse, they make use of normal vectors obtained from Fresnel's theory to improve the estimation of the normal maps obtained from the stereo images. Blanchon et al.[48] extend the monocular depth estimation network Monodepthv2[73] to consider polarization information by adding the azimuthal constraint to the deep-learning loss. Zhao et al.[50] extended the multi-view reconstruction system[74] by adding polarization cues to the optimization. They introduce a continuous function that has four minimum values, each of them at one of the ambiguous normal azimuth possibilities of Fresnel's theory. Kondo et al.[51] developed a polarimetric BRDF model that does not constrain the illumination nor the camera position during acquisition. This model is used to synthesize polarization images out of the RGB images, easing the dataset creation for data-driven algorithms. By acquiring images with different illumination of known Stokes vectors, they use Mueller calculus to model the object reflectance. Shakeri et al.[49] produced a dense 3D reconstruction using polarization cues. This is done by optimizing an initial depth map obtained from MiDaS,[75] and the coarse depth map from COLMAP.[76] The optimization routine constraints the normals with the ones from Fresnel's theory. The initial depth map is used to disambiguate the polarization normals, and the coarse map is used to regularize the optimization routine since they are metrically correct but sparse.

The previously presented works have all been developed for shape/depth estimation while leveraging the polarization information. Combining the polarization state of the light with any geometry problem developed for the RGB space results in a significant improvement in accuracy and image quality. This is because the polarization measurements are dense since they are provided pixel-wise, which means that the normal constraints are also dense. Thus, passive, high-quality far-field 3D reconstructions can be retrieved using a multi-modal RGB-polarization camera, which cannot be done with active sensors, such as LiDAR or Microsoft Kinect. However, these polarization constraints are still often dependent on knowing priors about the material (metallic versus insulator) and the reflection type (specular versus diffuse), thus sometimes they can produce low-quality results in the wild. To overcome this problem, some works decide to use only diffuse reflections[46] or to classify the pixels into either diffuse dominant or specular dominant (such as in Refs. 10 and 49). To deal with more complex cases, a better modeling of the reflection effect might be required.[44] For data-driven algorithms, this field also suffers from the lack of large-scale datasets that can be used as benchmarks for research. Most papers propose their own dataset by doing acquisitions,[43] or they model the polarization state assuming artificial conditions to simulate polarization over already existing RGB images.[51]

In summary, the polarization clearly provides valuable cues in the field of shape estimation and 3D reconstruction but two main problems need to be addressed. Foremost, the lack of large-scale, standard evaluation benchmarks hinders the development of techniques using this modality in the current era of data-driven algorithms. On the other hand, there is no generic model that can effectively handle generic types of reflections over any type of material. Therefore, the challenge of interpreting the measured data and determining which model to use remains open.

4.4 Pose Estimation

The polarization of light can also play a significant role in object pose estimation since it provides valuable geometric constraints in the determination of the vectors normal to the observed surfaces (as discussed in Sec. 4.3). Hence, this additional information can be used to overcome the ill-posedness condition of many RGB problems, such as estimating the relative rotation and translation of textureless objects between two images. Particularly, these additional constraints are useful when the objects to analyze are highly reflective and translucent since the polarization measurements are independent of the intensity of the light. Cui et al.[12] used the normal vectors estimated from the Fresnel equations to add geometric constraints for pose estimation (some visualizations of the pose estimation can be seen in Fig. 5). With this additional information, only two corresponding points in two views are required to estimate the rotation matrix and the translation vector. In the same direction, Gao et al.[52] proposed a data-driven algorithm to find the pose transformation of an object in the image with respect to the camera coordinate frame. The algorithm uses three ambiguous normals as inputs to one encoder, and the polarization parameters into another. Then, the features are fused at different levels, and given to the decoder. Tzabari and Schechner[54] present a static approach in which they use the AoLP and the DoLP to expand the optical flow theory. This new component accounts for the rotation speed estimation, which cannot be done with the classical optical-flow approach. Hu et al.[55] utilized a monochrome polarization camera to build a complete pipeline to estimate the sun's position based on the DoLP and the AoLP measured underwater. Jointly using the Snell and Fresnel theories, they revert the ray bending caused by the change in medium and handle the problem as if the measurements were done outside the water. Zou et al.[53] pushed forward the accuracy in the human shape and pose estimation by building a two steps network with polarization cues. Assuming the human cloth to be diffuse dominant, they retrieve the human features by using the raw polarization intensity images, and the ambiguous normal maps obtained from Fresnel theory. The first network produces a high-quality normal map, and the second one uses this result jointly with the output of the SMPL human shape model[77] to estimate the final shape and pose of a clothed person.

Due to the geometric nature of the pose estimation problem, the polarization state of the light provides valuable clues that can be used in any computer vision algorithm in this field. The applications included in this section of the review demonstrate the potential of accuracy gain obtained with the polarization constraint while at the same time relaxing the other algorithms
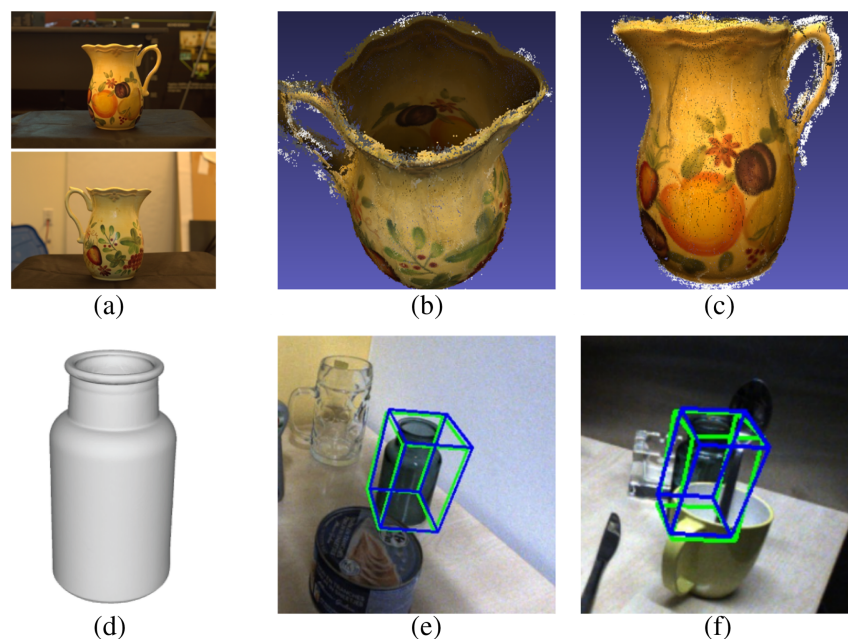

(a)　　　　　(b)　　　　　(c)

(d)　　　　　(e)　　　　　(f)

Fig. 5 Results of some of the reviewed algorithms for pose estimation. From left to right: (a)–(c) input RGB image, reconstructed surface with the algorithm provided in Ref. 12. (d)–(f) Glass vase model and the detected pose from two viewpoints. Results courtesy of Ref. 52.

hypothesis. In Ref. 12, it is shown that only two points are required to estimate the pose transformation between two views, resulting in an improvement in speed and accuracy when added to any structure from motion algorithm. Without any requirement in the type of clothes to be used,[53] they were able to estimate the human pose with a lower error and fewer constraints than competitors. In underwater applications, global positioning system (GPS) signals cannot be used because their intensities decrease rapidly with the depth in the water. To address this issue, Hu et al.[55] propose an autonomous underwater navigation system that uses polarization instead of a GPS signal. In their system, the camera's global position is estimated by applying geometrical constraints that link the sun's position to its known trajectory.[78,79] However, several limitations still remain when doing pose estimation with polarization. For example, in Ref. 52, only the position of one object can be done each time, whereas others adopt known object materials and physical properties. Furthermore, most algorithms only consider one type of reflection (either diffuse or specular), which limit their generalization to any type of scene.

