

MOTION ESTIMATION USING HYPERCOMPLEX CORRELATION IN THE WAVELET DOMAIN

V. Argyriou and T. Vlachos

Centre for Vision, Speech and Signal Processing
School of Electronics and Physical Sciences
University of Surrey, Guildford GU2 7XH, United Kingdom.
E-mail: v.argyriou@surrey.ac.uk , t.vlachos@surrey.ac.uk
Phone : +44 1483 686032, +44 1483 689854
Fax : +44 1483 68603

ABSTRACT

We present a novel frequency-domain motion estimation technique, which employs hypercomplex correlation in the wavelet domain. Our method involves wavelet decomposition followed by cross-correlation in the frequency domain using the discrete quaternion Fourier transform. Experiments using both artificially induced motion and actual scene motion demonstrate that the proposed method outperforms the state-of-the-art in frequency-domain motion estimation, in the shape of phase correlation, in terms of sub-pixel accuracy for a range of test material and motion scenarios.

1. INTRODUCTION

Motion estimation is a critical component of various video analysis and processing systems including compression where it allows redundancy reduction in the temporal domain. International standards for video communications such as MPEG-1/2/4 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 compensated prediction in such algorithms is based on regular block-based partitions of source frames while the emerging H.264 standard provides additional flexibility in that respect.

One of the main motivations for this work has been the recent interest in motion estimation techniques operating in the frequency domain. These are commonly based on the principle of cyclic correlation and offer well-documented advantages in terms of computational efficiency due to the employment of fast algorithms [1]. Perhaps the best known method in this class is phase correlation [2] which has become one of the motion estimation methods of choice for a wide range of professional studio and broadcasting applications [3]. In addition to computational efficiency, phase correlation 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 sub-pixel accuracy [3]. On the other hand, image analysis in the wavelet domain has emerged as a key tool in various applications such as compression, offering improved efficiency as well as useful functionalities such as scalability. Our work provides a natural point of convergence for the above two key concepts combining the implementation efficiency of correlation operations carried out in the frequency domain with the good signal decomposition properties that have made the wavelet transform one of the methods of choice for many image analysis problems.

Motion estimation in the wavelet domain is attractive for a number of reasons; it is less sensitive to photometric distortions; the directional nature of information inherent in certain wavelet decompositions can provide useful references for measuring motion; the multi-resolution nature of the wavelet decomposition facilitates the hierarchical estimation of motion [4]. Furthermore, the extraction of high-pass features such as edges which emerge in the detail subbands of the wavelet transform and the related fact that motion prediction residuals depend on the gradient magnitude [5] are likely to enhance the potential for high motion measurement accuracy.

For those reasons motion estimation based on the wavelet transform is by no means a novel concept and a significant amount of relevant work can be found in the literature. It is worth mentioning work by Magarey and Kingsbury [6], Cai and Adjouadi [7], Li [4], Zhen et al. [8], Mujica et al. [9], Cheong et al. [10] and Yu-Te Wu [11] to name but a few. Nevertheless none of the schemes reported in the literature operates in the wavelet domain using hyper-complex correlation, which is one of the key features of our work.

In this paper we propose a wavelet based motion estimation method for obtaining sub-pixel estimates of translational interframe motion using hyper-complex cross correlation. Our method includes two stages, wavelet decomposition and quaternion correlation. The wavelet decomposition stages provides a multi-band description of a frame extracting useful features for motion estimation purposes. In the proposed technique, we apply discrete wavelet

decomposition (DWT) to transform the successive frames into multi-band representation, and then we apply quaternion correlation filters to perform multi-band processing jointly.

This paper is organised as follows. In Section 3 we briefly review the principles underlying sub-pixel motion estimation using phase correlation. In Section 4 a short summary of quaternion algebra is presented and in Section 5 image decomposition in the wavelet domain is briefly reviewed. In Section 6 we formulate our wavelet approach, which is based on quaternion correlation. In Section 7 we present experimental results while in Section 8 we draw conclusions arising from this work.

