Extrapolation-free arbitrary-shape motion estimation using phase correlation

Vasileios Argyriou
Theodore Vlachos
University of Surrey
Centre for Vision, Speech and Signal Processing
School of Electronics and Physical Sciences
GU2 7XH United Kingdom
E-mail: V.Argyriou@surrey.ac.uk

Abstract. *We propose a frequency domain scheme for obtaining subpixel estimates of interframe motion for arbitrary shaped regions. Our scheme is based on phase correlation and the shape adaptive discrete Fourier transform and one of its key features is that it does not require extrapolation, which requires extra complexity and can dilute the accuracy of the estimated motion parameters. We demonstrate that our method outperforms in terms of subpixel accuracy and motion-compensated prediction error both conventional phase correlation but also shape adaptive techniques operating in the frequency domain and requiring extrapolation. © 2006 SPIE and IS&T.* [DOI: 10.1117/1.2170582]

1 Introduction

Motion estimation is a critical component of various video processing tasks, especially video compression, allowing redundancy reduction in the temporal domain. International standards for video communications such as MPEG-1/2 and H.261/3/4 employ the well-established hybrid two-component architecture, which relies on motion estimation and compensation as well as on the lossy compression of the motion-compensated prediction error. Motion estimation in such standards is carried out by means of block matching in the data domain with one motion vector portraying the motion of each block. Such block-based approaches offer well-documented implementation advantages such as low complexity and low overheads mainly due to their regularity. On the other hand, they have a number of well-known disadvantages. A block may contain more moving objects than just one, in which case a single motion vector derived from the most dominant object will cause large motion compensation errors in areas occupied by the other objects. Conversely, a moving object may be contained in more blocks than one. Any errors in estimating motion vectors for those blocks are likely to cause blocking artefacts. Various attempts at departing from established block-based approaches have been made, most notably in video coding systems like MPEG-4, while in H.264 a variable block size approach was preferred to shape coding. While motion estimation is not a normative element, most compliant architectures implement arbitrary shape motion estimation in the pixel domain using suitable modifications to the well-known block-matching algorithm, such as extrapolation of the arbitrary shaped area until the limits of a bounding rectangle are reached. Such pixel-domain modifications inherit some of the disadvantages of baseline block matching such as complexity, assuming an exhaustive search strategy for the identification of the minimum error location. In the wider literature, object-based motion estimation is well represented albeit most approaches operate in the pixel domain. An exhaustive review would be outside the scope of this paper but it is worth mentioning Refs. 1–4 to name but a few.

Recently there has been a lot of interest in motion estimation techniques operating in the frequency domain because they offer well-documented advantages in terms of computational efficiency due to the employment of fast algorithms. Perhaps the best-known method in this class is phase correlation (PC),[5] which has become one of the motion estimation methods of choice for a wide range of non-consumer applications.[6] In addition to computational efficiency, PC offers key advantages in terms of its strong response to edges and salient picture features, its immunity to illumination changes and moving shadows, and its ability to measure large displacements without sacrificing subpixel accuracy.

In this letter we propose an arbitrary-shape motion estimation algorithm based on PC. While in Ref. 7 we proposed a solution using conventional extrapolation, i.e., by padding with the average (mean) intensity of the arbitrary-shaped object, in this work we present an extrapolation-free algorithm using the shape adaptive discrete Fourier transform (SA-DFT) methodology.[8]

This letter is organized as follows. In Sec. 2 we briefly review the principles underlying subpixel motion estimation using PC. In Sec. 3 we present the arbitrary-shape phase correlation algorithm. In Sec. 4 we report experimental results while in Sec. 5 we draw conclusions arising from this work.