5 Pola4all: Polarization Toolkit

From the presented study on the advances in polarization imaging, we have observed that in all the reviewed applications there are no common standards either for data structure, acquisition, or visualization tools. We can also notice a lack of a common library to boost the development, evaluation, and deployment of algorithms using polarization. For these reasons, this section presents a software library that allows to visualize and analyze polarization images, as well as to develop and integrate any custom application in a structured manner. In what follows, we briefly describe the software, including its basic components and the implemented image-processing algorithms. The objective is to contribute to the community by providing several tools integrated with a single common GUI software that can interact with any DoFP RGB-polarization camera available on the market. Additionally, this software is meant to provide access to all the developed algorithms that showcase the power of polarization information, making it easier for anyone interested in working with polarization modality to get started. The library has been developed in C++ object oriented programming language and is built on top of three main libraries/frameworks: OpenCV, Robot Operating System (ROS) and QT5. An overview of the final GUI can be found in the GitHub repository of this project.

Throughout this section, we will show the output images of our program. To be able to compare the polarization properties through the proposed software library toolkit, the same test image of an urban scene is used in all the experiments. This image, corresponding to the raw image obtained from the camera, is shown in Fig. 6. The library is composed of several modules designed to be independent so as to enable easy maintenance and debugging. With such a structure the addition or removal of a functionality is straightforward. Regarding the architecture of

Fig. 6 Raw image of an urban scene used as a reference for all the algorithms implemented in the toolkit Pola4All.

the code, it is composed of two general components: the camera server, and the GUI client. They are detailed in the documentation of the library repository.

5.1 Core Components

The first component is an ROS camera server package that interacts directly with the camera: getting images and changing (or querying) its parameters, such as the pixel gain, exposure time, and frame rate. It has been designed to support the integration of existing modern micro-grid polarization cameras. The second component of the software is the GUI, which works as a client of the ROS server. The final user interface, with a detailed explanation, is included in the GitHub repository.[80] This interface allows the user to perform all the required tasks involving the camera such as changing the image parameters, and the super-pixels filter configuration. It also allows image processing, raw image display, and sensor calibration. To analyze the calibration performance, plotting functions are provided.

5.2 Basic Processing

The basic polarimetric processing techniques included in the toolkit are detailed below.

Raw split images: it produces as output four images, one per filter orientation independently of the color. An example of the image for filter orientation of 0 deg and red channel is shown in Fig. 7(a).

Polarized color images: it produces as output four images, which are the demosaiced version of the output from the raw split images mode obtained using the bilinear interpolation algorithm. Since the raw measurements of the camera are used, they are not white-balanced. Thus, the resulting images exhibit a greenish aspect, as shown in Fig. 7(b) for the 0 deg polarization-red color channel. To this image, the white-balance algorithm described in Sec. 5.2.1 can be applied to obtain the resulting white-balanced image, as shown in Fig. 7(c). Polarized color images allow to identify regions in the scene where light is polarized. To illustrate this, the demosaiced, white-balanced image for the channel 135 deg is shown in Fig. 7(d). Note that the windshields of the cars appear dark in this image compared to the one corresponding to filter orientation of 0 deg.
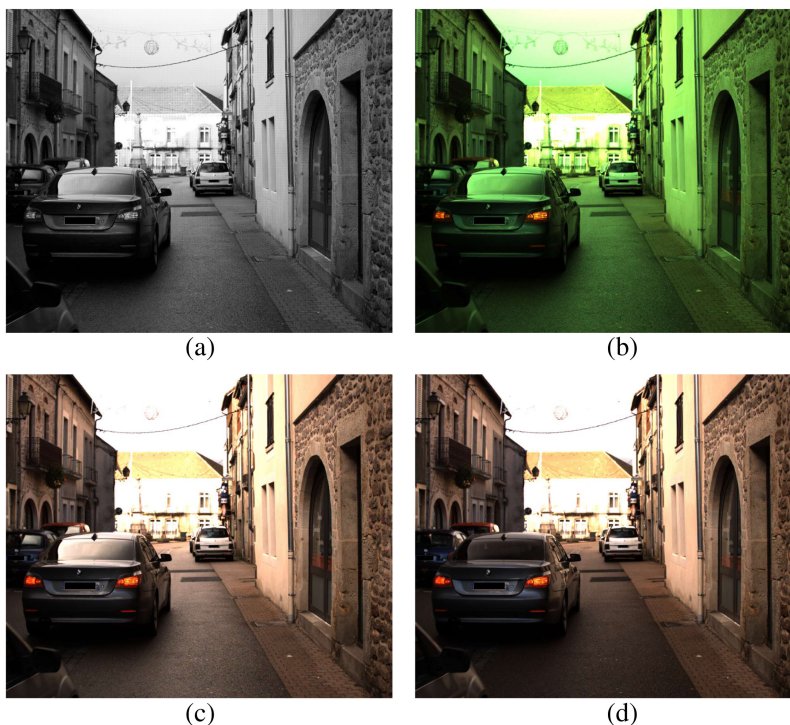

Fig. 7 (a) Raw mosaiced version of the reference image, for the polarization channel 0 deg. (b) Demosaiced version of the reference image, without white-balance, for the 0 deg polarization channel. (c) White-balanced demosaiced images for the polarization channel 0 deg. (d) White-balanced demosaiced images for the polarization channel 135 deg.

This indicates that the light reflected by the windshields is polarized because polarized light is filtered by polarimetric filters according to their orientations. As explained in Sec. 2, the intensity of linearly polarized light that passes through a rotating linear polarizer will exhibit a sine wave curve with respect to the rotation angle. This sine wave is maximum when the orientation of the filter is equal to the angle of the linearly polarized light and it decreases as the filter rotates. The intensity reaches a minimum when the orientation of the filter is shifted by $\pi/2$ radians with respect to the angle of linear polarization of the light. In the case of the RGB-polarization camera, four filter orientations are considered. If any of these orientations matches the AoLP of the incoming polarized light, then the corresponding image will have a bright spot, and the one that is shifted by $\pi/2$ radians will have a dark spot at the same position. The windshields of the cars in Fig. 7(d), for a filter orientation of 135 deg, are dark, whereas in Fig. 7(c), for a filter orientation of 0 deg, it is rather bright; this means that the 135 deg filter has filtered out a great part, if not all, of the polarized reflections. Thus, we can deduce that the reflected light has an AoLP close to 45 deg, and that it is highly linearly polarized.

