encoded at a rate of 10kbit/s. Image quality was evaluated by peak signal-to-noise ratio (PSNR). For comparison, BMA (block matching algorithm) and the conventional 2D triangular mesh based [3] methods were also implemented in our experiments. In BMA, mean absolute difference is used as the prediction error criterion and the search area is up to [–15, +15] pixels in the horizontal and vertical directions around the original 16 × 16 macroblock position. The PSNR and the normalised execution time produced by these three methods are provided in Figs. 1 and 2, respectively.

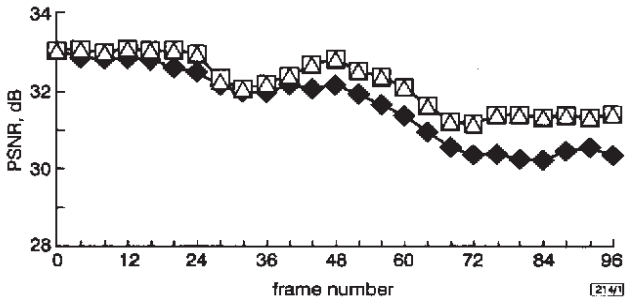

Fig. 1 *PSNR of results for 'Mother and Daughter' at 10kbit/s and 7.5f/s*

◆ BMA
■ 2D mesh
△ proposed

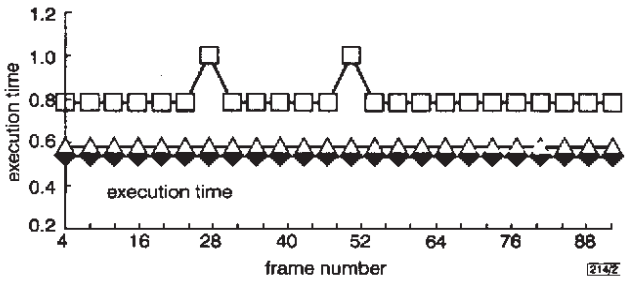

Fig. 2 *Execution time for different methods*

◆ BMA
□ 2D mesh
△ proposed

From Fig. 1, it is observed that the PSNR performance of the proposed method is better than that of the BMA by a factor of 0.55dB. Conversely, the PSNR of the proposed method is very close to that of the conventional 2D triangular mesh based method. From Fig. 2, it can also be seen that the processing speed of the proposed method is nearly 37% better than that of the 2D triangular mesh based method, and is close to that produced by the BMA method.

Conclusions: A fast motion estimation method for 2D triangular mesh based image coding has been presented. The proposed method solves two big problems of previous methods. One is the low image quality in the BMA method, and the other is the expensive computational cost in the conventional 2D triangular mesh based method. The simulation results show that the average PSNR of the proposed method is better than that of the BMA method, and the computational cost of the proposed method is cheaper than that of the conventional 2D triangular mesh based method. Therefore, the proposed algorithm can be very useful for very low bit rate real-time video coding applications.

© IEE 1998
5 January 1998
Electronics Letters Online no: 19980400

Hak Soo Kim, Taekhyun Yun and Kyu Tae Park (*Computer Application Laboratory, Department of Electronic Engineering, Yonsei University, 134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, Korea*)

E-mail: hskim@eve.yonsei.ac.kr

References

1 NAKAYA, Y., and HARASHIMA, H.: 'An iterative motion estimation method using triangular patches for motion compensation'. SPIE Visual Commun. and Image Processing '91: Visual Commun., Boston, MA, November 1991, Vol. 1605, pp. 546–557

2 CHOI, C.S., AIZAWA, K., HARASHIMA, H., and TAKEBE, T.: 'Analysis and synthesis of facial image sequences in model-based image coding', *IEEE Trans. Video Technol.*, 1994, **4**, pp. 257–275

3 NAKAYA, Y., and HARASHIMA, H.: 'Motion compensation based on spatial transforms', *IEEE Trans. Video Technol.*, 1994, **4**, pp. 339–356

4 ROSS, S.M.: 'Introduction to probability and statistics for engineers and scientists' (John Wiley & Sons, 1987)

Splitting criterion for hierarchical motion estimation based on perceptual coding

J. Malo, F.J. Ferri, J. Albert and J.M. Artigas

A new entropy-constrained motion estimation scheme using variable-size block matching is proposed. It is known that fixed-size block matching as used in most video codec standards is improved by using a multiresolution or multigrid approach. In this work, it is shown that further improvement is possible in terms of both the final bit rate achieved and the robustness of the predicted motion field if perceptual coding is taken into account in the motion estimation phase. The proposed scheme is compared against other variable- and fixed-size block matching algorithms.