2. MOTION ESTIMATION USING PHASE CORRELATION

Baseline phase correlation (PC) operates on a pair of images (or co-sited 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 correlation surface is defined as

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

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 co-ordinates (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:

$$(k_m, l_m) = \arg \max \text{Re} \left\{ c_{t,t+1}(k,l) \right\} \quad (2)$$

Sub-pixel accuracy of motion measurements is obtained by separable-variable fitting performed in the neighbourhood of the maximum using one-dimensional quadratic functions [3]. Using the notation in (2) above, prototype functions are fitted to the triplets:

$$\left\{ c_{t,t+1}(k_m-1, l_m), c_{t,t+1}(k_m, l_m), c_{t,t+1}(k_m+1, l_m) \right\} \quad (3)$$

$$\left\{ c_{t,t+1}(k_m, l_m-1), c_{t,t+1}(k_m, l_m), c_{t,t+1}(k_m, l_m+1) \right\} \quad (4)$$

The location of the maximum of the fitted function provides the required sub-pixel motion estimate (dx, dy) . For example fitting a parabolic function horizontally to (3) yields a

closed-form solution for the horizontal component of the motion estimate dx as follows:

$$dx = \frac{c_{t,t+1}(k_m+1, l_m) - c_{t,t+1}(k_m-1, l_m)}{2(c_{t,t+1}(k_m, l_m) - c_{t,t+1}(k_m+1, l_m) - c_{t,t+1}(k_m-1, l_m))} \quad (5)$$

The fractional part dy of the vertical component can be obtained in a similar way using (4).

3. QUATERNION ALGEBRA

The quaternions (also referred as hypercomplex numbers) were introduced by Hamilton in 1843 [12] and provide a generalization of complex number notation. While a complex number has two components (the real and the imaginary part) the quaternion has four components, one real and three imaginary parts: $q = q_r + q_i \cdot i + q_j \cdot j + q_k \cdot k$ where q_r , q_i , q_j , and q_k are real and i , j , and k are complex operators which obey the following rules $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$, $jk = -kj = i$ and $ki = -ik = j$

Based on the concept of quaternion, the quaternion Fourier transform (QFT) has been introduced recently. There are many different types of QFT. The earliest definition of QFT is the two-side form as follows

$$H_1(w, u) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\mu_1 w x} \cdot h(x, y) \cdot e^{-\mu_2 u y} \cdot dx dy \quad (6)$$

where w and u are quaternion frequencies and μ_1 and μ_2 are two unit pure quaternions that are orthogonal to each other. Except for (6) there are also other types of QFT. Recently, the left and the right side forms of QFT were introduced

$$H_2(w, u) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\mu_1 (w x + u y)} \cdot h(x, y) \cdot dx dy \quad (7)$$

$$H_3(w, u) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot e^{-\mu_1 (w x + u y)} \cdot dx dy \quad (8)$$

where μ_i is any unit pure quaternion. Also, there are at least three types of discrete quaternion Fourier transform (DQFT) and the two-side DQFT is below.

$$H_3(p, s) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{-\mu_1 2\pi(p m / M)} \cdot h(m, n) \cdot e^{-\mu_2 2\pi(s n / N)} \quad (9)$$

The cross-correlation of two images $f(m, n)$ and $g(m, n)$ was defined in [13]:

$$r(m, n) = \sum_{q=0}^{M-1} \sum_{p=0}^{N-1} f(q, p) \overline{g(q-m, p-n)} \quad (10)$$

where \bar{g} denotes complex conjugate and the shift operation on $g(m, n)$ is carried out cyclically using modulo arithmetic for the subtractions. If the mean value of each image is first subtracted, the cross-covariance is obtained. A hypercomplex generalization of the Wiener-Khintchine theorem is as follows:

$$cr(m, n) = F^{-1} \{ F(v, u) \overline{G_{\parallel}(v, u)} \} + F \{ F(v, u) \overline{G_{\perp}(v, u)} \} \quad (11)$$

where v and u are quaternion frequencies, \bar{G} denotes quaternion conjugate and $G(v, u) = F\{g(m, n)\}$ and $G_{\parallel}(v, u) \parallel \mu$ and $G_{\perp}(v, u) \perp \mu$ and $G(v, u) = G_{\parallel}(v, u) + G_{\perp}(v, u)$.

For example using the right side quaternion Fourier transform

$$H_3(p, s) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} h(m, n) \cdot e^{-\mu_1 2\pi((pm/M) + (sn/N))} \quad (12)$$

the cross-correlation is as follows:

$$cr(m, n) = F^{-R} \{ F^R(v, u) \overline{G_{\parallel}^R(v, u)} \} + F^R \{ F^R(v, u) \overline{G_{\perp}^R(v, u)} \} \quad (13)$$

where F^R (G^R) and F^{-R} denote the right side forward and inverse Fourier transform respectively.