Original color: only one image is returned. This image is equivalent to the one obtained with a conventional RGB camera. It corresponds to the first component of the Stokes vector S_0. It is computed by applying Eq. (9) to the output of the Raw split images mode. The obtained image is then demosaiced to obtain the three-channel color image.

Stokes images: after separating the input image by polarization channel, the three Stokes components images can be obtained by applying Eq. (9). Since the polarization state of the light depends on the frequency of the light, the Stokes vectors are split by color channel. As a consequence, this function will provide $3 \times 4 = 12$ images. Four color channels are considered since the Bayer patterns consist of 2×2 patterns of red, green, green, and blue color filters. The three linear components, S_0, S_1, and S_2, of the Stokes vector for the red channel are shown in Figs. 8(a)–8(c). The images for the other channels are included in the GitHub repository.[80]
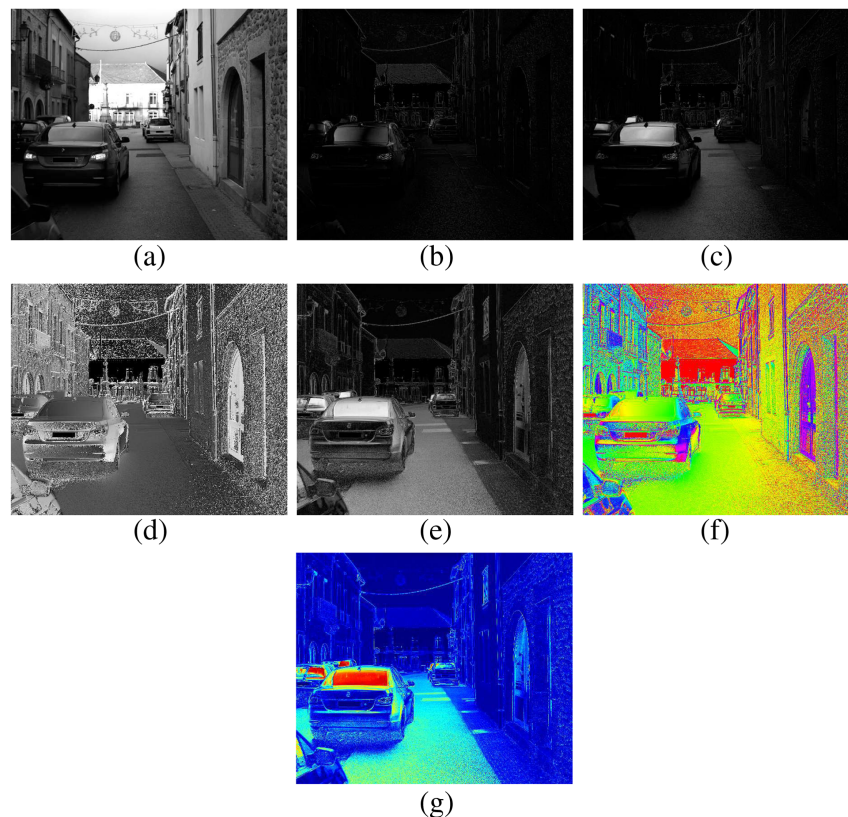

Fig. 8 Polarization images for the red channel. (a)–(c) $S_{0,1,2}$ components of the Stokes vector. The parameter S_0 corresponds also to the total intensity image. (d) and (e) Raw AoLP and Raw DoLP. (f) Raw AoLP of the red channel, colored with the HSV palette. (g) Raw DoLP of the red channel colored with the Jet palette.

The S_0 image is the red channel image of the original color image, whereas the S_1 and S_2 images are functions of the orientation of the polarizer filters. Since S_1 and S_2 can have negative values, it is their absolute values that are considered. In the same way as polarized color images, Stokes component images can be used to detect regions with polarized light. Indeed, polarized light will be characterized by a bright or high Stokes value region in one of the S_1 and S_2 images and a corresponding dark or low Stokes value region in the other image. Conversely, regions that appear dark in both images correspond to areas where the reflected light is weakly polarized. In Figs. 8(a) and 8(c), we can deduce that the windshields exhibit highly polarized reflections. This is consistent with what we have already deduced from the polarization color images.

Raw I - Rho - Phi: as also explained in Sec. 2, the Stokes vector can be represented as a function of three physical parameters: the total intensity I, the degree of linear polarization ρ, and the AoLP ϕ. The equations to compute these parameters as a function of the Stokes vector are given in Eq. (1). Again, these three physical parameters can be computed for each color channel, and thus 12 images are also returned in this mode. Since each parameter has a different interval of values, all of them have been normalized to be in the range [0,255]. The resulting images for the red channel only are shown in Figs. 8(a), 8(d), and 8(e), respectively. These images are a good representation of all the objects that reflect linearly polarized light. Note that a pixel value in the AoLP image has a meaning only if its corresponding pixel in the DoLP image has a non-zero value. In this set of images, it is possible to confirm that the road, the windshields, part of the body of the car, and some door glasses have a large DoLP. Thus, the AoLP and DoLP features are extremely valuable and can be used, for example, in deep learning models to improve the accuracy of the network results for these objects.

I - Rho - Phi: in the Raw I - Rho - Phi mode, all the images are single channeled, thus they are displayed in gray-scale. Particularly for the AoLP, this is not an adequate representation since it is a circular variable. A proper representation would be one that assigns the same color to the maximum and minimum values. In this mode, this is done by creating a color palette based on the HSV color space. Let X be a gray-scale value in the range [0,255]. The HSV palette is defined as a function that assigns a 3D vector $\mathbf{c_{hsv}}$ to each value X, such that

$$\mathbf{c_{hsv}} = \left[\frac{179X}{255}, 255, 255 \right]^T. \tag{10}$$

Then, the obtained three-channeled image is converted from the HSV to the RGB color space. On the other hand, the degree of linear polarization is colorized using the Jet palette, which maps blue colors to low values, green colors to middle values, and red colors to high values. As in the Raw I - Rho - Phi mode, 12 images are returned in this mode. The results of this mode for the colored AoLP and DoLP are shown in Figs. 8(f) and 8(g), respectively, for the red channel.

Fake colors: when dealing with polarization imaging, a correspondence is established between the HSV (hue, saturation, and value) color space and the intensity, degree of linear polarization, and AoLP.[81] Since the AoLP is a circular variable, it is considered to be the hue of the color. The saturation is the purity of that color, meaning that a saturation of 100% is a pure color, and a saturation of 0% means that it is a gray-level value. Similarly, the degree of linear polarization indicates the "purity" of the light: if it is totally unpolarized, ρ is equal to zero, and if it is totally linearly polarized, ρ is equal to 1. Finally, if a conventional color image is considered, and the value channel is extracted from it in the HSV space, a gray-scale version of the original image is obtained. This information is similar to the total intensity measured by the I parameter. Thus, a color image can be obtained if the I, ϕ, and ρ images are stacked together, considered to be in the HSV space, and then converted back to the RGB space. The result obtained in this way is called a fake color image. The colors obtained with this processing algorithm have the following properties.

