USING GRADIENT CORRELATION FOR SUB-PIXEL MOTION ESTIMATION OF VIDEO SEQUENCES

V. Argyriou and T. Vlachos
University of Surrey, Guildford GU2 7XH, UK

ABSTRACT

A highly accurate and computationally efficient method is presented suitable for the estimation of motion in video sequences. The method is based on the maximization of the spatial gradient cross-correlation function, which is computed in the frequency domain and therefore can be implemented by fast transformation algorithms. We present enhancements to the baseline gradient-correlation algorithm, which further improve performance, especially in the presence of manually induced additive Gaussian noise. We also present a comparative performance analysis, which demonstrates that the proposed method outperforms the state-of-the-art in frequency-domain motion estimation, in the shape of phase correlation.

1. INTRODUCTION

Motion estimation is a key element of various video processing tasks such as standards conversion, frame-rate up-conversion and mosaicing. More importantly it is a critical component of video compression systems allowing redundancy reduction in the temporal domain.

Recently there has been a lot of interest in motion estimation techniques operating in the frequency domain. Perhaps the best-known method in this class is phase correlation [1], which has become one of the motion estimation methods of choice for a wide range of professional studio and broadcasting applications [2].

A key performance issue in motion estimation is sub-pixel accuracy. Theoretical and experimental analyses [11] have established that sub-pixel accuracy has a significant impact on motion compensated error performance. Sub-pixel accuracy mainly can be achieved through the use of bilinear interpolation. Interpolation methodologies are also applicable to frequency domain motion estimation methods. However, in this case, interpolation can only be applied as a separate pre-processing step and, in contrast to data domain approaches, cannot benefit from the use of fast algorithms. To circumvent this difficulty alternative approaches have been developed and these operate in synergy with the processing steps required for cyclic correlation.

In this paper we introduce a high-performance version of the gradient cross-correlation algorithm that was first presented in its basic form in [12]. The key advances introduced by this paper are the use of optimal filtering, the use of zero padding and windowing in the frequency-domain and the use of alternative curve fitting approaches on the gradient correlation surface. A significant component of this paper is a comprehensive comparative performance study between our method and a number of state-of-the-art sub-pixel motion estimation strategies that employ phase correlation.

This paper is organised as follows. In Section 2 the principle of motion estimation using gradient cross-correlation is reviewed and a number of novel performance-enhancing features are introduced. In Section 3 the key features of three state-of-the-art sub-pixel motion estimation algorithms are briefly summarised. In Section 4 experimental results are reported both in relation to the impact of the performance enhancing features to the baseline gradient-correlation method. In addition a comparative performance assessment is carried out with respect to three state-of-the-art frequency-domain motion estimation methods based on phase correlation. Finally, conclusions are drawn and final remarks are made.

2. SUB-PIXEL MOTION ESTIMATION USING GRADIENT CROSS-CORRELATION

In this section we review the principle of motion estimation using gradient cross-correlation and introduce the novel performance-enhancing features of our scheme. The use of image gradient information for the purpose of estimating motion is a well-established concept originating in very early work on image registration [5], [6] and today featuring in a number of popular algorithms of the non-matching variety [7].

Combining the natural advantages of good feature selection, offered by gradient-based methods with the speed and computational efficiency that typifies frequency domain processing, owing to the use of fast algorithms, is an idea that has only occasionally been explored in the literature [8], [9].

2.1 Computation of the image gradient

It is common ground that the computation of a spatial gradient of a discrete signal can only provide an approximation to the ideal differentiation operator whose frequency response is of the form $g(f) = j2\pi f$ for $|f| < f_s/2$, where f_s is the sampling frequency. In digital image processing common approximations rely on the use of forward or central pixel differencing, whose frequency response is of the form $g_c(f) = jf_s \sin(2\pi f / f_s)$. Using central differencing, at each pixel location of a given frame $f_t(x,y)$ discrete approximations to the horizontal and vertical gradients can be estimated as $g_t^h(x,y) = f_t(x+1,y) - f_t(x-1,y)$

and $g_t^v(x,y)=f_t(x,y+1)-f_t(x,y-1)$. The two terms are combined in a single complex representation of the form $g_t(x,y)=g_t^h(x,y)+jg_t^v(x,y)$, which retains magnitude and phase information at each pixel location.