2 Motion Estimation Using Phase Correlation

Baseline PC operates on a pair of images (or cosited blocks) f_t and f_{t+1} of identical dimensions belonging to consecutive frames or fields of a moving sequence sampled at t, $t+1$. The estimation of motion relies on the detection of the maximum of the cross-correlation function between f_t and f_{t+1}. Since all functions involved are discrete, cross-correlation is circular and can be carried out as a multiplication in the frequency domain using fast implementations. The real-valued correlation surface is defined as

$$c_{t,t+1}(k,l) = F^{-1}\left(\frac{F_t^* F_{t+1}}{|F_t^* F_{t+1}|}\right) \tag{1}$$

where F_t and F_{t+1} are respectively the two-dimensional discrete Fourier transforms of f_t and f_{t+1}; F^{-1} denotes the inverse Fourier transform, and * denotes complex conjugate. The coordinates (k_m, l_m) of the maximum of the real-valued array $c_{t,t+1}$ can be used as an estimate of the horizontal and vertical components of motion at integer-pixel precision between f_t and f_{t+1} as follows:

Paper 05174LR received Oct. 4, 2005; revised manuscript received Dec. 9, 2005; accepted for publication Dec. 19, 2005; published online Feb. 28, 2006.
1017-9909/2006/15(1)/010501/3/$22.00 © 2006 SPIE and IS&T.

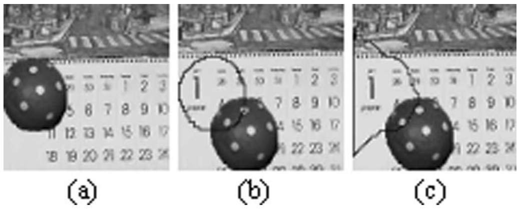

Fig. 1 Artificial example: (a) original, (b) showing accurate segmentation mask, and (c) showing inaccurate segmentation mask relative to the new object location.

$$(k_m, l_m) = \arg\max\{c_{t,t+1}(k,l)\}. \tag{2}$$

Subpixel accuracy of motion measurements is obtained by variable-separable fitting performed in the neighborhood of the maximum using one-dimensional quadratic functions.[6] Using the notation in Eq. (2) above, prototype functions are fitted to the triplets:

$$\{c_{t,t+1}(k_m-1,l_m), c_{t,t+1}(k_m,l_m), c_{t,t+1}(k_m+1,l_m)\} \text{ and}$$
$$\{c_{t,t+1}(k_m,l_m-1), c_{t,t+1}(k_m,l_m), c_{t,t+1}(k_m,l_m+1)\}. \tag{3}$$

The location of the maximum of the fitted function provides the required subpixel motion estimate (dx, dy). For example, fitting a parabolic function horizontally to the left-hand side of Eq. (3) yields a closed-form solution for the horizontal component of the motion estimate dx as follows:

$$dx = \frac{c_{t,t+1}(k_m+1,l_m) - c_{t,t+1}(k_m-1,l_m)}{2(2c_{t,t+1}(k_m,l_m) - c_{t,t+1}(k_m+1,l_m) - c_{t,t+1}(k_m-1,l_m))}. \tag{4}$$

The fractional part dy of the vertical component can be obtained in a similar way using the right-hand side of Eq. (3).

3 Arbitrary Shape Phase Correlation

The proposed arbitrary shape phase correlation (ASPC) scheme assumes that an object is available (i.e., obtained using segmentation) for an object of interest in the target (i.e., next) frame. The motion of this object needs to be estimated relative to a reference (i.e., current) frame. The ASPC algorithm consists of two main steps. First, SA-DFT is applied to pixel values inside the object both in the target and reference frames. Then, PC is applied using Eqs. (1)–(4) in order to obtain motion estimates of subpixel accuracy.

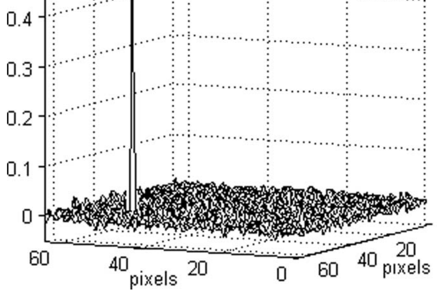

Fig. 2 Correlation surface using proposed ASPC (inaccurate mask).

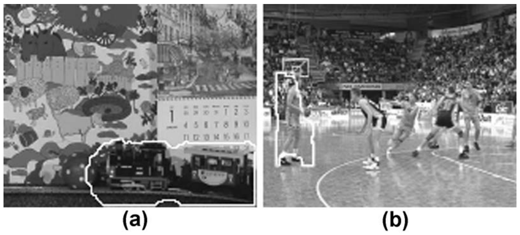

Fig. 3 Object masks for ‘train’ (a) and ‘player’ (b).

