QUAD-TREE MOTION ESTIMATION IN THE FREQUENCY DOMAIN

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

ABSTRACT

We propose a quad-tree scheme for obtaining sub-pixel estimates of interframe motion in the frequency domain. Our scheme is based on phase correlation and uses motion compensated prediction error to control the partition of a parent block to four children quadrants. This criterion guarantees a monotonic decrease of the motion compensated prediction error with an increasing number of iterations making our scheme suitable for embedded coding applications. Our results show that our scheme provides a better level of adaptation to scene contents and outperforms fixed block size phase correlation in terms of total motion compensated prediction error for the same number of motion vectors and also in terms of number of motion vectors for the same level of motion compensated prediction error.

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 employ motion compensated prediction 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 advantages of such partitions is that they require little or no overhead information. On the other hand they provide little or no adaptation to picture content.

Motion estimation using non-regular, non-block based partitions has been an active area of research over the past few years culminating in a wide range of proposals including mesh-based, polygon-based and object-based schemes. Such schemes achieve a higher degree of adaptation to picture contents and are better suited to portray the evolution of moving objects in a dynamic scene. Unfortunately they also require considerable overhead information as well as computational complexity.

Motion estimation based on quad-tree partitions achieves a good balance between a degree of adaptation to motion content on the one hand and low-complexity, low-overhead implementation on the other. Indeed quad-tree image partitions can be described very economically while offering a wealth of useful features for practical implementations. Quad-tree hierarchies also allow a natural and efficient portrayal of motion at different levels, from global to local.

Quad-tree motion estimation is by no means a novel concept and a significant amount of relevant work can be found in the literature. An exhaustive review would be outside the scope of this paper but it is worth mentioning work by Strobach [1], Nicolas and Labit [2], Jensen and Anastassiou [3], Banham et al. [4], Sullivan and Baker [5], Seferidis and Ghanbari [6], Lee [7], Schuster and Katsaggelos [8], Packwood et al. [9], Tredwell and Evans [10] and Cordell and Clarke [11] to name but a few. Nevertheless none of the schemes reported in the literature operates in the frequency domain, which is one of the key features of our work.

One of the main motivations for this paper 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 [12]. Perhaps the best-known method in this class is phase correlation [13], which has become one of the motion estimation methods of choice for a wide range of professional studio and broadcasting applications [14]. In addition, 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 [14]. In this paper we propose a quad-tree motion estimation method for obtaining sub-pixel estimates of interframe motion using phase correlation. In Section 2 we briefly review the principles underlying sub-pixel motion estimation using phase correlation. In Section 3 we formulate our quad-tree approach, which includes addressing the problem of correlating blocks of dissimilar sizes. In Section 4 we present experimental results while in Section 5 we draw conclusions arising from this work.

2. MOTION ESTIMATION USING PHASE CORRELATION

Baseline phase correlation operates on a pair of images or, more commonly a pair of co-sited rectangular 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 it can be carried out as a multiplication in frequency domain using fast implementations. The correlation surface is defined as

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

0-7803-9332-5/05/$20.00 ©2005 IEEE

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:

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

where Re{ } is the real part of the complex phase correlation surface array.

2.1 Sub-pixel accuracy

Sub-pixel accuracy of motion measurements is obtained by separable-variable fitting performed in the neighbourhood of the maximum using one-dimensional quadratic functions [14]. Using the notation in (2) above, 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 yields a closed-form solution for the horizontal component of the motion estimate *dx* as follows:

$$dx = \frac{c_{t,t+1}(k_m+1, l_m) - c_{t,t+1}(k_m-1, l_m)}{2(2c_{t,t+1}(k_m, l_m) - c_{t,t+1}(k_m+1, l_m) - c_{t,t+1}(k_m-1, l_m)} \quad (3)$$

The fractional part *dy* of the vertical component can be obtained in a similar way.