Introduction: Motion estimation is a crucial step in video coding, mainly because optical flow is used to remove temporal redundancy from the signal. In current standards, the fixed-size block matching algorithm (BMA) is commonly used for motion compensation due to its trade-off of simplicity, computational load and performance [1].

Nevertheless, a fixed-size BMA usually implies a bad trade-off between motion information and prediction error volume because many motion vectors are wasted in stationary areas while more precision is needed along moving edges. Also, it exhibits a certain tendency towards unstable estimates [2].

To partially overcome these drawbacks, several hierarchical or variable-size BMAs have been proposed [2 – 5]. By operating in a top-down manner at different resolution levels they can give reliable, locally adapted motion estimates. Apart from different search strategies and similarity measures used, the splitting criterion is the most important part of most multigrid schemes because it allows the algorithm to locally refine the motion field. Several splitting criteria have already been used:

(i) A measure of the magnitude of the prediction error (MSE or MAE) [3, 4].

(ii) A measure of the volume of the signal to be sent to the decoder. The entropy of the motion field plus the zero-order entropy of the error signal in the spatial domain has been used [2, 5].

It has been pointed out that the entropy-based criterion is better than other criteria based on frame differences because, by its nature, it is well suited for coding applications. Moreover, it does not require any pre-specified threshold as in the first criterion.

As all of the above criteria use spatial domain measures, it cannot be guaranteed that the splitting process will be stopped when no further coding benefit is obtained. Moreover, this over-resolution may give rise to unstable motion estimates and false alarms. In fact, the need for a merging step has been pointed out to avoid this problem [3].

New splitting criterion: The use of entropy-based measures adapts the motion prediction to the particular goal of coding, but it should be emphasised that reducing the volume of the error signal in the spatial domain [2, 5] does not exactly match the behaviour of most image coding schemes in which the final step is a frequency-dependent quantisation. The actual volume of the prediction error is the entropy of its quantised transform coefficients (*spectral entropy*) and not the entropy of the pixels of the error (*spatial entropy*). In this Letter, a new splitting criterion based on spectral entropy is proposed, in the context of entropy-constrained variable-size BMA, in the following way:

Split a block at a given resolution in a multigrid scheme if:

$$H\left[\vec{D}_{split}\right]+H_f[DFD_{split}] < H\left[\vec{D}_{nosplit}\right]+H_f[DFD_{nosplit}]$$

where $H(\vec{D})$ is the entropy of the displacement vectors and H_f (DFD) stands for spectral entropy of the displaced frame difference (DFD), which is defined as the entropy of the quantised error signal in the DCT domain:

$$H_f[DFD] = H[Q[DCT[DFD(\mathbf{x})](\mathbf{f})]]$$

This change really makes a difference because splitting a block using previous criteria will give rise to a reduction in complexity of the error signal in the spatial domain that may be completely ignored by the quantiser which usually carries out frequency perceptual information [1, 6].

If this perceptual information is taken into account in the splitting criterion (using the quantised signal), it is possible to obtain a more compact motion description and a better trade-off between motion information and prediction error: first, because the splitting criterion is more closely connected to the way in which the coder performs, thus increasing the precision only if the corresponding quantiser can take advantage of it; and secondly, because perceptual information gives rise to a motion estimate that implies more accurate compensations from an observer's point of view. This means that perceptually important features of the scene (seen as sharp moving edges) receive more effort for prediction. Even though this scheme obviously requires more computation than spatial-based algorithms, this additional effort is restricted to the (usually off-line) encoding phase while the decoding phase remains unchanged.

Results and discussion: Several experiments have been carried out using different standard image sequences. Fixed- and variable-size BMAs using both spatial and spectral entropy criteria have been considered at the same bit rate (200kbit/s) for comparison purposes. Blocks of size 8×8 have been used in a fixed-size BMA. A quadtree multigrid scheme with five resolution levels (blocks from 64×64 to 4×4) has been used in a variable-size BMA. The n-step displacement search [2], has been applied to compute motion vectors at all resolution levels. The usual correlation has been used as a similarity measure. A nonlinear perceptually-weighted DCT quantiser has been used to take into account the frequency and amplitude response of the human viewer [6].