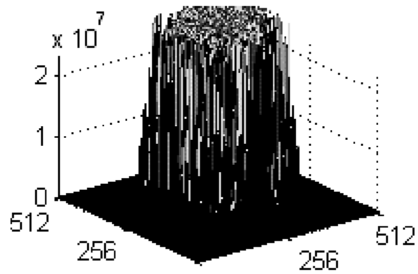

Fig. 1. A 3D representation of the magnitude after zero padding by a factor of 2

2.1.1 Optimal filtering

While central pixel differencing was the approach taken in our earlier work reported in [12], elementary filter design suggests that the addition of more terms can provide a better approximation. Work reported in [10] demonstrated that more sophisticated discrete approximations to the gradient are possible by using a filter optimization approach that favours a better spectral match at lower frequencies. This is an intuitively plausible approach given that a significant proportion of a typical image spectrum is clustered in a lower frequency range and decreases with a rate of $1/f$. In our work we adopt the approach advocated in [10] and compute horizontal and vertical gradients using the filter coefficients shown in Table I.

TABLE I

Coefficient for central difference estimators up to 3rd order

Order	c_{-3}	c_{-2}	c_{-1}	c_0	c_1	c_2	c_3
1			-1	0	-1		
2		1/12	-2/3	0	2/3	-1/12	
3	-1/60	3/20	-3/4	0	3/4	-3/20	1/60

2.2 Cross-correlation in the frequency-domain

For pairs of consecutive frames f_t and f_{t+1} discrete gradients g_t and g_{t+1} are respectively computed as above. The estimation of motion relies on the detection of the maximum of the cross-correlation function between g_t and g_{t+1}. Because all functions involved are discrete, cross-correlation is circular and for computational efficiency it can be carried out as a multiplication in the frequency domain using fast implementations:

$$c_{t,t+1}(k,l) = IFFT\left(G_t^{*}G_{t+1}\right) \qquad (1)$$

where G_t and G_{t+1} are respectively the two-dimensional discrete Fourier transforms of complex arrays g_t and g_{t+1} and * denotes complex conjugate. The coordinates (k_m, l_m) of the maximum of the real part of $c_{t,t+1}$ can be used as an estimate of the horizontal and vertical components of motion between f_t and f_{t+1} as follows:

$$(k_m, l_m) = \arg\max \operatorname{Re}\{c_{t,t+1}(k,l)\} \qquad (2)$$

where Re{} denotes the real part of complex array $c_{t,t+1}$.

2.3 Sub-pixel accuracy

Sub-pixel performance is a critical element of the proposed algorithm. With reference to our previously published work [12] we are introducing a number of important new features, which improve the accuracy of the motion estimates.

2.3.1 Zero-padding

The first such feature is zero-padding in the frequency domain. We denote by $G_{t,p}$ the two-dimensional symmetric extension of array G_t carried out by inserting zeros, where p is the zero-padding factor. This is illustrated in Fig. 1 for a factor of 2 and for a sample transformed data field. Circular cross-correlation of zero-padded arrays is carried out as in (1) above. In this case the estimate of the horizontal and vertical components of motion is given by $(k_m/p,\ l_m/p)$ where $k_m,\ l_m$ are defined as in (2) above.

2.3.2 Separable-variable fitting

Zero-padding as an interpolation mechanism has a number of limitations. It cannot provide estimates of floating point accuracy, only estimates whose accuracy is associated with negative powers of two. Also, it should be taken into account that high values of p carry an implementation penalty with regard to storage requirements as well computational complexity associated with FFT operations.

In contrast to zero-padding, separable-variable fitting on the gradient correlation surface used in [12], is free of many of the above constraints. In [12] we applied separable-variable fitting in the neighbourhood of the maximum using one-dimensional quadratic functions fitted to the triplet $\{c_{t,t+1}(k_m-1,l),\ c_{t,t+1}(k_m,l),\ c_{t,t+1}(k_m+1,l)\}$. In this paper we show that improved performance can be obtained by using a Gaussian function fitted to the same data triplet.