3. QUAD-TREE MOTION ESTIMATION

The starting point for a quad-tree partition of frame f_t involves a global phase correlation operation between f_t and f_{t+1}. Here we use the conventional notion of f_{t+1} being the target frame (the frame whose motion we seek to estimate) and f_t being the reference frame (the frame just been decoded in a moving sequence). An artificial example with 3 moving objects and the resulting correlation surface $c_{t,t+1}$ obtained by (1) and showing the corresponding 3 peaks is shown in Figure 1 below.

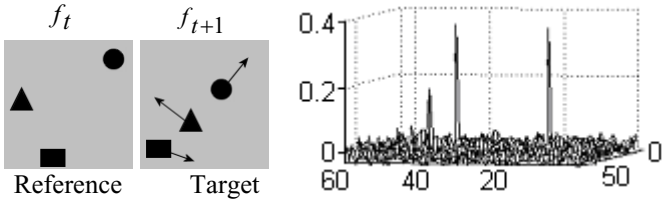

Fig. 1. Artificial example with 3 objects and corresponding correlation surface.

3.1 Partition criterion

The next step is to examine whether target frame f_{t+1} should be partitioned or not to four quadrants $\{ f_{t+1}^i \}$ $i = 1, \cdots, 4$. The partition criterion is based on the translational motion parameters $(k_m + dx, l_m + dy)$ obtained by (2) and (3) above to form a motion compensated prediction $\hat{f}_{t+1}$ of f_{t+1} using f_t i.e.

$$\hat{f}_{t+1}(x, y) = f_t(x + k_m + dx,\ y + l_m + dy) \quad (4)$$

where (x, y) are pixel locations.

A motion compensated prediction is also formed for each of the four quadrants as follows:

$$\hat{f}_{t+1}^i(x, y) = f_t^i\left(x + k_m^i + dx^i,\ y + l_m^i + dy^i\right) \quad i = 1, \cdots, 4 \quad (5)$$

where $\left(k_m^i + dx^i,\ l_m^i + dy^i\right)$ are sub-pixel accurate motion parameters for each of the four quadrants.

Finally, if the mean squared motion compensated prediction error (MSE) of the split is lower than the MSE error before the split, which is equivalent to the following holding true:

$$\sum_{i=1}^{4} \sum_{x,y} \left[\hat{f}_{t+1}^i(x, y) - f_{t+1}^i(x, y) \right]^2 < \sum_{x,y} \left[\hat{f}_{t+1}(x, y) - f_{t+1}(x, y) \right]^2 \quad (6)$$

then the target image/block is allowed to split into four quadrants. This criterion guarantees a monotonic decrease of the motion compensated prediction error with an increasing number of iterations making our scheme suitable for progressive/embedded coding applications.

3.2 Correlating unequal size image blocks

Assuming that the target frame has been partitioned as above, motion parameters need to be estimated for each of the resulting four quadrants. One obvious course of action would have been to partition the reference frame f_t in a similar manner i.e. four quadrants $\{ f_t^i \}$ $i = 1, \cdots, 4$ and perform phase correlation between co-sited quadrants f_t^i and f_{t+1}^i. However, this would have restricted the range of motion parameters accordingly i.e. inside the i-th quadrant. A consequence of this would be that the motion parameters of a fast moving object traversing quadrant boundaries during a single frame period would be impossible to estimate. For this reason it is preferable to correlate quadrant f_{t+1}^i in the target frame with the entire reference frame f_t, Fig 2 (a). Nevertheless this has the obvious disadvantage of requiring a phase correlation operation to be performed between two images of unequal sizes i.e. the reference being four times larger than the target. One straightforward way to go round this problem would be to increase the size of the target image by a factor of 2 in each dimension by extrapolative padding i.e. by symmetric insertion zeros or mid-grey values to the unknown pixel locations outside f_{t+1}^i until the latter assumes equal dimensions to f_t. In our work we have opted to carry out this operation in the frequency domain by interpolative upsampling of F_{t+1}^i i.e. the Fourier transform of f_{t+1}^i. We have used linear interpolation to obtain F_{t+1}^i whose

dimensions are now identical to F_t hence allowing (1) to be possible. Interpolative upsampling in the frequency domain has the obvious practical advantage that the Fourier transform of the target image requires far less computations than otherwise.

