GLOBAL MOTION ESTIMATION USING VARIABLE BLOCK SIZES AND ITS APPLICATION TO OBJECT SEGMENTATION

Marina Georgia Arvanitidou [(1)], Alexander Glantz [(1)], Andreas Krutz [(1)], Thomas Sikora [(1)]
Marta Mrak [(2)], Ahmet Kondoz [(2)]

[(1)] Communication Systems Group, Technische Universität Berlin, Germany
[(2)] Centre for Communication Systems Research, University of Surrey, United Kingdom

ABSTRACT

Global motion is estimated either in the pixel domain or in block based domain. Until now, all the approaches regarding the latter are based on fixed sized blocks while the recent compression methods tend to use variable block sizes during motion estimation. In this paper we present a new procedure for global motion estimation based on a variable block size motion vector field. A block matching algorithm which is able to adapt the block size according to the motion complexity within the frame is used. The resulting motion vectors are employed for global motion estimation. Furthermore, binary foreground-background masks are created based on the frame-by-frame motion compensated differences by exploiting spatial conditions through anisotropic diffusion filtering. For global motion estimation the performance evaluation in terms of background PSNR shows an enhancement of more than 2.5 dB in the well-known "Stefan" sequence, compared to the conventional case of fixed block size, at a reasonable implementation complexity.

1. INTRODUCTION

Global motion is commonly used to describe the motion of the background in video sequences. It is generally induced by camera motion and modeled by parametric transforms of two-dimensional images. The process of estimating the transform parameters is called global motion estimation (GME) and is widely used for video coding [1] as well as for object segmentation [2], [3] and other applications.

A number of GME algorithms have been proposed in the past, and generally they can be categorized into direct and indirect methods. Direct GME methods are pixel-based and try to minimize the prediction error in the pixel domain. Indirect GME methods contain two stages where GME is performed at the second stage, based on the motion vectors resulted from the first motion estimation stage. This first motion estimation stage, which might be quite time-consuming, could also be skipped when exploiting the motion vectors available in a coded stream. [4], [5], [6].

The advantage in this case compared to direct (pixel-based) methods is the alleviation of the computational burden. Smolic et al. presented in [5] their low-complexity GME algorithm based on P-frame motion vectors using an M-estimator for outlier rejection. In [6] the global motion model is estimated using the Newton–Raphson method with outlier rejection for minimizing the error between the input and the estimated motion vector field. Another compressed domain approach is presented in [7] where the same problem is tackled using DCT coefficients instead of motion vectors. All the techniques mentioned above estimate global motion directly from MPEG compressed video sequences and all of them are built for the case that the video frames are partitioned in a predefined number of fixed size blocks.

Recent compression techniques tend to use variable blocks sizes during motion estimation and compensation stages. These methods are capable of adapting the block size according to the motion complexity within the frame. The proposed work shows how the global motion model can be estimated based on a motion vector field of variable block size partitions. The blocks are created using the binary partition tree method while the affine model is used for describing the global motion model as in [5].

In the following section the block matching algorithm is briefly described. Sub-Section 3.1 presents the implementation of the global motion estimation algorithm, while the object segmentation approach based on anisotropic diffusion is discussed in Sub-Section 3.2. Experimental results and the overall procedure performance are evaluated in Section 4, before conclusions are drawn in Section 5.

2. VARIABLE SIZE BLOCK MATCHING ALGORITHM USING BINARY PARTITION TREE

In video coding the motion is generally represented by motion vectors that are associated with the blocks. Frequently, the frame partitioning into blocks is realized using fixed-size blocks, e.g. blocks of 16 × 16 pixels in

978-1-4244-3610-1/09/$25.00 ©2009 IEEE.

MPEG-2. Recently it has been shown that the actual compression can be enhanced when variable size blocks (i.e. flexible motion models) are employed. Variable size blocks models are capable of changing the block size according to the motion complexity within the frame. Small blocks for frame regions with complex motion are used when higher final bit-rates are available. On the other hand, large blocks are used for regions with simple motion such as background, or when exceptionally low bit-rates are targeted, such that even motion bit-rates have to be reduced.

The currently most efficient video-compression standard H.264 / AVC also uses variable size blocks with limited flexibility [8]. In H.264 / AVC the block sizes are varied between 4 × 4 and 16 × 16 pixels, while for block dimensions only multiples of 4 pixels are used. A model that offers higher flexibility of frame partitioning, and therefore better adaptation to the actual content, is often referred to as binary partition tree (BPT) model, (see [9]). This method enables adaptive partitioning of video frames, originally motivated by rate-distortion optimization requirements in compression. Recently, it has been shown that this method is very suitable for application in 3D video coding [10], especially for depth-map images, since BPT model enables excellent adaptability to the actual frame content.