4. WAVELET DECOMPOSITION

The wavelet transform decomposes the original signal into different scales and resolutions, providing more insight of the joint space-frequency characteristics of the original signal. More formally the wavelet transform is written as:

$$\gamma(x, y) = \sum_j \sum_k \psi_{s, \tau}(x, y) f(k-x, j-y) \quad (14)$$

This equation shows how a function f is decomposed into a set of basis functions $\psi_{s, \tau}$ called the wavelets. The discrete wavelet decomposition (DWT) can be viewed in a tree structure, where the original signal is passed through a low-pass filter and a high-pass filter, and then downsampled by a factor of 2 to obtain respectively the low-frequency and high-frequency components of the original signal. The decomposition can be iteratively applied to the low-frequency band to generate a wavelet decomposition tree. For 2-D signals such as images a 4-quadrant decomposition can be similarly obtained corresponding to LL, HL, LH and

HH frequency bands respectively, (Figure 1). Finally it should be noted that in our work best results were obtained using the well-known Daubechies family of wavelets [14].



Figure 1. Original image and 1-level wavelet transform without downsampling.

5. WAVELET CROSS-CORRELATION

In this paper we introduce a method that combines wavelet decomposition and quaternion correlation filters for fast and low complexity motion estimation. The proposed algorithm consists of two main stages.

In the first stage two successive frames f and g of a video sequence are selected and a conventional wavelet decomposition to four quadrants (LL, HL, LH and HH) is applied to both of them. We do not perform downsampling to avoid aliasing whose presence reduces the motion estimation accuracy. The resulting quadrants are encoded into two 2-D quaternion arrays as follows:

$$q_f = q_f^{LL} + q_f^{HL} \cdot i + q_f^{LH} \cdot j + q_f^{HH} \cdot k \quad (15)$$

and

$$q_g = q_g^{LL} + q_g^{HL} \cdot i + q_g^{LH} \cdot j + q_g^{HH} \cdot k \quad (16)$$

In the second stage the resulting hypercomplex images q_f and q_g are segmented into non-overlapping blocks. For each pair of blocks quaternion cross-correlation is performed using (13) or using the left side Fourier transform [15] followed by hypercomplex inverse right side Fourier transform

$$cr(m, n) = F^{-R} \{ CR(m, n) \}$$

where

$$CR(m, n) = F^R(v, u) \overline{G_{\parallel}^L(v, u)} + F^{-R}(v, u) \overline{G_{\perp}^L(v, u)} \quad (17)$$

is the correlation of the images $f(m, n)$ and $g(m, n)$. The real (scalar) part of the resulting 2-D quaternion array c_{fg} is selected. At the resulting correlation surface (scalar part), the location of the maximum corresponds to the motion parameters between the corresponding blocks. In order to

obtain subpixel accuracy the method suggested in Section 2 is be used in combination with the equation (5).

6. EXPERIMENTAL RESULTS

To illustrate the efficiency of the proposed scheme we compare its performance with the phase correlation technique. In our experiments we used the well-known broadcast resolution MPEG test sequences ‘Mobcal’, ‘Basketball’ and ‘Foreman’. Only the luminance component was considered and to avoid complications due to interlacing only even-parity field data were retained. Both global and local (block-based) motion estimation performance was assessed. In the first case the estimation area was limited to the central 256x256 pixels of the retained field while in the second case the central 320x256 pixel area was retained and then further partitioned to blocks.

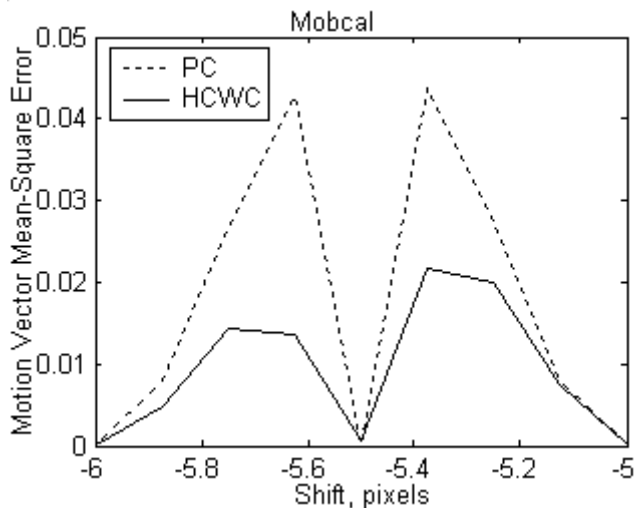


Figure 2. Error performance comparison for manually induced motion for a frame of sequence ‘Mobcal’.