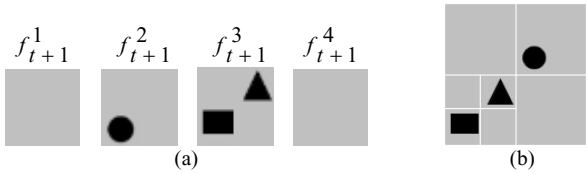

Fig.2. (a) Quadrants of target frame. (b) Quad-tree structure.

3.3 Further iterations

The algorithm proceeds in an iterative fashion to determine whether or not f^i_{t+1} will be partitioned any further. We will assume that each subsequent partition is denoted by an additional index i.e. $\left\{f^{i,j,k,\cdots}_{t+1}\right\}$ $i, j, k, \cdots = 1, \cdots, 4$ so that the number of indices is equal to the levels of partitions performed. For example $\left\{f^{i,j}_{t+1}\right\}$ denotes the *j*-th quadrant at the 2nd level of partition located inside the *i*-th quadrant at the 1st level of partition. Due to the fact that some quadrants may occasionally fail the criterion and hence remain undivided not all $\left\{f^{i,j,k,\cdots}_{t+1}\right\}$ will exist for all possible combinations of i, j, k, …= 1,…4, Fig. 2 (b).

Using the above notation, during a second iteration of the algorithm quadrant f^i_{t+1} will be further partitioned to four sub-quadrants $\left\{f^{i,j}_{t+1}\right\}$ $j = 1, \cdots, 4$. Reference frame f_t is also partitioned to four quadrants $\left\{f^i_t\right\}$ $i = 1, \cdots, 4$ and target sub-quadrant $f^{i,j}_{t+1}$ is phase-correlated with its co-sited parent quadrant f^i_t according to 3.2 above. It should be noted that the minimum block size is 16x16 pixels since phase-correlation yields unreliable estimates for smaller blocks.

4. EXPERIMENTAL RESULTS

In our experiments we used 2:1 downsampled versions of the well-known broadcast resolution (720x576) MPEG test sequences 'Mobcal' and 'Basketball'. We discarded even parity fields to avoid complications due to interlacing. The resolution was further restricted to the central 256x256 pixels to facilitate the computation of the Fourier transform and for easier implementation of the quad-tree partition, without any loss of generality, because phase correlation can be applied either on square or on rectangular blocks.

The proposed algorithm is compared with fixed block size phase correlation that uses 16x16 pixel blocks and 256 motion vectors per frame. A visual comparison demonstrating adaptivity to scene contents is shown in Figure 3. The partition criterion of the quad-tree scheme was adjusted to obtain a similar MSE to the fixed block size scheme. The results are shown in Figure 4 where it can be seen that our scheme uses far less motion vectors for the same amount of prediction error.

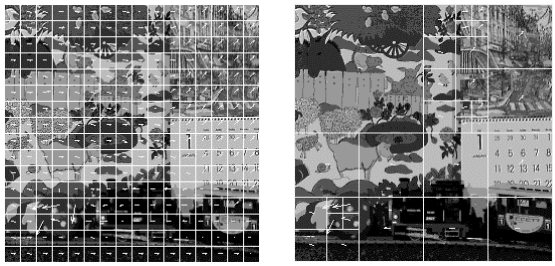

Fig.3. Vector field comparison between fixed-block size and quad-tree phase correlation.