In the schemes based on variable size block models, like here the BPT, the block sizes are not predefined. Therefore, block sizes can be varied to optimize the trade-off between number of bits used to encode motion vectors and residual (rate-distortion requirements). The partitioning of a frame into blocks is described with a tree-structure and can be achieved using a two-step algorithm. First step is the growing of the tree by frame partitioning (top-down approach). Second step is the pruning of the tree which finds the optimal partitioning with respect to given requirements (bottom-up approach).

During the tree growing step, the entire picture is repeatedly split up to a target number of N blocks. Initially the whole frame is considered as one block. Optimal partitioning is achieved using motion estimation and its actual partitioning is described in the tree root and with its two new "branches" that represent two new blocks. Then the iterative procedure continues.

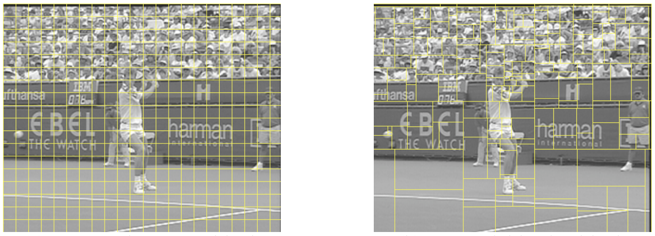
At the bottom up step, the tree is pruned in order to find the optimal partitioning, in the R-D sense.

The BPT model has demonstrated promising results in video coding. Its main advantage is its capability to partition the frame along actual motion boundaries.

3. GLOBAL MOTION ESTIMATION AND MOTION BASED SEGMENTATION

3.1. Global Motion Estimation based on Motion Vectors

In order to estimate the global motion in each frame, we use the motion vectors derived from the BPT procedure as described above. As one can notice in Fig. 1 large size blocks, which correspond to homogeneous areas (e.g. tennis court), tend to belong to the background. Respectively, more detailed areas, which correspond to smaller sized blocks (e.g. tennis player), tend to belong to the moving foreground object. Our proposed approach takes advantage of this fact, and is based on [5], using a more realistic 6-parameter affine motion model instead of the simplified affine with 4 parameters.



(a) 396 regular 16x16 blocks

(b) 198 variable-size blocks

Fig. 1 Block partitioning on frame 13 of Stefan sequence. In (a) a block matching algorithm is used and in (b) the variable block matching algorithm, using BPT, is used

Let M be the total number of blocks that a frame is divided into. In contrast with [5] (N = M), we consider only the N<M blocks which cover greater than the average surface that a block covers. The relation between the affine motion parameters and the motion vectors can be formulated as

$$V = H \cdot \Xi \qquad (1)$$

Or equally

$$\begin{bmatrix} v_x^{(1)} + x^{(1)} \\ v_y^{(1)} + y^{(1)} \\ \vdots \\ v_x^{(N)} + x^{(N)} \\ v_y^{(N)} + y^{(N)} \end{bmatrix} = \begin{bmatrix} x^{(1)} & y^{(1)} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x^{(1)} & y^{(1)} & 1 \\ \vdots & & & & & \\ x^{(N)} & y^{(N)} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x^{(N)} & y^{(N)} & 1 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} \qquad (2)$$

where $\left(v_x^{(n)}, v_y^{(n)}\right)$ and $\left(x^{(n)}, y^{(n)}\right)$ denote the motion vector and the block center coordinates of the n^{th} block respectively (in the N number of participating blocks). The vector $\Xi = [a_1, a_2, a_3, a_4, a_5, a_6]^T$ includes the motion parameter set, when the transformed position (x', y') of a point (x, y) is described as $\left(a_1 + a_2 x + a_{3y} \ , \ a_4 + a_5 x + a_6 y \right)$.

This is a suitable approximation of the global motion over a short period of time. The least squares solution of Equation 1 with respect to Ξ is

$$\Xi = \left(\mathrm{H}^T \cdot \mathrm{H}\right)^{-1} \cdot \mathrm{H}^T \cdot V \qquad (3)$$

The accuracy of the affine motion parameters achieved from a motion vector field is often influenced by the existence of outliers. In order to eliminate this effect, a robust M-estimator with its diagonal weighting matrix W is used within the computation of the affine motion parameters (see [5]), which leads to

$$\Xi = \left(\mathrm{H}^{T} \cdot W \cdot \mathrm{H}\right)^{-1} \cdot \mathrm{H}^{T} \cdot W \cdot V \qquad (4)$$

In the case of variable block sizes method, the consideration of only the set of blocks with sufficient size, serves as an outlier rejection method. Therefore, even without the use of M-estimator the prediction in this case outperforms the case using fixed size blocks.