3. SUB-PIXEL MOTION ESTIMATION USING PHASE CORRELATION

In this section we consider three state-of-the-art sub-pixel motion estimation algorithms that have recently appeared in the literature. The methods considered are based on the phase correlation methodology and as such have similar computational complexity with our scheme.

3.1 Subspace Identification Extension (Hoge) [14]

This method is based on the observation that a 'noise-free' phase correlation matrix is a rank one, separable-variable matrix. As a consequence, for a 'noisy' phase correlation matrix, the sub-pixel motion estimation problem can be recast as finding the rank one approximation to that matrix.

This can be achieved by using singular value decomposition (SVD) followed by the identification of the left and right singular vectors. These vectors allow the construction of a set of normal equations, which can be solved for the required estimate. The author reports good results with registration of MRI scanned data.

3.2 Frequency-Domain Masking (Stone) [13]

After obtaining an integer-precision alignment of the input images the method takes steps towards alias cancellation by eliminating certain frequency components. Elimination is based on two criteria: (i) distance from highest peak and (ii) amplitude in relation to a threshold. The latter is dynamically determined and then a plane fitting operation on the frequencies that have survived the above two criteria yields the required motion estimates. The authors report good results with registration of aerial photographs.

3.3 Polyphase Decomposition (Foroosh) [4]

This method uses a model according to which images with subpixel shifts are obtained by integer pixel displacement on a higher resolution grid followed by subsampling. This assumption allows the analytic computation of the normalized cross-power spectrum of a pair of downsampled images as a polyphase decomposition of a filtered unit impulse. By approximating the Dirichlet kernel with a *sinc* function after taking the inverse Fourier transform of the cross-power spectrum, a closed-form solution for the subpixel shift estimate can be obtained.

4. EXPERIMENTAL RESULTS

In our experiments we used 2:1 downsampled versions of the well-known broadcast resolution MPEG test sequences 'Mobcal' and 'Basketball'. Both global and local motion estimation performance was assessed by applying motion compensation using the estimated motion parameters and computing either the Mean-Square Error (MSE) or the equivalent Peak-Signal-to-Noise Ratio (PSNR).

4.1 Optimal filtering

In Fig. 2 (a) we assess the impact of optimal gradient filters as discussed in 2.1.1. We compare the 3 filters and our results, obtained for the noise-contaminated 'Basketball' sequence (Gaussian noise 18dB), show that the optimal filter of size 5 performs significant better by achieving the best balance between faithful spectral matching at lower frequencies and rejection of unwanted windows of the spectrum. Similar results were obtained using the 'Mobcal' sequence and for other noise power values.

4.2 Gaussian fitting

In Fig. 2 (b) we assess performance for two different types of fitted surfaces as discussed in 2.3.2. We use variable-separable fitting independently in the horizontal and vertical dimensions. Our results show that Gaussian fitting consistently provides a higher level of accuracy.

4.3 Zero-padding

In Fig. 2 (c) we assess the impact of zero-padding in the frequency domain as discussed in 2.3.1. We use variable-separable Gaussian fitting and the optimal gradient filter of size 5 both of which yielded the best performances so far. Our results, obtained for the noise-contaminated 'Basketball' sequence, show that the use of zero padding consistently yields the best performance with respect to the non-zero padded baseline approach. Similar results were observed for the 'Mobcal' sequence and for other noise power values.

TABLE II
Average MSE over all measured even-parity field pairs

	Mobcal			Basketball		
	Global	**Local**	**Noise**	**Global**	**Local**	**Noise**
Foroosh	316.8	194.2	1185.7	269.9	229.1	1235.7
Stone	374.9	315.3	1285.9	334.9	342.9	1306.1
Hoge	276.9	236.4	1723.9	273.2	278.7	1783.5
Gradient	<u>234.9</u>	<u>145.3</u>	<u>1103.9</u>	<u>258.8</u>	<u>184.4</u>	<u>1176.9</u>

TABLE III
The entropy of each algorithm for block based motion estimation with and without noise