Global Motion

First, estimation accuracy was assessed when global motion was manually induced and hence known a priori. Sub-pixel motion was simulated using bi-linear interpolation to displace test images ‘Mobcal’ and ‘Basketball’ at desired positions. Various positive as well as negative sub-pixel shifts were used whose accuracy was limited to 1/8 of a pixel, (Figure 2). Table I summarises the results obtained for the two sequences, by calculating the average MSE of the estimated motion vectors over all tested field pairs.

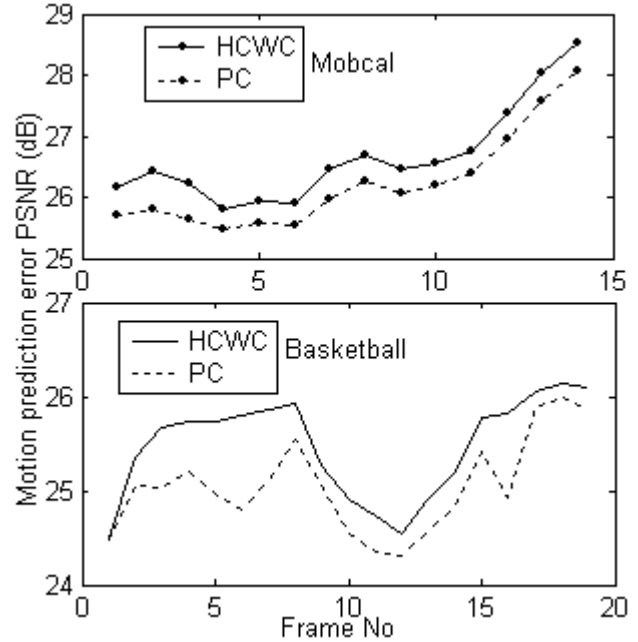


Figure 3. PSNR versus frame number for local (block-based) motion compensated prediction.

Local Motion

To measure local motion, the image is typically partitioned into non-overlapping blocks and one set of motion parameters (i.e. a motion vector) is obtained for all pixels in the same block. Both phase correlation and the proposed technique, were applied to 32x32 pixel blocks of two successive even-parity fields of the ‘Mobcal’ sequence. MSE differences between the two competing algorithms as a function of block number are shown in Figure 3 (upper). Similar experiments were performed for the ‘Basketball’ and ‘Foreman’ sequences. The results are shown in Figure 3 and further confirm that the proposed algorithm consistently outperforms phase correlation. Tables I summarizes the results obtained for the three sequences, by calculating the average MSE over all tested field pairs. A more comprehensive performance assessment in terms of rate-distortion characteristics is shown in Figure 4. Here the motion-compensated prediction MSE is plotted as a function of zero-order two-dimensional vector field entropy.

TABLE I

Manually induced motion and corresponding squared error in pixels and Average MSE over all measured even-parity field pairs for phase correlation and the proposed method.

	Artificial Motion		Real Motion		
	Mobcal	Basket	Mobcal	Basket	Foreman
PC	0.1565	0.1577	157.294	204.525	116.786
HCWC	0.0826	0.0916	142.293	185.764	93.4392

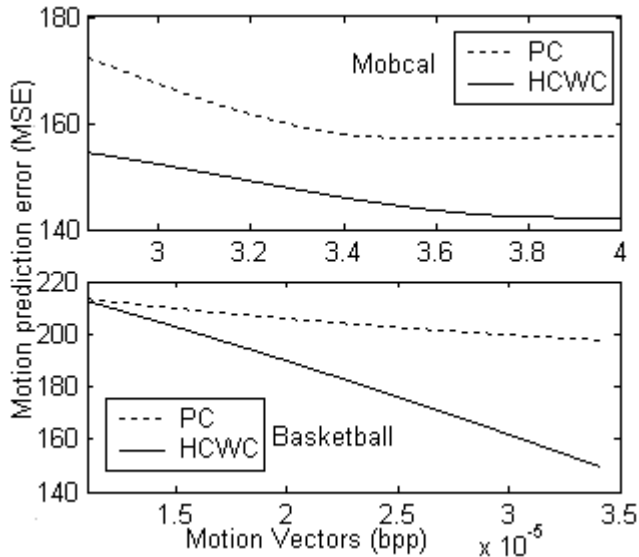


Figure 4. Rate-distortion plots for ‘Mobcal’ and ‘Basketball’.

Furthermore, we perform a short computational complexity comparison based on [15] and [16] in terms of the number of real multiplications required by each scheme. For PC and HCWC $M \times N$ is the block size (or the frame size). Consequently, the total amount of real number multiplications required is $MN \log_2 MN$ and $3MN \log_2 MN$ for phase correlation and wavelet-based hypercomplex cross-correlation, respectively. It should be noted that for block-based methods that operate in the pixel domain (e.g. full search block matching) with block size $M \times N$ and search area size $2M \times 2N$, the amount of real number multiplications required is $4M^2N^2$ suggesting that the computational complexity is significantly higher and hence not attractive for real time applications.