3.2. Segmentation using anisotropic diffusion

Once the warping matrix Ξ is calculated for every pair of adjacent frames, both in case of variable and fixed-size block partition, the compensated frame-by-frame difference is computed as the absolute value of the subtraction of the luminance values of any frame I_n with the camera motion compensated consecutive frame I_{n+1}. On these frame-to-frame differences the segmentation algorithm as described in [11] is applied where anisotropic filtering is used for exploitation of spatial conditions. An overview of the algorithm is presented in Table 1.

Since segmentation requires accurate GME and compensation, the object segmentation based on our method outperforms the corresponding one using fixed size blocks as shown in the following section.

4. EXPERIMENTAL RESULTS

We have evaluated our approach using two sequences. The well known "Stefan" (352x240, 300 frames) as well as the "Biathlon" (352x288, 200 frames) sequence, which was recorded from a German TV broadcaster.

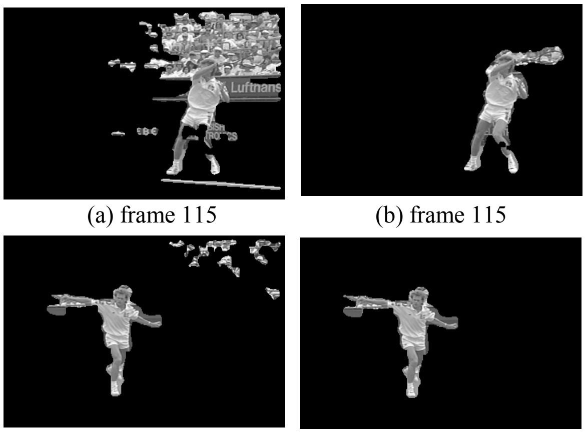
"Stefan" sequence is especially complex both in terms of content and motion. It consists of low-frequency (tennis court) and high-frequency parts (audience), while the camera motion is composed of translation, scaling and perspective transformation.

The objective evaluation of the background quality is done by computing the background PSNR of the error frames using the luminance information of the sequence.

In order to compute the PSNR values only between the background pixels of the original reference frame and the warped reference frame, we have created and used ground truth binary object masks for both sequences.

Table 1

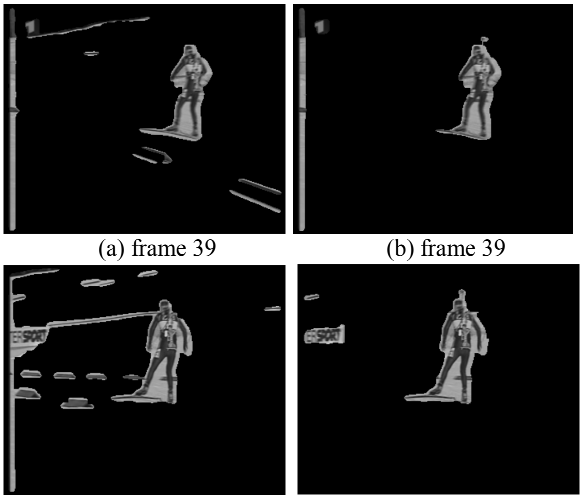
(i)	Anisotropic lowpass filtering using diffusion
(ii)	Intensity rescaling [0, 1] – image binarization using threshold τ
(iii)	Small objects removal and hole filling using morphological operators
(iv)	Binary foreground/background mask creation



(a) frame 115 (b) frame 115

(c) frame 27 (d) frame 27

Fig. 2 Object segmentation results on "Stefan. (a), (c) with reference approach and (b), (d) with proposed approach



(a) frame 39 (b) frame 39

(c) frame 87 (d) frame 87

Fig. 3 Object segmentation results on "Biathlon. (a), (c) with reference approach and (b), (d) with proposed approach.

In Fig. 4 and Fig. 5 the background PSNR is illustrated over the "Stefan" and "Biathlon" sequences respectively. Dashed line corresponds to the case of GME based on 8x8 block sized motion vectors and continuous line to the proposed GME as described in Section 3. As it can be observed and also listed in Table 2, our proposed algorithm outperforms the one described in [5] on both sequences that it is applied.

In Fig. 2 and Fig. 3 object segmentation results are illustrated and the proposed algorithm ((b) and (d)) is compared with the case of regular 8x8 blocks used for GME, compensation and segmentation (cases (a) and (c)).

The latter has significantly improved performance, since the outliers are less dominant compared to the first case.