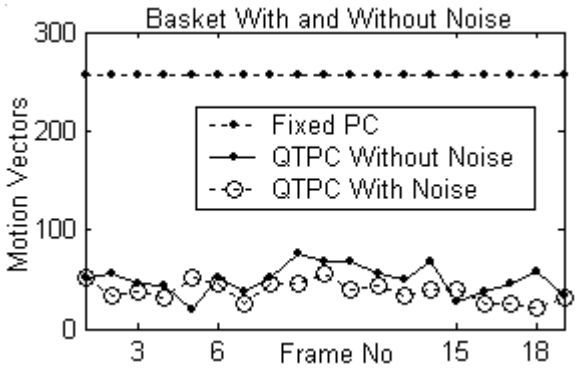

Fig.4. Motion vectors vs frame no. for a constant level of motion compensated prediction error for sequence 'Basket' with and without noise.

FIXED	CASE	Mobcal		Basket	
		W/out Noise	With Noise	W/Out Noise	With Noise
MSE	Fixed PC	176.59 (256)	289.12 (256)	222.84 (256)	331.72 (256)
	QTPC	(176.34) 26	(288.97) 22	(222.53) 64	(330.93) 40
MVs	Fixed PC	196.49 (64)	302.46 (64)	238.00 (64)	341.43 (64)
	QTPC	154.92 (64)	266.31 (64)	222.53 (64)	319.99 (64)

Table I. Performance comparison between fixed-block size and quad-tree phase correlation. Each entry contains the MSE value (upper) and the corresponding number of motion vectors (lower).

Subsequently we have adjusted the partition criterion so that both schemes yield 64 vectors. Results shown in Figure 5 demonstrate the superiority of our scheme. In both Figures 4 and 5 we have also included results obtained using artificially induced additive white gaussian noise of 30dB to demonstrate the reliability of partition criterion and phase correlation. Time-averaged results are summarised in Table I where MSE values for a fixed number of motion vectors and vice versa are shown for 'Mobcal' and 'Basket' with and without noise. Best cases are shown underlined confirming the superiority of our scheme. It is worth noting that a PSNR-based performance comparison with techniques based on exhaustive block-matching is rather meaningless because the latter scheme by definition achieves the global minimum of distortion metrics like the MSE. Nevertheless, we compare the proposed QTPC scheme with a Quad-Tree Block Matching (QTBM) scheme operating under the same quad-tree formation rules.

PSNR (dB)	Mobcal				Basketball			
Resolution	**1/1**	**1/2**	**1/4**	**1/8**	**1/1**	**1/2**	**1/4**	**1/8**
QTPC	26.4427	26.4427	26.4427	26.4427	25.3227	25.3227	25.3227	25.3227
QTBM	25.3034	26.4170	27.1416	27.2719	25.2141	26.1227	26.3922	26.4681

Table II. PSNR performance comparison with block matching.

Our results were obtained for the same number of motion vectors for each of the two schemes under comparison (256 vectors). Results for different subpixel accuracies were obtained for the block matching scheme while for QTPC there is no such accuracy issue because the function fitting procedure described in 2.1 yields interpolation-free floating point accuracy.

QTPC	$MN\left(4\log_2(MN)+9L\max\log_2(MN)-L\max(9L\max-7)\right)$
QTBM	$\frac{16}{3}\cdot M^2N^2\cdot\left(1-\left(\frac{1}{4}\right)^{L\max+1}\right)+(L\max+1)A^2MN$

Table III. Complexity comparison with block matching.

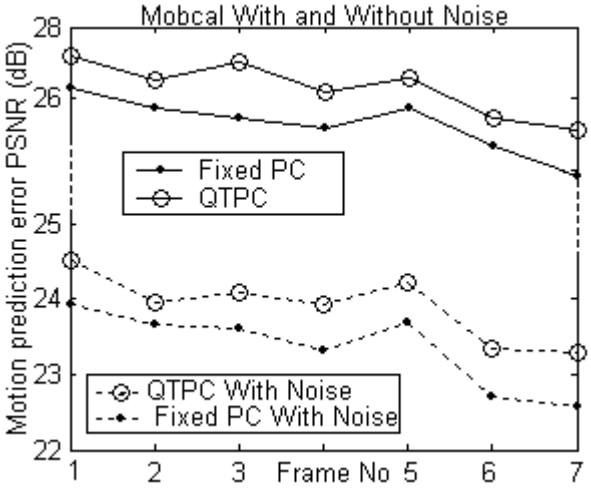

Fig.5. PSNR vs frame no. for the same number of motion vectors for sequence 'Mobcal' with and without noise.