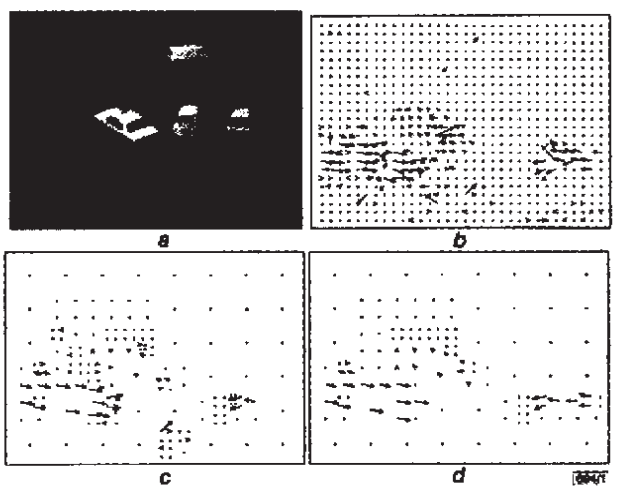

Fig. 1 *Estimated motion field using three BMA considered*

a Frame 6 of *Taxi* sequence
b Fixed-size
c Spatial entropy criterion
d Spectral entropy criterion

Frame six of the *Taxi* sequence is shown in Fig. 1, along with the motion field estimate using the three algorithms considered. It is worth noting that with the proposed criterion, fewer motion vectors are needed to properly describe the motion in the scene. Fewer false alarms are obtained in the static areas, and significant motions are better described. Note how in the static area at the bottom of Fig. 1*c*, the spatial entropy-based algorithm gives the same false alarm as the uniform resolution BMA (Fig. 1*b*), due to an inadequate increase in resolution. The motion of the incoming vehicle on the right is better described by the proposed method.

A side-effect of using perceptual information is a more conservative splitting strategy which gives rise to more robust estimates. This is because perceptual quantisers show bandpass behaviour [1, 6] neglecting certain aspects of the prediction error that would have been considered by the objective spatial domain based criteria. In fact, resolution is increased only if the corresponding motion structures are perceptually relevant. Most moving objects can be properly identified with the proposed criterion, giving rise to coherent fields without spurious vectors, unlike in the other cases.

Table 1: Motion field volumes (kbit/s) for different sequences

	Taxi	Rubik	Yosemite	Trees
Fixed-size BMA	25.15	43.42	92.37	78.42
Spatial entropy	3.78	6.58	17.38	9.94
Spactral entropy	1.87	3.73	8.64	6.48

Table 1 shows the volume of the motion description using the three algorithms. With the proposed criterion, the motion information needed for compensation is approximately reduced by factors of 2 and 10 with regard to spatial entropy-based variable-size BMA and fixed-size BMA, respectively. Savings in motion information can be spent by the new approach in a more accurate encoding of the prediction errors, giving rise to subjectively better reconstructions.

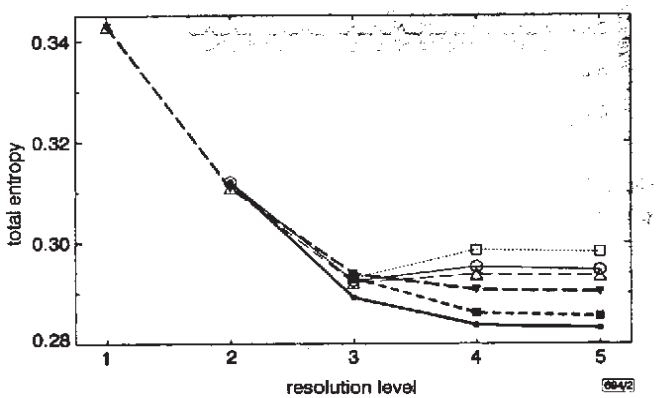

Fig. 2 *Volume of flow and prediction errors (in bit/pix) while refining motion estimate for two variable-size BMA considered at three different starting resolution levels*

▼— —▼ spectral criterion (i)
●——● spectral criterion (ii)
■- - -■ spectral criterion (iii)
△- - -△ spatial criterion (i)
○——○ spatial criterion (ii)
□·····□ spatial criterion (iii)

At each frame, the entropy of the encoded signal (flow and quantised errors) has been computed at the different resolution levels of the quadtree during the application of the variable-size BMA. Fig. 2 shows these results averaged over all frames. Different starting resolution levels are considered with both splitting criteria. As expected, the proposed criterion does minimise the volume of the signal in the meaningful encoding domain, while the spatial entropy criterion gives rise to an increase in volume during the final steps, presumably because too many blocks are unnecessarily split.