Finally, it should be noted that we have not carried out any comparisons with other wavelet-based motion estimation methods (such as those mentioned in the Introduction) due to the fact that these either operate in the pixel domain and involve significantly higher computational costs or are engineered to yield dense motion vector fields and hence are not directly comparable to the proposed method.

7. CONCLUSIONS

In this paper a wavelet-based hypercomplex cross-correlation technique for sub-pixel motion estimation in the frequency domain was presented. The proposed method takes full account of all the subimages obtained by a wavelet decomposition ensuring the selection of useful and reliable image features. Wavelet cross-correlation yields very accurate sub-pixel motion estimates for a variety of test material and motion scenarios and outperforms phase correlation, which is the current motion estimation method of choice in the frequency domain. 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.

8. REFERENCES

- [1] A.J. Fitch, A. Kadyrov, W.J. Christmas and J. Kittler, “Orientation Correlation”, Proc. BMVC, vol 1, pp 133-142, 2002.
- [2] J. J. Pearson, D. C. Hines, S. Goldman, and C. D. Kluging, “Video rate image correlation processor”, Proc. SPIE, Vol. 119, Application of Digital Image Processing, 1977.
- [3] G. A. Thomas, “Television motion measurement for DATV and other applications”, BBC Res. Dept. Rep., No. 1987/11.
- [4] X. Li, “Fast and Efficient Block Motion Estimation in the Wavelet Space”, Proceedings. ITCC 2003. International Conference on 28-30 Pages: 488 – 494, April 2003.
- [5] B. Tao and M.T. Orchard, “Gradient-Based Residual Variance Modeling and Its Applications to Motion-Compensated Video Coding”, Image Processing, IEEE Transactions on Vol. 10, Issue 1, Pages: 24 – 35, Jan. 2001.
- [6] J. Magarey and N. Kingsbury, “Motion Estimation Using a Complex-Valued Wavelet Transform”, Signal Processing, IEEE Transactions on Vol. 46, Issue 4, Pages: 1069 – 1084, April 1998
- [7] W. Cai and M. Adjouadi, “Wavelet-Domain Shift Invariant Motion Estimation and Compensation”, IMVSP, 2004. Proceedings of 2004 International Symposium on 20-22 Pages: 49 – 52, Oct. 2004
- [8] J. Zhen, J. Yifeng and Z. Jihong, “Automatic Registration Algorithm Based on Wavelet Transform”, WCCC-ICSP 2000. 5th International Conference on Vol. 2, 21-25 Pages: 979 – 982, Aug. 2000.
- [9] F.A. Mujica, J.P. Leduc and M.J. Smith, “A New Motion Parameter Estimation Algorithm Based on the Continuous Wavelet Transform”, Image Processing, IEEE Transactions on Vol. 9, Issue 5, Pages: 873–888, May 2000
- [10] C.K. Cheong, K. Aizawa, T. Saito and M. Hatori, “Motion Estimation With Wavelet Transform and the Application to Motion Compensated Interpolation”, ICASSP, IEEE International Conference on Vol.5, 27-30 Pages: 217–220, April 1993
- [11] Y. Wu, T. Kanade, J. Cohn and C. Li, “Optical Flow Estimation Using Wavelet Motion Model”, Computer Vision, 1998. Sixth International Conference Pages: 992 – 998, 4-7 Jan. 1998
- [12] W.R. Hamilton, Elements of Quaternions, London, U.K.: Longman 1866.
- [13] S. J. Sangwine and T. A. Ell, “Hypercomplex auto- and cross-correlation of color images”, in IEEE International Conference on Image Processing (ICIP99), Kobe, Japan, October 24-28 1999, pp. 319-322
- [14] I. Daubechies, “Ten lectures on wavelets”, CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, 1992.
- [15] C. E. Moxey, S. J. Sangwine and T. A. Ell, “Hypercomplex Correlation Techniques for Vector Images” Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions Vol. 51, Issue 7, Pages: 1941–1953, July 2003
- [16] S. Pei, J.J Ding and J. Chang, “Efficient Implementation of Quaternion Fourier Transform, Convolution, and Correlation by 2-D Complex FFT”, Signal Processing, IEEE Transactions Vol. 49, Issue 11, Pages: 2783 – 2797, Nov. 2001