SubPixel	1/1	1/2	1/4	1/8
Full-BM / QTPC	666.38	666.42	666.53	666.99

Table IV. Comparative complexity performance with block matching for selected parameter values of interest.

In Tables II we show averages (over all the processed frames of the two test sequences) of PSNR values and in Table III based on [12] we compare computational complexity in terms of number of real multiplications required by each scheme. For block matching *MxN* is the search area size and *M/2xN/2* the block size and *A* takes values in {2,4,8} for 1/2, 1/4, and 1/8 subpixel accuracy respectively. *Lmax* is the maximum number of levels. Finally in Table IV we tabulate computational complexity savings for selected parameter values of interest.

5. CONCLUSIONS

In this paper a quad-tree motion estimation algorithm based on phase correlation (QTPC) was presented. Owing to the fact that QTPC operates in the frequency domain it enjoys a high degree of computational efficiency and can be implemented by fast algorithms (FFT). In comparison to fixed-block schemes it achieves a significant degree of scene adaptation and offers better performance in terms of number of motion vectors for a fixed level of motion prediction error as well as PSNR for the same number of vectors with or without manually induced noise.

6. REFERENCES

[1] P. Strobach, "Tree-Structured Scene Adaptive Coder," IEEE Trans. Comm., vol. 38, no. 4, April 1990.

[2] H. Nicolas and C. Labit, "Region-based motion estimation using deterministic relaxation schemes for image sequence coding," Proc. ICASSP, vol. 3, pp. 265-268, 1992.

[3] K. Jensen and D. Anastassiou, "Digitally assisted deinterlacing for EDTV," IEEE Trans. Circ. Syst. Video Tech., vol. 3, no. 2, pp. 99-106, April 1993.

[4] M. R. Banham, J. C. Brailean, C. L. Chan, and A. K. Katsaggelos, "Low bit rate video coding using robust motion vector regeneration in the decoder," IEEE Trans. Image Proc., vol. 3, pp. 652-665, Sept. 1994.

[5] G. J. Sullivan and R. L. Baker, "Efficient quadtree coding of images and video" IEEE Trans.Image Proc,vol.3, pp.327, May 94.

[6] V. Seferidis and M. Ghanbari, "Generalised block-matching motion estimation using quad-tree structured spatial decomposition," Proc. IEE-I., vol. 141, no. 6, Dec. 1994.

[7] J. Lee, "Optimal quadtree for variable block size motion estimation," in Proc. ICIP, vol. 3, pp. 480-483, Oct. 1995.

[8] G. M. Schuster and A. K. Katsaggelos, "An Optimal Quad-Tree-Based Motion Estimation and Motion-Compensated Interpolation Scheme for Video Compression," IEEE Trans. Image Proc., vol. 7, no. 11, pp. 1505-1523, Nov. 1998.

[9] R. A. Packwood, M. K. Steliaros and G. R. Martin, "Variable Size Block Matching Motion Compensation for Object-Based Video Coding," Proc. IEE.Conf. Image Proc. and Appl., no. 443, pp.56-60, July 1997.

[10] S. Tredwell and A. N. Evans, "Embedded Quad-Tree Motion Estimation for Low Bit Rate Video Coding,", IEE Europ. Workshop Distributed Imag., No.1999/109, pp.3/1-3/5, Nov. 1999.

[11] P.J.Cordell and R.J.Clarke, "Low bit rate image sequence coding using spatial decomposition' Proc.IEE I,vol.139,pp.575, 92.

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

[13] 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.

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