- Unpolarized light is represented by gray-scale colors.
- Highly polarized light will be colored.
- The colors in the image depend mainly on the AoLP, thus, to the surface orientation.

This processing is interesting since it helps to quickly identify the objects that reflect polarized light. As explained before, the polarization parameters depend on the frequency, thus the

Fig. 9 (a) White-balanced color image. (b) Fake colors image representation for the red channel.

fake color images are separated by color channel. Therefore, this operation returns four images. The resulting image for the red channel is shown in Fig. 9.

5.2.1 *White-balance module*

The observed light by a camera depends on two parameters: the observed object reflectance, and the illumination of the color.[22] Thus, a white-balance algorithm needs to be used in order to restore the true colors in the scene. In some polarization cameras, and particularly in the Basler RGB-polarization camera, this type of algorithm is not correctly implemented. Thus, our software includes an implementation of a white-balance algorithm. The white balance is a gain applied to each color channel to get the true color of the objects despite the environment where the image is taken. In our implementation, the automatic white search is done globally, and not in a user-defined region of interest. The algorithm computes the average of all the color channels of a single orientation, and it searches for the pixels whose average is the highest. If there is a white piece in the scene, even with the color gains unbalanced, its average will be the highest. Thus, the pixel whose average is the highest is considered as white. Then, the highest channel value is left untouched (gain equal to 1), and the other channel gains are computed such that their values equal the highest channel value. This algorithm of automatic white balance is constrained to work with no high-level saturated images. It can be deactivated and the different gains set manually. An example of the input image to this algorithm, and its corresponding output are shown in Figs. 7(b) and 7(c), respectively.

5.3 Polarimetric Camera Calibration Module

This module is an implementation of the polarimetric camera calibration algorithm described in Ref. 7. The calibration algorithm is used to compute a series of matrices that are applied by a correction function to rectify the measurement errors due to manufacturing imperfections. In this way, two super-pixels of the same color channel that receive the same light source will provide the same output measurements. The module also provides a function to correct the image, once the calibration algorithm has been run. This function takes a raw image from the camera, and if the calibration matrices have been computed, the function will return another image, with the same structure as the input, but with all the pixel measurements corrected by the calibration algorithm.

The calibration problem can be solved by taking several images of a uniform and linearly polarized light. If the light source is uniform but unpolarized, it can be polarized using a linear polarization filter. By turning the filter, the light received by the camera at each filter position will have a different AoLP. A sample of a polarized light source with an AoLP of 40 deg is shown in Fig. 10, for the 0 deg and 45 deg polarization channels. These images correspond to the results before and after applying the calibration algorithm.

The calibration procedure will compute the pixel parameters given the model defined in Ref. 7, i.e., (T_i, P_i, θ_i). In this model: T_i is the pixel gain; P_i is a parameter that accounts for the non-ideality of the micro-polarization filter implemented on the pixel; θ_i is the effective orientation of the micro-polarization filter of the pixel; i is the position of the pixel considered. From
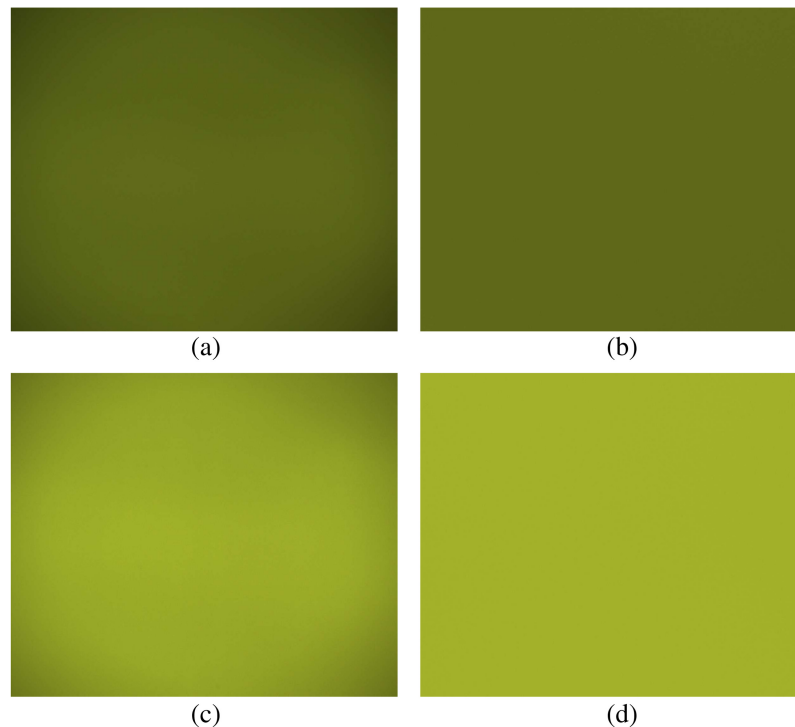
Fig. 10 Calibration effects over a uniform, linearly polarized light, with an AoLP of 40 deg. These images are the demosaiced RGB polarization channels. (a) and (c) Uncalibrated polarization channels 0 deg and 45 deg, respectively. (b) and (d) Calibrated polarization channels 0 deg and 45 deg, respectively.

Mueller calculus, these parameters have an ideal value of $(T_i, P_i) = (0.5, 1.0)$ for all i, and $\theta_i \in \{0 \text{ deg}, 45 \text{ deg}, 90 \text{ deg}, 135 \text{ deg}\}$. However, in general, each pixel will have a set of parameters that will be different from these ideal values. Thus, the calibration algorithm allows to compensate for these differences.

To assess the acquisition quality, the different plot functions of the software can be used. With these plots, the user can determine if the acquisition is correct, and if the camera measurements are valid. It can also assess the quality of the sensor and confirm the effectiveness of the correction on the camera measurements.

Among the available plots, the histograms of the intensity, the DoLP, and the AoLP of the incident light can be computed from the image currently displayed by the software. This information allows to evaluate the quality of the calibration results. After correction, these three parameters (intensity, DoLP, and AoLP) will have a narrower distribution than the uncalibrated case. These histograms before and after calibration are represented in Figs. 11(a) to 11(f). They have been computed with the same sample image, mentioned before, of a polarized light with an AoLP of 40 deg.

A real-time plot of these three parameters (intensity, DoLP, and AoLP) can be done for a given row of pixels of the sensor. These graphs are displayed in the last two columns of Fig. 11, and they correspond to the measurements before and after applying the calibration. These plots illustrate the vignetting effect and show how calibration can reduce its impact over the three polarization parameters. It is important to note that this correction is achieved since the pixel model used considers the polarization parameters of the pixel, and not only the unbalanced sensing gain. A simple gain correction will only affect the intensity image but not the AoLP nor the DoLP images.