3.1 *Shape Adaptive Discrete Fourier Transform*

SA-DFT is based on the notion that the DFT of a vector can be interpreted as the transform of a periodic signal, one period of which is processed. The signal doesn’t need to be shifted, it suffices to extend it periodically and to compute the DFT starting from the beginning of a row/column.

We consider a one-dimensional vector of data samples $x(n)$ of size N with $n=0,1,\ldots,N$. We assume that N_S samples belong to the object of interest with $n=m$, $m+1,\ldots,N_s+m-1$ while the remaining samples are considered to belong to another object or the background. The application of SA-DFT consists of the following steps:

- Periodic extension of the N_S object samples to span the entire range of values $n=0,1,\ldots,N$, i.e., set $x'(n+kN_S)=x(n)$ with $k=\ldots,-2,-1,0,1,2,\ldots$, and $0<n+kN_S<N$.
- Computation of the N_S-point DFT.
- Scaling of the results. In this case the scaling factor is $1/\sqrt{N_S}$, which makes the 2-D transform orthogonal.[9]

The above can be extended to two dimensions in a variable-separable fashion (i.e., application of the above steps first horizontally then vertically). This allows the computation of SA-DFT for any arbitrary shaped region.

Given an arbitrary-shaped region in frame $t+1$ (i.e., the next frame) motion is estimated between the image data f_{t+1} inside this region and image data f_t inside a cosited region in frame t (i.e., the current frame). If F_t and F_{t+1} denote respectively the SA-DFT of f_t and f_{t+1}, the final step is to apply the phase correlation method using Eqs. (1)–(4).

4 Experimental Results

In our experiments we used the well-known broadcast resolution (720 × 576 pixels, 50 fields per second) MPEG test sequences ‘Mobcal’ and ‘Basketball.’ Only the luminance component was considered and to avoid complications due to interlacing, only even-parity field data were retained. We highlight the principles underlying our scheme using an artificial example. We manually segment the ‘ball’ object from a single frame of ‘Mobcal’ and superimpose it on the ‘calendar’ object from the same sequence without low-pass filtering object boundaries. We then create an artificial sequence by displacing the two objects relative to each other by progressively varying shifts of subpixel accuracy. These shifts are then used as ground truth. We consider two cases (shown in Fig. 1) corresponding respectively to accurate

Table 1 Average ground truth MSE performance comparison for artificial motion.

	'calendar'	
Artificial Motion	Accurate mask	Inaccurate mask
Phase Correlation	0.1363671	0.1363671
SAPC[7]	0.1316538	0.1302917
ASPC (proposed)	0.1277144	0.1268891

and inaccurate segmentation of the 'ball' object to reflect the real possibility of a segmentation algorithm providing variable quality results. In Fig. 2 we show a magnified portion of the correlation surface for the proposed ASPC scheme. While for conventional phase correlation two peaks would have emerged (each corresponding to the motion of each of the two objects 'ball' and 'calendar'), for ASPC only a single peak emerges corresponding to the motion of the 'ball' object. This becomes a critical consideration when the motion parameters of two (or more) objects are similar (but not identical) and hence the proximity of the corresponding peaks on the correlation surface may prevent the accurate extraction of the dominant motion parameters.

Next, we compare the proposed scheme with conventional PC and the shape adaptive phase correlation (SAPC) algorithm presented in Ref. 7, which is based on padding using the average intensity of the object of interest. In Table 1 we show the average MSE relative to the ground truth subpixel displacement parameters (computed over the total number of frames) obtained for both accurate and inaccurate segmentation scenarios of object 'calendar.' Overall, these results demonstrate (best cases underlined) that the proposed scheme achieves consistently higher accuracy compared to both alternative schemes.