	Mobcal	Basketball	Mobcal	Basketball
	Local		**Local with Noise**	
Foroosh	3.5864299	2.8209298	4.414523	4.411116
Stone	4.8358054	4.3832988	5.305393	5.248677
Hoge	2.9309682	2.2519736	5.022979	5.322726
Gradient	<u>2.7860467</u>	<u>1.9111562</u>	<u>3.278353</u>	<u>3.537585</u>

4.4 Comparative performance assessment

We compare the performance of the optimized gradient correlation scheme above with the three state-of-the-art sub-pixel motion estimation algorithms based on phase correlation as discussed in Section 3, namely the work of Foroosh [4], Stone [13] and Hoge [14].

Global motion estimation
The first set of experiments aimed at measuring actual global motion parameters between even-parity fields for the two test sequences. Table II (Global column) summarizes the results obtained for the two sequences, by calculating the average MSE over all tested field pairs. Our results demonstrate that the optimized gradient-correlation method outperforms as far as the measurement of actual global scene motion is concerned achieving higher precision and a significantly smaller corresponding measurement error.

Local motion estimation
To measure local motion, the image is typically partitioned into blocks (32x32) and one set of motion parameters is obtained for all pixels in the same block. The results for the 'Mobcal' and 'Basketball' sequences are tabulated in Table II (Local column) and confirm the efficiency of the gradient

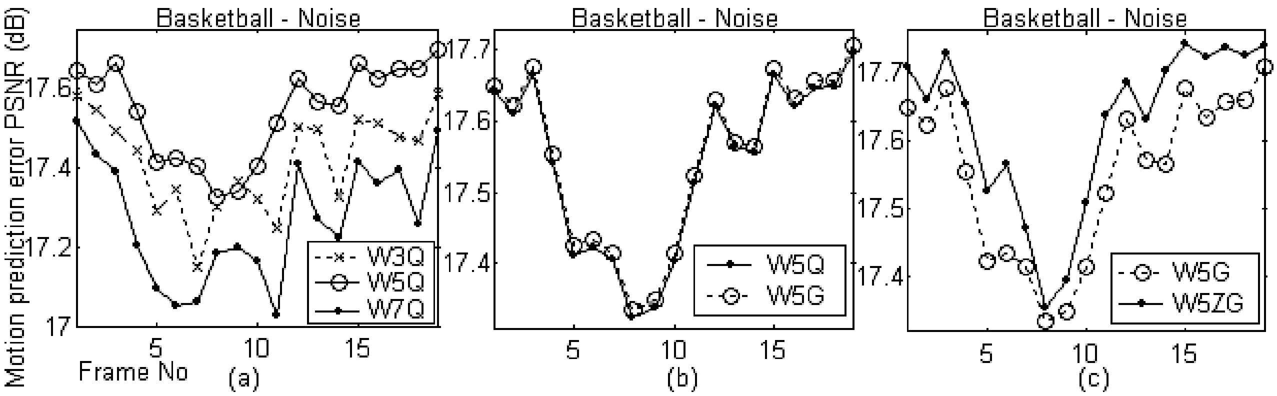

Fig. 2. (a) 1st (W3Q), 2nd (W5Q) and 3rd (W7Q) order filtering using Quadratic fitting (b) Gaussian (W5G) and Quadratic (W5Q) fitting for the 2nd order Gradient filter (c) With (W5ZG) and without (W5G) zero-padding using the 2nd order Gradient filter.

correlation method. Further experiments were performed, by manually inducing additive Gaussian noise of varying power. Figure 3 illustrates the average motion compensated prediction error computed over all available field blocks and for manually induced Gaussian noise of 18 dB. Our results demonstrate that gradient-based cross-correlation is substantially more immune to noise.

Another metric that was used to assess the performance of each algorithm for block based motion estimation is the entropy of the motion vectors. The results are shown in Table III and further confirm that the proposed algorithm consistently results higher levels of compression and motion vector coherence.