Finally, the consequence of applying the calibration can be observed in the reference urban scene image. The effects of the calibration over the intensity, AoLP, and DoLP images are shown in Fig. 12. A zoom has been added to the top and bottom right areas of the image where it is possible to see that the vignetting effect has been corrected. Particularly from these images, the contribution of the calibration can be observed mostly in the AoLP and the intensity images. In
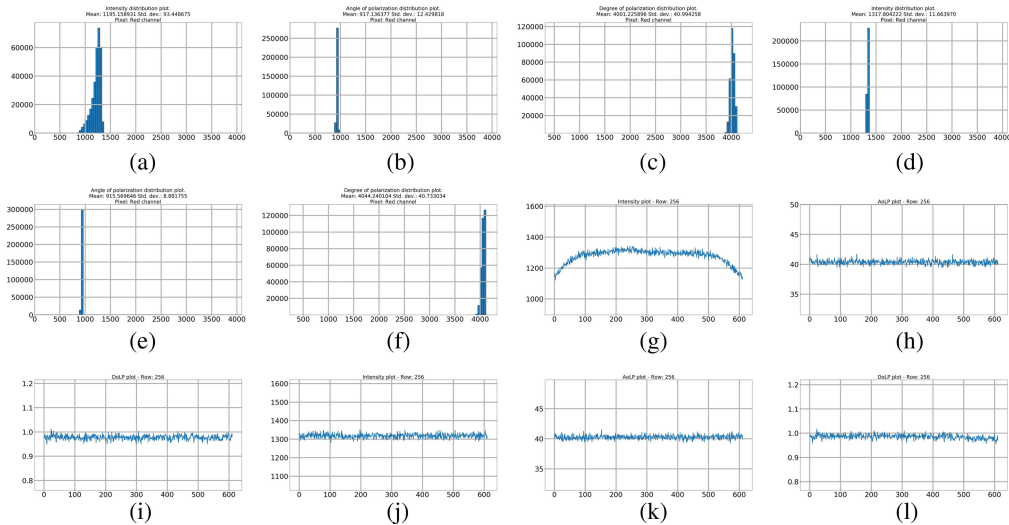
Fig. 11 Calibrated camera quality evaluation. All the plots are computed before (first and third columns) and after (second and fourth columns) calibration. (a) and (d) Intensity histograms. (b) and (e) AoLP histograms. (c) and (f) DoLP histograms. (g) and (j) Intensity measurements plots over a single row. (h) and (k) AoLP measurements plots over a single row. (i) and (l) DoLP measurements plots over a single row.
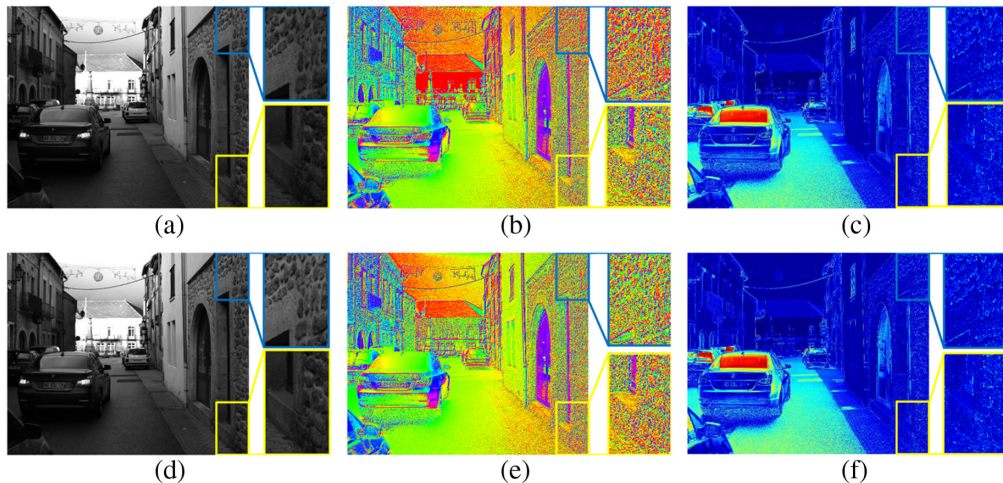

Fig. 12 Calibration effects over the polarization images, for the red channel. Top row: uncalibrated images. Bottom row: calibrated images. (a) and (d) Total intensity. (b) and (e) Angle of linear polarization, colored with the HSV palette. (c) and (f) Degree of linear polarization, colored with the Jet palette.

the scene, there are several walls that act as planes. Thus, they should reflect the same AoLP, i.e., they should have the same color. This happens only after calibration, mainly in the building situated in the far region of the image, and in the walls on both sides of the road. In the intensity image, since the pixel model includes a gain factor, the vignetting effect is also corrected with this system, making the darker areas in the borders to be as bright as the center of the image.

5.4 Polarization Processing Algorithms

Differently from Sec. 5.2 where basic polarimetric operations can be done, in this module two applications of the polarization concepts are implemented. The first one is the simulated polarization filter. As explained in Sec. 2, each super-pixel allows the computation of the Stokes vector $\mathbf{s} = [S_0, S_1, S_2]^T$ of the incident light. Now, let us consider a light, described by the Stokes vector $\mathbf{s}_{\text{in}}$ that passes through a linear polarizer. The filter is oriented at an angle θ and modeled by a

Mueller matrix **M**, as explained in Sec. 2. Therefore, the effects of a linear polarizer in front of a normal camera can be simulated by computing Eq. (7). Thus in this functionality, the inputs are a raw image from the camera and the orientation of the filter θ that one would like to simulate. Then, this algorithm returns two images: the input image and the filtered image after applying Eq. (7) to all the super-pixels. This functionality is commonly used in photography to remove annoying polarized reflections from the environment. In a real system with a conventional RGB camera, the filter is physically placed on top of the lens, and turned until the reflection is removed. With this software, the filter and its effects are simulated after the image has been captured, and the exact angle for reflection removal can be found. This orientation is equal to the AoLP of the incident light shifted by 90 deg. As an example, the results of simulating a linear filter with the angles $\theta = 60$ deg and $\theta = 120$ deg are shown in Figs. 13(b) and 13(c), respectively. Note that the choice of an angle or another will either reinforce or erase the reflections on the windshields.

