

ROBUST GLOBAL MOTION ESTIMATION FROM MPEG STREAMS WITH A GRADIENT BASED REFINEMENT

David Corrigan, Anil Kokaram*

Renan Coudray, Bernard Besserer

Department of Electronic and Electrical Engineering,
University of Dublin, Trinity College,
Dublin, Ireland.

L3I,
University of La Rochelle,
La Rochelle, France.

ABSTRACT

In this paper a new robust algorithm for Global Motion Estimation from an MPEG stream is proposed. The approach is to use a hybrid of two previous Global Motion Estimation algorithms in order to improve the accuracy and robustness of the estimate. The first algorithm is a non-parametric technique which uses the block based motion vectors obtained from an MPEG2 stream and which also generates a coarse background segmentation of the frame. The second algorithm is a gradient based technique which estimates the parameters from the image data directly. In this work, an initial estimate for the global motion is obtained using the non-parametric approach. The parameters are then refined using the gradient based estimation and the segmentation field. This hybrid approach shows significant reduction in mean compensation error over the non-parametric approach. The algorithm is then applied to the problem of mosaicking in sports sequences in which the global motion parameters can be used to make a panorama of an entire shot.

1. INTRODUCTION

Mosaicking [1, 2] is the arranging of the background for each frame of a sequence into a single image. It is a well established application in Image Processing and is also incorporated into MPEG4 as the notion of sprites. In order to form a mosaic, the global motion parameters between the current and previous frames must be known. Robust estimation of these parameters is essential to the quality of the mosaics as errors in the estimates propagate in the mosaic over time.

Typically the best quality estimate is obtained through direct matching of consecutive image frames [3, 4]. However these approaches are computationally expensive. Extracting the parameters from the MPEG2 stream is an attractive prospect proposed in [5] since it allows the quick estimation of the motion parameters. This method has a robust to local motion and enables a coarse background segmentation to be achieved. Furthermore motion can be estimated over a

*Thanks to the Irish Research Council for Science, Engineering and Technology for funding.

longer time period, from I-frame to I-frame, which are typically 12 frames apart. However as the estimate is generated from block based motion field, the accuracy is dependent on the vector field and is not sufficiently accurate over 12 frames.

The gradient based global motion estimator [6] is an example of a direct matching technique. This approach accurately estimates small displacements but it cannot estimate large displacements as accurately. Furthermore it is not as robust to local motion as [5] and so requires a background segmentation. The novelty in this paper is to exploit the quick and robust estimation and background segmentation of the non-parametric method but to improve the estimate using the precision of the gradient based estimator giving an accurate estimate of global motion from I-frame to I-frame.

The next sections give a brief description of the nonparametric and gradient based techniques while section 4 outlines the mosaicking process. Finally, experimental results and some final remarks are presented.

2. NON-PARAMETRIC GLOBAL MOTION ESTIMATION FROM A MOTION VECTOR FIELD

The non-parametric algorithm attempts to estimate zoom, rotation and translation parameters from a motion vector field $V(x, y)$ where (x, y) are the spatial co-ordinates of the motion vector. It is achieved by modeling motion fields caused by pure zoom, rotation and translation which are given by

$$V_T(x, y) = \begin{bmatrix} t_x \\ t_y \end{bmatrix}, V_Z(x, y) = \begin{bmatrix} z \cdot x \\ z \cdot y \end{bmatrix}, V_R(x, y) = \begin{bmatrix} -r \cdot y \\ r \cdot x \end{bmatrix}. \quad (1)$$

where t_x and t_y are the horizontal and vertical components of the translation and where z and r are the zoom and rotation factors respectively. From these models it can be deduced that

$$\frac{\partial V_{Z_x}}{\partial x} = \frac{\partial V_{Z_y}}{\partial y} = z, \quad -\frac{\partial V_{R_x}}{\partial y} = \frac{\partial V_{R_y}}{\partial x} = r, \\ V_{T_x} = t_x, V_{T_y} = t_y. \quad (2)$$

where V_{*x} and V_{*y} are the horizontal and vertical components of each vector field. The algorithm proceeds by using Eq. 2