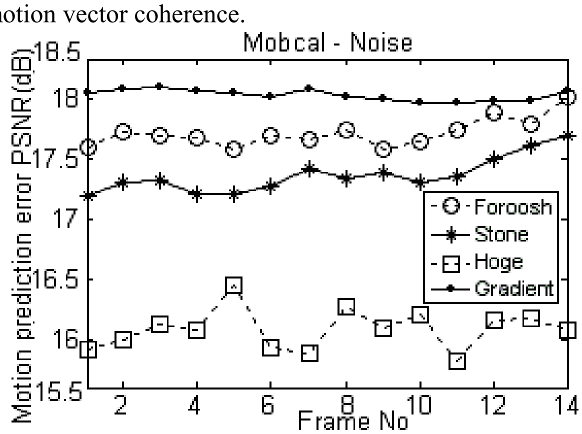

Fig. 3. PSNR versus frame number for local (block-based) motion compensated prediction with Gaussian noise

5. CONCLUSIONS

In this paper we presented a gradient-based cross-correlation technique for sub-pixel motion estimation in the frequency domain. By virtue of a number of enhancements, namely optimal gradient filtering, gaussian surface fitting and zero-padding, performance advantages over the baseline method were achieved in terms of measured PSNR of the motion compensated prediction error. In addition our method was shown to outperform the state-of-the-art in frequency-domain motion estimation using phase correlation. One of the most attractive features of the proposed scheme is that it enjoys a high degree of computational efficiency and can be implemented by fast transformation algorithms in the frequency domain.

6. REFERENCES

[1] J. J. Pearson, D. C. Hines, S. Goldsman, and C. D. Kuglin, "Video rate image correlation processor", Proc. SPIE , Vol. 119, Application of Digital Image Processing, 1977.

[2] G. A. Thomas, "Television motion measurement for DATV and other applications", BBC Res. Dept. Rep., No. 1987/11.

[3] R. Manduchi and G. A. Mian, "Accuracy analysis for correlation-based image registration algorithms", Proc. ISCAS, pp. 834-837, 1993.

[4] H. Foroosh, J. Zerubia and M. Berthod, "Extension of phase correlation to sub-pixel registration" , IEEE Trans. Image Processing, vol. 11, no. 3, pp. 188-200, 2002.

[5] D. I. Barnea and H. F. Silverman, "A class of algorithms for fast digital image registration", IEEE Trans. Computers, vol. C-21, no. 2, pp. 179-186, Feb. 1972.

[6] J. F. Andrus, C. W. Campbell and R. R. Jayroe, "Digital image registration method using boundary maps", IEEE Trans. Computers, vol. C-24, no. 9, pp. 935-940, Sep. 1975.

[7] S.S. Beauchemin and J.L. Barron, "The computation of optical flow", ACM Computing Surveys, vol. 27, no.3, pp.433-467,1996.

[8] C. Morandi, F. Piazza and R. Capancioni, "Digital image registration by phase correlation between boundary maps", IEE Proc., Vol. 134, pt. E, no. 2, pp. 101-104, Mar. 1987.

[9] A.J. Fitch, A. Kadyrov, W.J. Christmas and J. Kittler, Orientation Correlation, Proc. BMVC, vol 1, pp. 133-142, 2002.

[10] W. J. Christmas, "Spatial filtering requirements for gradient-based optical flow measurement", IEEE Trans. Image Processing, vol. 9, no. 10, pp. 1817-1820, Oct. 2000.

[11] B. Girod, "Motion-compensating prediction with fractional-pel accuracy", IEEE Trans. Comm., vol.41, no.4, pp. 604, Apr.'93.

[12] V. Argyriou and T. Vlachos, "Estimation of sub-pixel motion using gradient cross-correlation", Elec. Lett., Vol. 39, No. 13, 980-982, June 2003.

[13] H. S. Stone, M. T. Orchard, E-C. Chang and S. A. Martucci, "A fast direct Fourier-based algorithm for subpixel registration of images", IEEE Trans. Geo. and Rem. Sensing, vol. 39, no. 10, pp. 2235-2243, Oct. 2001.

[14] W. S. Hoge, "A subspace identification extension to the phase correlation method", IEEE Trans. Med. Imag., vol. 22, no. 2, pp. 277-280, Feb. 2003.