The second functionality of this module is the polarized specularity removal. It is an extension of the simulated polarization filter, explained above. In the previous case, all the pixels are affected by a single polarization filter. But, in a scene, there might be several objects that produce this type of reflection, with different AoLP. To erase them all at once, let us consider the Stokes vector of the observed light. This vector can be split into two other Stokes vectors: one that represents totally unpolarized light and another that represents a totally linearly polarized light $\mathbf{s} = \mathbf{s_{unpol}} + \mathbf{s_{pol}}$ such as

$$
\begin{bmatrix} S_0 \\ S_0 \rho \cos(2\phi) \\ S_0 \rho \sin(2\phi) \end{bmatrix} = (1-\rho) \begin{bmatrix} S_0 \\ 0 \\ 0 \end{bmatrix} + \rho \begin{bmatrix} S_0 \\ S_0 \cos(2\phi) \\ S_0 \sin(2\phi) \end{bmatrix}. \tag{11}
$$

Removing the polarized reflection means erasing the component corresponding to $\mathbf{s_{pol}}$. This is equivalent to computing the $\mathbf{s_{unpol}}$ vector. This functionality returns two images: the input and the filtered images, both demosaiced. In contrast to the previous case, this functionality does not
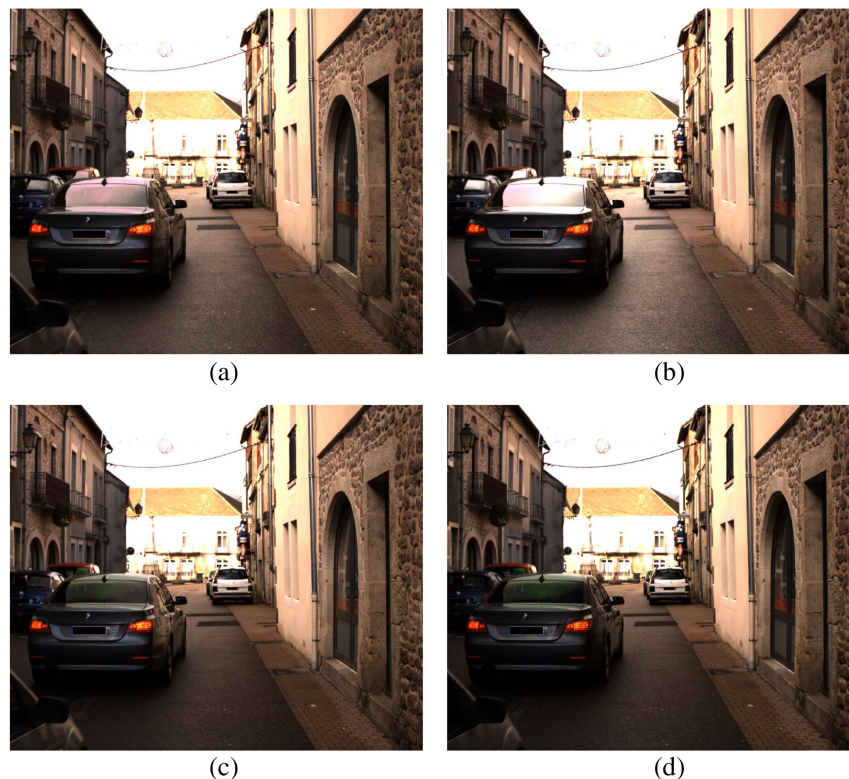

Fig. 13 Specularity removal application results. (a) Unfiltered total intensity. (b) and (c) Filtered image by simulating a linear polarization filter oriented at 60 deg and 120 deg, respectively. (d) Polarized specularity removal by filtering with the degree of linear polarization.

require the user to enter an angle to each filter. The filtering is done based on the measured DoLP at each super-pixel. The results of the filtered image are shown in Fig. 13(d). One can note that most reflections from the shiny surfaces, such as the windshield, the road, and the door glasses, have been removed.

6 Discussion

Recent technological advances have made possible new sensors that allow for capturing RGB and polarization information in a single snapshot. This is an important step forward in the development of the polarization field since it reduces the movement constraints and the acquisition time for this modality. This is one of the main reasons why many research works leveraging polarization have been published in recent years with impressive results. From a global point of view, all of the works make use of the AoLP and the DoLP as principal added features to an RGB system. The Fresnel equations are used as geometrical constraints to better guide an optimization algorithm or a data-driven model training toward more consistent results. This last point can be used in two ways: either by computing the normal maps out of the Fresnel equations and providing these maps as input cues for a deep learning network or using these relationships to compute an energy function that will be later optimized. Going further, some works[27,29] use the raw measurements to avoid adding non-linearities to the relationships between the polarization state and the intensity measurements. However, there is still much to explore since one of the reasons why polarization has not been widely adopted in the computer vision field is the required controlled acquisition conditions. In other words, many applications have been developed in laboratory conditions or in environments where there are no perturbations that might affect the measurements. One example of this is in shape from polarization algorithms,[11,42,44] where often the properties of the object being reconstructed are known beforehand. Even though experiments can be done in outdoor conditions, the fact of having a single object to analyze prevents their generalization to a more generic scene in which several objects, with different properties, are present at the same time. Another example of the generalization limitation are algorithms developed under simulated conditions, such as the underwater image enhancement algorithm Li et al.[30] In this particular case, the results show a clear improvement with respect to RGB systems, but the experiments have been done in a small-sized water tank, with known light, and no perturbations, such as waves or the presence of other objects inside the tank. These constraints are not in line with the requirements of a real-world underwater situation, leaving its actual applicability unknown. It is worth noting that for data-driven algorithms, it is hard to make acquisitions of large number of images of highly reflective surfaces. These types of surfaces are known to be challenging for RGB-only algorithms and the polarization information can provide valuable cues. However, to obtain a ground truth regarding the shape or the depth, sensors such as the Microsoft Kinect or LiDAR must be used, but the reflective or transparent nature of the objects can result in inaccurate measurements by the sensors. Therefore, there are still challenges in handling these types of situations. As also mentioned in the different categories of Sec. 4, there is a lack of common standards to share data, acquisition and basic signal processing tools to process polarization information. In this context, we have proposed a software library toolkit that can serve to the objectives of facilitating the acquisition, application deployment, and comparison of techniques with polarization cues. We expect this toolkit to be a step further to help promoting the polarization imaging field in less constrained conditions.

7 Conclusion

In this paper, we have done a comprehensive review of recent works that have leveraged the polarization modality in robotics and computer vision fields, summarizing some of their main contributions and limitations. The rich information brought by RGB-polarization cameras has allowed a huge improvement in the analysis of scenes where RGB-only systems generally fail. The number of works in the recent years has increased due to the advent of new and cheap RGB-polarization sensors, and data-driven algorithms. These devices facilitate data acquisition, in less constrained environments, and in real-time with a single snapshot. However, there is still a long way to the vision community to fully exploit the potential and the benefits of this

modality. There is no general rule to follow to extend most existing conventional RGB algorithms to RGB-polarization cameras. As discussed before, just providing the polarization information as an additional input to an algorithm is unlike to provide better results. A good understanding of the physics and principles of light polarization is required to be able to get the most out of this modality. Furthermore, the lack of standards, common processing tools, and large-scale benchmarks limit the development of methods that extend the results obtained for traditional vision systems since it is often not possible to train and compare different approaches over a common dataset. To allow a greater number of researchers to easily and rapidly acquire and analyze polarimetric images, we have developed an open-source software library toolkit that makes available commonly used processing algorithms. This software has been carefully developed to allow easy maintenance and the addition of features. We hope this work will encourage other researchers to contribute to this field and to participate in the expansion of the functionalities included in the presented toolkit, with the aim of making it accessible to anyone who wants to explore polarization.