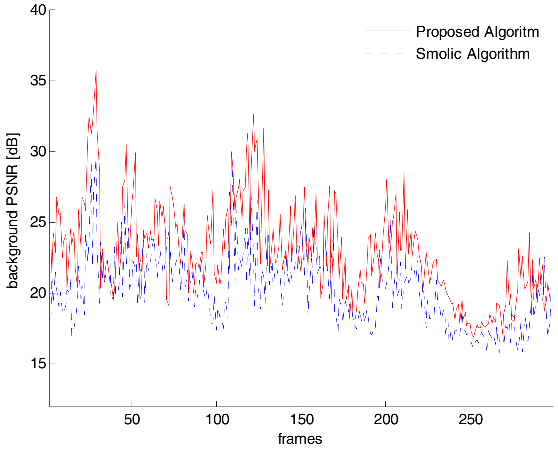


Fig. 4 Comparison of background PSNR for the "Stefan" sequence.

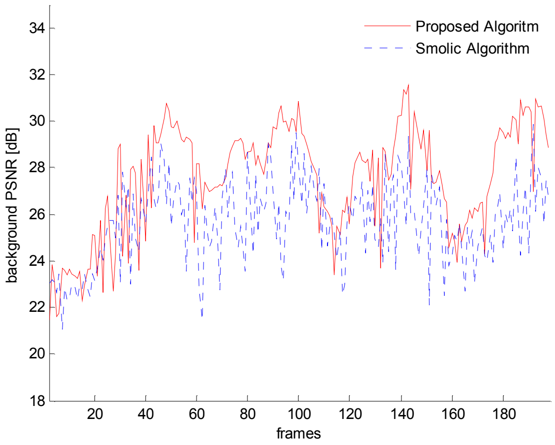


Fig. 5 Comparison of background PSNR for the "Biathlon" sequence.

Table 2 Mean background PSNR comparing reconstruction using GME based on FSB and VSB

Sequence	Algorithm	PSNR [dB]
"Stefan"	GME using FSB (Smolic) GME using VSB (proposed)	20.2342 **22.8906**
"Biathlon"	GME using FSB (Smolic) GME using VSB (proposed)	25.6220 **27.4758**

5. CONCLUSIONS

We have presented a new procedure for global motion estimation based on variable size motion vector field and its application to object segmentation. The objective evaluation shows that our approach improves the global motion estimation compared with the case of fixed sized blocks. This is reasonable since the motion diversities in the frame define the size/shape of the blocks used for the GME in a way that blocks obviously belonging to the background affect more the calculation of the camera motion model.

6. ACKNOWLEDGMENTS

The work presented was developed within VISNET II, a European Network of Excellence and funded by the European Commission IST FP6 program.

7. REFERENCES

[1] T. Sikora, "Trends and perspectives in image and video coding", *Proc. of the IEEE*, Vol. 93, January 2005, pp. 6-17.

[2] A. Smolic, T. Sikora and J.-R. Ohm, "Long-term global motion estimation and its application for sprite coding, content description, and segmentation", *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 9, no. 8, pp. 1227-1242, December 1998

[3] B. Qi, M. Ghazal and A. Amer, "Robust Global Motion Estimation Oriented to Video Object Segmentation", *IEEE Transactions on Image Processing,* vol. 17, no. 6, pp. 958-967, June 2008.

[4] J. Heuer and A. Kaup, „Global Motion Estimation in Image Sequences Using Robust Motion Vector Field Segmentation", *in Proc. ACM Multimedia*, Orlando FL, November 1999*,* pp. 264-264

[5] A. Smolic, M. Hoeynck, and J. R. Ohm, "Low-complexity global motion estimation from P-frame motion vectors for MPEG-7 applications," *in Proc. of IEEE International Conf. on Image Processing,* Vancouver, September 2000.

[6] Y. Su, M.-T. Sun and V. Hsu, "Global motion estimation from coarsely sampled motion vector field and the applications", *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 15, no. 2, pp. 232-242, February 2005

[7] E. Saez, J.M. Palomares, J. I. Benavides and N. Guil, "Global motion estimation algorithm for video segmentation", *in Proc., Visual Communications and Image Processing 2003, Lugano, July 2003, vol 5150, pp. 1540-1550*

[8] M. Wien, "Variable Block-Size Transforms for H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 13, no. 7, pp. 604 - 613, July 2003

[9] M. Servais, T. Vlachos and T. Davies, "Motion Compensation using Variable -Size Block-Matching with Binary Partition Trees," *in Proc. of IEEE International Conference on Image Processing***,** Genova, September 2005

[10] B. Kamolrat, A. Fernando, M. Mrak, and A. Kondoz, "Flexible motion model with variable size blocks for depth frames coding in colour-depth based 3D video coding," *in Proc. of IEEE International Conference on Multimedia & Expo (ICME 2008),* June 2008

[11] A. Krutz, M. Kunter, M. Mandal, M. Frater, and T. Sikora, "Motion-based Object Segmentation using Sprites and Anisotropic Diffusion", *in Proc. of IEEE International Workshop on Image Analysis for Multimedia Interactive Services*, Santorini, June 2007