Conclusion: A new splitting criterion for hierarchical BMA has been proposed. From the viewpoint of video coding, the new method improves on other approaches because the behaviour of the quantiser is taken into account for motion estimation. In this way, motion estimation effort is directed towards certain areas in such a way that the final bit rate is minimised. Conversely, due to perceptual weighting, the final flow exhibits some interesting properties, such as coherence, compactness, robustness and reliability, that could make it suitable for more general applications.

Acknowledgment: This work was partially supported by the Spanish CICYT project TIC95-676-C02-01.

© IEE 1998
27 November 1997
Electronics Letters Online No: 19980404

J. Malo and J.M. Artigas (*Dept. d'Òptica, Universitat de València, Dr. Moliner, 50, 46100 Burjassot, Spain*)

F.J. Ferri and J. Albert (*Dept. d'Informàtica i Electrònica, Universitat de València, Dr. Moliner, 50, 46100 Burjassot, Spain*)

E-mail: ferri@uv.es

References

1 LEGALL, D.: 'MPEG: a video compression standard for multimedia applications', *Commun. ACM*, 1991, **34**, (4), pp. 47–58
2 DUFAUX, F., and MOSCHENI, F.: 'Motion estimation techniques for digital TV: A review and a new contribution', *Proc. IEEE*, 1995, **83**, (6), pp. 858–875
3 CHAN, M.H., YU, Y.B., and CONSTANTINIDES, A.G.: 'Variable size block matching motion compensation with applications to video coding', *IEE Proc.*, 1990, **137**, (4), pp. 205–212
4 LI, J., and LIN, X.: 'Sequential image coding based on multiresolution tree architecture', *Electron. Lett.*, 1993, **29**, (17), pp. 1545–1547
5 MOSCHENI, F., DUFAUX, F., and NICOLAS, H.: 'Entropy criterion for optimal bit allocation between motion and prediction error information'. SPIE Proc. Visual Communication and Image Processing, 1993, pp. 235–242
6 MALO, J., PONS, A.M., and ARTIGAS, J.M.: 'Bit allocation algorithm for codebook design in vector quantisation fully based on human visual system non-linearities for suprathreshold contrasts', *Electron. Lett.*, 1995, **31**, (15), pp. 1222–1224

Video object motion representation using run-length codes

M.K. Steliaros, G.R. Martin and R.A. Packwood

A method of runlength coding motion vector data for object based video coding is proposed. It is shown that significant improvements in coding efficiency can be gained for certain classes of source material.

Introduction: Evolving object-based video coding standards, such as MPEG-4, permit arbitrary-shaped objects to be encoded and decoded as separate video object planes (VOPs). In MPEG-4, the exact method of producing VOPs from the source imagery is not defined, but it is assumed that 'natural' video objects are represented by shape information in addition to the usual luminance and chrominance components. Shape data is provided as a binary segmentation mask, or a grey scale alpha plane to represent multiple overlaid objects. The MPEG-4 video verification model [1] proposes that motion compensation (MC)-based predictive coding is applied to each VOP. Conventional fixed size block matching (FSBM) motion estimation, of the type originally proposed by Jain and Jain [2], is employed. In this Letter, we propose two improvements: a best vector selection strategy, and an alternative method of coding motion vector data, which results in a considerable improvement in coding efficiency.

Block matching: In MPEG-4 it is proposed that FSBM is used to estimate the motion in presegmented video objects.

The VOP is constructed by placing 16 × 16 pixel macro-blocks over the shape description, and then further extended by one macro-block to cater for movement of the object between frames. A repetitive padding technique is used to fill out those areas which are inside the block structure but outside the object's shape mask.