Next we turn our attention to the estimation of real motion. We identify arbitrary-shaped objects of interest such as the 'train' object in 'Mobcal' and the 'player' object in 'Basketball.' The objects are determined manually and are shown in Fig. 3. We have deliberately allowed the objects to be inaccurate, i.e., not following closely the outline of the object under consideration or even omitting part(s) of it. This is a reflection of the fact that object definition will occasionally be inaccurate in a practical situation if this is obtained from automatic segmentation while this may not be true for manual segmentation. Performance is compared to conventional PC using a rectangular block (of the same size in pixels as the arbitrary-shaped object), which includes a large section of the object. Performance is assessed either using the motion-compensated prediction MSE (average values computed over the total number of frames are shown in Table 2) or the peak signal-to-noise ratio (PSNR) as a function of frame number (shown in Fig. 4) for the two sequences under consideration. It is worth noting that Table 2 includes a comparison with spatial-domain block-matching, which further underlines the potential of our method. Our results confirm the superiority of the proposed scheme, i.e., by as much as 4 dB compared to conventional phase correlation.

Table 2 Average motion-compensated prediction MSE performance comparison for real motion.

Real Motion	'train'	'player'
Block Matching	281.949546	313.995271
Phase Correlation	312.856572	266.257170
SAPC[7]	181.317667	202.621286
ASPC (proposed)	153.140966	201.528410

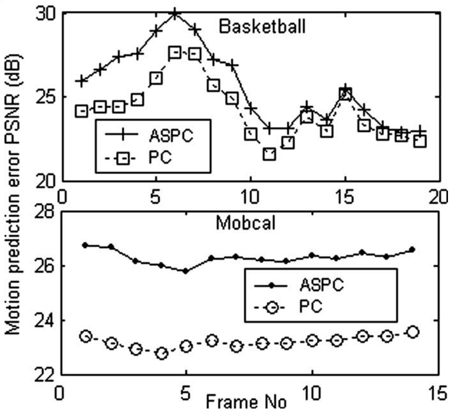

Fig. 4 PSNR vs frame number for conventional phase correlation and ASPC for objects 'player' (Basketball) and 'train' (Mobcal).

5 Conclusions

In this letter an arbitrary shape motion estimation algorithm based on phase correlation was presented. Owing to the fact that the scheme operates in the frequency domain it enjoys a high degree of computational efficiency and can be implemented by fast algorithms such as the FFT. Our approach avoids extrapolation and uses information only from moving objects of interest thereby yielding higher accuracy motion estimates. Our results have shown that the proposed method outperforms conventional phase correlation as well as shape adaptive phase correlation using extrapolation both for artificially-induced motion using manually extracted objects as well as actual interframe motion.

References

1. P. Kauff and K. Schüür, "Fast motion estimation for real-time shape-adaptive MPEG-4 encoding," *Proc. 2000 ACM Workshops on Multimedia* (Nov. 2000).
2. C. Stiller, "Object-based estimation of dense motion fields," *IEEE Trans. Image Process.* **6**(2), 234–250 (Feb. 1997).
3. A N. Delopoulos and A. G. Constantinides, "Object oriented motion and deformation estimation using composite segmentation," *Proc. IEEE-ICIP*, 2217–2220 (1995).
4. S. Panis and J. P. Cosmas, "Motion estimation with object based regularisation," *Electron. Lett.* **32**(9), 872–873 (May 1996).
5. J. J. Pearson, D. C. Hines, S. Goldsman, and C. D. Kuglin, "Video rate image correlation processor," *Proc. SPIE* **119**, 197–205 (1977).
6. G. A. Thomas, "Television motion measurement for DATV and other applications," *BBC Res. Dept. Rep.* (1987).
7. L. Hill and T. Vlachos, "Motion measurement using shape adaptive phase correlation," *Electron. Lett.* **37**(25), 1512–1513 (Dec. 2001).
8. R. Stasinski, "Shape adaptive discrete Fourier transform for coding of irregular image segments," *NORSIG-99* (Sept. 1999).
9. R. Stasinski and J. Konrad, "A new class of fast shape-adaptive orthogonal transforms and their application to region-based image compression," *IEEE Trans. Circuits Syst. Video Technol.* **9**, 16–34 (Feb. 1999).