References

1. M. Garcia et al., "Bio-inspired color-polarization imager for real-time in situ imaging," *Optica* **4**, 1263–1271 (2017).
2. M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, Elsevier (2013).
3. D. Goldstein, *Polarized Light*, CRC Press (2011).
4. G. Horváth and D. Varjú, *Circulary Polarized Light in Nature*, pp. 100–103, Springer Berlin Heidelberg, Berlin, Heidelberg (2004).
5. Z. Ding et al., "Calibration method for division-of-focal-plane polarimeters using nonuniform light," *IEEE Photonics J.* **13**(1), 3900309 (2021).
6. C. Lane, D. Rode, and T. Roesgen, "Calibration of a polarization image sensor andinvestigation of influencing factors," *Appl. Opt.* **61**, C37–C45 (2021).
7. J. Rodriguez et al., "A practical calibration method for RGB micro-grid polarimetric cameras," *IEEE Rob. Autom. Lett.* **7**(4), 9921–9928 (2022).
8. J. S. Tyo, "Design of optimal polarimeters: maximization of signal-to-noise ratio and minimization of systematic error," *Appl. Opt.* **41**, 619–630 (2002).
9. E. Hecht, *Optics*, Pearson Education, Addison-Wesley (2002).
10. W. A. P. Smith, R. Ramamoorthi, and S. Tozza, "Linear depth estimation from an uncalibrated, monocular polarisation image," *Lect. Notes Comput. Sci.* **9912**, 109–125 (2016).
11. D. Zhu and W. A. P. Smith, "Depth from a polarisation + RGB stereo pair," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 7578–7587 (2019).
12. Z. Cui, V. Larsson, and M. Pollefeys, "Polarimetric relative pose estimation," in *IEEE/CVF Int. Conf. Comput. Vision (ICCV)*, pp. 2671–2680 (2019).
13. R. Rojas, *The Backpropagation Algorithm*, pp. 149–182. Springer Berlin Heidelberg, Berlin, Heidelberg (1996).
14. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Adv. Neural Inf. Process. Syst. 25* (2012).
15. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd Int. Conf. Learn. Represent. (ICLR 2015)*, Computational and Biological Learning Society, pp. 1–14 (2015).
16. D. Sinha and M. El-Sharkawy, "Thin mobilenet: an enhanced mobilenet architecture," in *IEEE 10th Annu. Ubiquit. Comput., Electron. & Mob. Commun. Conf. (UEMCON)*, IEEE, pp. 0280–0285 (2019).
17. B. Koonce, *EfficientNet*, pp. 109–123, Apress, Berkeley, California (2021).
18. S. C. Yurtkulu, Y. H. Åžahin, and G. Unal, "Semantic segmentation with extended deeplabv3 architecture," in *2019 27th Signal Process. and Commun. Appl. Conf. (SIU)*, Sivas, Turkey, pp. 1–4 (2019).

19. K. He et al., "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 770–778 (2016).

20. A. Vaswani et al., "Attention is all you need," in *Adv. in Neural Inf. Process. Syst.*, I. Guyon et al., Eds., Curran Associates, Inc., Vol. **30**, pp. 5998–6008 (2017).

21. A. Dosovitskiy et al., "An image is worth 16x16 words: transformers for image recognition at scale," in *Int. Conf. Learn. Represent.* (2021).

22. T. Ono et al., "Degree-of-linear-polarization-based color constancy," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 19708–19717 (2022).

23. M. Morimatsu et al., "Monochrome and color polarization demosaicking using edge-aware residual interpolation," in *IEEE Int. Conf. Image Process. (ICIP)*, pp. 2571–2575 (2020).

24. S. Wen, Y. Zheng, and F. Lu, "A sparse representation based joint demosaicing method for single-chip polarized color sensor," *IEEE Trans. Image Process.* **30**, 4171–4182 (2019).

25. J. Zhang et al., "Sparse representation-based demosaicing method for microgrid polarimeter imagery," *Opt. Lett.* **43**, 3265–3268 (2018).

26. H. Hu et al., "IPLNet: a neural network for intensity-polarization imaging in low light," *Opt. Lett.* **45**, 6162–6165 (2020).

27. C. Zhou et al., "Learning to dehaze with polarization," in *Adv. Neural Inf. Process. Syst.*, M. Ranzato et al., Eds., Curran Associates, Inc., Vol. 34, pp. 11487–11500 (2021).

28. J. Liu et al., "Semantic-guided polarization image fusion method based on a dual-discriminator GAN," *Opt. Express* **30**, 43601–43621 (2022).

29. C. Lei et al., "Polarized reflection removal with perfect alignment in the wild," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 1747–1755 (2020).

30. X. Li et al., "Underwater image restoration via Stokes decomposition," *Opt. Lett.* **47**, 2854–2857 (2022).

31. H. Hu et al., "Polarimetric underwater image recovery via deep learning," *Opt. Lasers Eng.* **133**, 106152 (2020).

32. K. Tanaka, Y. Mukaigawa, and A. Kadambi, "Polarized non-line-of-sight imaging," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 2133–2142 (2020).

33. K. O. Amer et al., "Enhancing underwater optical imaging by using a low-pass polarization filter," *Opt. Express* **27**, 621–643 (2019).

34. L. Shen and Y. Zhao, "Underwater image enhancement based on polarization imaging," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XLIII-B1-2020**, 579–585 (2020).

35. S. Wen, Y. Zheng, and F. Lu, "Polarization guided specular reflection separation," *IEEE Trans. Image Process.* **30**, 7280–7291 (2021).

36. Y. Liang et al., "Multimodal material segmentation," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 19768–19776 (2022).

37. H. Mei et al., "Glass segmentation using intensity and spectral polarization cues," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 12612–12621 (2022).

38. A. Kalra et al., "Deep polarization cues for transparent object segmentation," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 8599–8608 (2020).

39. N. Li et al., "Illumination-invariant road detection and tracking using LWIR polarization characteristics," *ISPRS J. Photogramm. Remote Sens.* **180**, 357–369 (2021).

40. N. Li et al., "Polarization-guided road detection network for LWIR division-of-focal-plane camera," *Opt. Lett.* **46**, 5679–5682 (2021).

41. K. Xiang, K. Yang, and K. Wang, "Polarization-driven semantic segmentation via efficient attention-bridged fusion," *Opt. Express* **29**, 4802–4820 (2021).

42. T. Ichikawa et al., "Shape from sky: polarimetric normal recovery under the sky," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 14827–14836 (2021).

43. C. Lei et al., "Shape from polarization for complex scenes in the wild," in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 12632–12641 (2022).

44. Y. Fukao et al., "Polarimetric normal stereo," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 682–690 (2021).

45. Y. Ba et al., "Deep shape from polarization," *Lect. Notes Comput. Sci.* **12369**, 554–571 (2020).

46. V. Deschaintre, Y. Lin, and A. Ghosh, "Deep polarization imaging for 3D shape and SVBRDF acquisition," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 15562–15571 (2021).

47. K. Berger, R. Voorhies, and L. H. Matthies, "Depth from stereo polarization in specular scenes for urban robotics," in *IEEE Int. Conf. Rob. and Autom. (ICRA)*, pp. 1966–1973 (2017).

48. M. Blanchon et al., "P2D: a self-supervised method for depth estimation from polarimetry," in *25th Int. Conf. Pattern Recognit. (ICPR)*, pp. 7357–7364 (2021).