The padding is designed to feed the best possible input into the next stage of coding, especially when material that did not exist in the reference frame is introduced. The resulting VOP is used as the reference frame for the block matching operation. Motion estimation is performed by matching the 16 × 16 macro-blocks within a prescribed search area in the reference frame. For maximum accuracy, an exhaustive search is used, and matching is conducted to 1/2 pixel precision. There are several options for the matching criterion [3], although that most commonly used is the sum absolute difference (SAD), since it is not as computationally demanding as others and yet achieves similar performance. It provides a measure of the error between a block of size $n \times n$ pixels in the current frame (C) and a block with displacement (x, y) in the reference frame (R) and can be described as

$$SAD(x,y) = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} |C(i,j) - R(i+x, j+y)| \quad (1)$$

where $C(i, j)$ and $R(i, j)$ represent the intensity of the pixel at location (i, j) in the current and reference frames, respectively. Both ITU-T H.263 [4] and MPEG-4 modify the SAD for the zero vector in order to favour it when there is no significant difference between vectors:

$$SAD(0,0) = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} |C(i,j) - R(i,j)| - \left(\frac{N_B}{2} + 1\right) \quad (2)$$

where N_B is the number of pixels in the block that are inside the VOP. These 'best' vectors are subsequently differentially coded, typically using a two-dimensional (2D) prediction strategy described in more detail in the next section.

Even with the zero bias, however, the above definition does not uniquely identify a 'best' choice when multiple vectors produce identical errors. A naive implementation of the exhaustive search algorithm would perform the search in a raster scan order. If each motion vector is denoted by $\mathbf{v}_i\ (x_i, y_i)$, where x_i, y_i (or $2x_i$, $2y_i$ for half pixel search) are integers, $-r \le x_i, y_i < r$ and r is the search radius, then for a raster scan, all $\mathbf{v}_i$ need to be ordered such that:

$$\mathbf{v}_i < \mathbf{v}_{i+1} \Leftrightarrow (y_i < y_{i+1}) \vee ((y_i = y_{i+1}) \wedge (x_i < x_{i+1}))$$
$$\forall\, i \in [0, 4r^2 - 1] \quad (3)$$

Since vector quality affects the 2D prediction, choosing the smallest vector in Euclidean terms will improve coding efficiency. The simplest solution in terms of implementation costs is to impose an optimal ordering in the search algorithm which ensures that smaller vectors are tested before larger ones. We therefore propose that the ordering condition has to be modified to:

$$\mathbf{v}_i < \mathbf{v}_{i+1} \Leftrightarrow (|\mathbf{v}_i|^2 < |\mathbf{v}_{i+1}|^2) \vee \Big((|\mathbf{v}_i|^2 = |\mathbf{v}_{i+1}|^2) \wedge ((x_i < x_{i+1}) \vee ((x_i = x_{i+1}) \wedge (y_i < y_{i+1})))\Big)$$
$$\forall\, i \in [0, 4r^2 - 1] \quad (4)$$

2D motion vector prediction: In MPEG-4 it is proposed that FSBM is used to perform motion estimation in pre-segmented video objects. The motion vector (MV) x and y components for each macro-block are transmitted as Huffman-type variable length codes (VLC) which depend on the spatial neighbourhood of three previously transmitted motion vectors and are calculated as the median of the vectors immediately to the left, above and to the above right of the block being considered, with special rules applying near the VOP boundaries.

Image sequences which exhibit small changes in motion from block to block result in small MVDs, and thus require few bits of VLC to represent the motion. In the extreme case, when motion vectors are identical, two bits are used for each block. This might arise in a panning shot or a still picture.

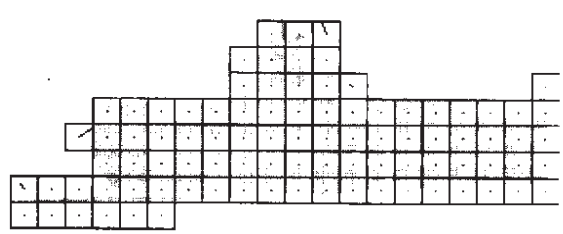

Fig. 1 *FSBM block structure example for 'container ship'*

Fig. 1 shows an example block structure produced by FSBM for a frame from the MPEG-4 test sequence 'container ship'. The short lines shown in the centre of the blocks indicate the motion vector associated with the block. Although most vectors are