49. M. Shakeri et al., "Polarimetric monocular dense mapping using relative deep depth prior," *IEEE Rob. Autom. Lett.* **6**(3), 4512–4519 (2021).

50. J. Zhao, Y. Monno, and M. Okutomi, "Polarimetric multi-view inverse rendering," *Lect. Notes Comput. Sci.* **12369**, 85–102 (2020).

51. Y. Kondo et al., "Accurate polarimetric BRDF for real polarization scene rendering," *Lect. Notes Comput. Sci.* **12364**, 220–236 (2020).

52. D. Gao et al., "Polarimetric pose prediction," *Lect. Notes Comput. Sci.* **13669**, 735–752 (2022).

53. S. Zou et al., "Human pose and shape estimation from single polarization images," *IEEE Trans. Multimedia* **25**, 3560–3572 (2023).

54. M. Tzabari and Y. Y. Schechner, "Polarized optical-flow gyroscope," *Lect. Notes Comput. Sci.* **12361**, 363–381 (2020).

55. P. Hu et al., "Solar-tracking methodology based on refraction-polarization in Snell's window for underwater navigation," *Chin. J. Aeronaut.* **35**(3), 380–389 (2022).

56. X. Li et al., "Polarimetric imaging via deep learning: a review," *Remote Sens.* **15**(6), 1540 (2023).

57. S. B. Powell and V. Gruev, "Calibration methods for division-of-focal-plane polarimeters," *Opt. Express* **21**, 21039–21055 (2013).

58. S. Kajiyama et al., "Separating partially-polarized diffuse and specular reflection components under unpolarized light sources," in *IEEE/CVF Winter Conf. Appl. of Comput. Vision (WACV)*, pp. 2548–2557 (2023).

59. S. Umeyama and G. Godin, "Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(5), 639–647 (2004).

60. S. Boyd et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.* **3**, 1–122 (2011).

61. D. Kiku et al., "Beyond color difference: residual interpolation for color image demosaicking," *IEEE Trans. Image Process.* **25**(3), 1288–1300 (2016).

62. L. Wolff, "Polarization-based material classification from specular reflection," *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(11), 1059–1071 (1990).

63. H. Chen and L. Wolff, "Polarization phase-based method for material classification and object recognition in computer vision," in *Proc. CVPR IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognit.*, pp. 128–135 (1996).

64. S. Tominaga and A. Kimachi, "Polarization imaging for material classification," *Opt. Eng.* **47**(12), 123201 (2008).

65. V. Thilak, D. G. Voelz, and C. D. Creusere, "Polarization-based index of refraction and reflection angle estimation for remote sensing applications," *Appl. Opt.* **46**, 7527–7536 (2007).

66. M. P. Khaing and M. Masayuki, "Transparent object detection using convolutional neural network," in *Big Data Analysis and Deep Learning Applications*, T. T. Zin and J. C.-W. Lin, Eds., pp. 86–93, Springer Singapore, Singapore (2019).

67. Z. Peng et al., "Conformer: local features coupling global representations for visual recognition," in *IEEE/CVF Int. Conf. Comput. Vision (ICCV)*, pp. 357–366 (2021).

68. M. Cordts et al., "The cityscapes dataset for semantic urban scene understanding," in *IEEE Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 3213–3223 (2016).

69. D. Miyazaki et al., "Polarization-based inverse rendering from a single view," in *Proc. Ninth IEEE Int. Conf. Comput. Vision*, Vol. 2, pp. 982–987 (2003).

70. O. Morel et al., "Polarization imaging applied to 3D reconstruction of specular metallic surfaces," *Proc. SPIE* **5679**, 178–186 (2005).

71. G. A. Atkinson and E. R. Hancock, "Surface reconstruction using polarization and photometric stereo," *Lect. Notes Comput. Sci.* **4673**, 466–473 (2007).

72. L. Zhang and E. R. Hancock, "A comprehensive polarisation model for surface orientation recovery," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR2012)*, pp. 3791–3794 (2012).

73. C. Godard et al., "Digging into self-supervised monocular depth estimation," in *IEEE/CVF Int. Conf. Comput. Vision (ICCV)*, pp. 3827–3837 (2019).

74. K. Kim, A. Torii, and M. Okutomi, "Multi-view inverse rendering under arbitrary illumination and albedo," *Lect. Notes Comput. Sci.* **9907**, 750–767 (2016).

75. R. Ranftl et al., "Towards robust monocular depth estimation: mixing datasets for zero-shot cross-dataset transfer," *IEEE Trans. Pattern Anal. Mach. Intell.* **44**(3), 1623–1637 (2022).

76. J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *IEEE Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 4104–4113 (2016).

77. M. Loper et al., "SMPL: a skinned multi-person linear model," *ACM Trans. Graphics* **34**, 248 (2015).

78. T. Du et al., "An autonomous initial alignment and observability analysis for sins with bio-inspired polarized skylight sensors," *IEEE Sens. J.* **20**(14), 7941–7956 (2020).

79. J. Li et al., "Bio-inspired attitude measurement method using a polarization skylight and a gravitational field," *Appl. Opt.* **59**, 2955–2962 (2020).

80. J. Rodriguez et al., "Pola4all: a survey of polarimetric applications and an open-source software to analyze polarimetric images – repository," 2023, https://github.com/vibot-lab/Pola4all_JEI_2023.

81. L. B. Wolff, "Polarization vision: a new sensory approach to image understanding," *Image Vision Comput.* **15**(2), 81–93 (1997).

Joaquin Rodriguez is a PhD student at the University of Burgundy, for the VIBOT team of the ImViA laboratory. He received his electrical engineering and master's degree at the National University of Rosario, Argentine, in 2017. In 2023, he finished his PhD in computer vision for robotics in the field of polarimetric imaging. He is the author of two journal papers. His current research interests include polarimetric imaging applied to depth estimation and object classification.

Lew-Fock-Chong Lew-Yan-Voon is an associate professor at the Université de Bourgogne, France. He received his PhD in computer aided design of VLSI circuits from Université Montpellier II, France, in March 1992. Since then he has been a temporary research scientist at laboratory LIRMM until September 1993 when he joined the laboratory ImViA of the Université de Bourgogne, France. His research interests lie in the field of signal and image processing for pattern recognition.

Renato Martins, PhD, is an assistant professor at the Université de Bourgogne, France. He holds BSE degree in control and automation engineering and MSc degree in electrical engineering, all from the University of Campinas, Brazil and a PhD in computer science carried-out at INRIA Sophia Antipolis, France. His main research interests are in computer vision and machine learning, more specifically on 3D vision, geometric deep learning, image descriptors, and image understanding and synthesis.

Olivier Morel is an associate professor in the VIBOT team of the ImViA laboratory. He received his engineering degree in automatic and industrial computing system from Université de Savoie (Polytech'Savoie) in 2001. He graduated from the University of Burgundy in 2002 and 2005 respectively with an MSc degree and PhD in computer vision and image processing. His main research interests are polarization imaging, 3D vision, and applications of these techniques to robotics navigation.