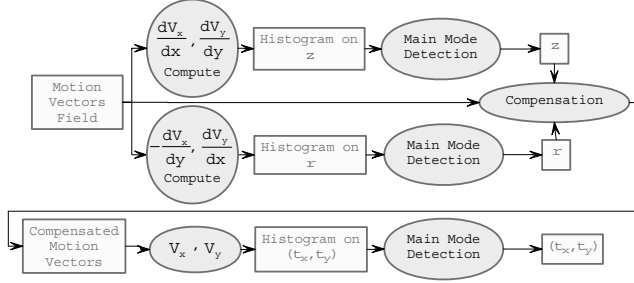


Fig. 1. The non-parametric Global Motion Estimation Algorithm

to generate estimates of the parameters from the vector field $V(x, y)$. Since the t_x and t_y add a constant value to the vector field, it does not affect the calculation of z and r . On the other hand, a zoom or rotation will affect the estimation of the translation. Therefore zoom and rotation must be compensated before translation can be calculated. Histograms on r and z are estimated from the partial derivatives and the values of z and r are given by the modes. The translation parameters are then estimated by first compensating $V(x, y)$ for zoom and rotation according to Eq. 1 and then by computing a 2D histogram on the compensated field. A flow chart of the algorithm is given in figure 1.

This result is equivalent to the six parameter affine motion model where the affine matrix A and translation \mathbf{d} are given by $A = [z - r; r z]$ (matlab notation) and $\mathbf{d} = [t_x, t_y]$. In [7] an implementation of this algorithm is given which estimates each of the 6 parameters independently. Although this approach gives more precise results than the 4 parameter case, it is not as robust. Furthermore since the gradient based approach also estimates the parameters independently and hence the 4 parameter implementation of the non-parametric approach is used.

The method for performing the background segmentation is also outlined in [7]. The segmentation map is defined on a coarse scale and segments the vectors in $V(x, y)$ into those contributing to the background motion and those to the foreground. This is performed by using a watershed algorithm [8] on the 2D histogram to generate a motion based segmentation. The background segment is chosen by the vectors which contributed to the modal peak. In addition [7] proposes an alternative segmentation technique to the standard watershed method. This method takes into account the sparsity of the histogram space in order to optimise the segmentation process.

3. GRADIENT BASED GLOBAL MOTION ESTIMATION

The gradient based algorithm employs an image sequence model as follows

$$I_n(\mathbf{x}) = I_{n-1}(f(\mathbf{x}, \mathbf{a})) + \epsilon(\mathbf{x}) \quad (3)$$

where I_k is the k^{th} frame of the sequence, $f(\mathbf{x}, \mathbf{a})$ is a linear affine transformation, where $\mathbf{a} = (a_1, a_2, a_3, a_4, d_x, d_y)^T$ and where $\epsilon(\mathbf{x})$ is the model error. The transformation used is $f(\mathbf{x}, \mathbf{a}) = A\mathbf{x} + \mathbf{d}$ where A is of the form $[a_1 \ a_2; a_3 \ a_4]$ and $\mathbf{d} = (d_x, d_y)^T$. By expanding $I_{n-1}(\cdot)$ using a Taylor series expansion about an initial guess for the motion \mathbf{a}_0 , equation 3 becomes

$$I_n(\mathbf{x}) = I_{n-1}(f(\mathbf{x}, \mathbf{a}_0)) + \mathbf{u}^T \nabla I_{n-1}(f(\mathbf{x}, \mathbf{a}_0)) + e_{n-1}(\mathbf{x}) \quad (4)$$

where $\mathbf{a} = \mathbf{u} + \mathbf{a}_0$ and where $e_{n,n-1}(\mathbf{x})$ is the sum of the model error and the higher order terms of the expansion. The gradient operator ∇ is given by

$$\nabla = \left(\mathbf{x}_x \frac{\partial}{\partial x}, \mathbf{x}_y \frac{\partial}{\partial x}, \mathbf{x}_x \frac{\partial}{\partial y}, \mathbf{x}_y \frac{\partial}{\partial y}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right)^T \quad (5)$$

where \mathbf{x}_x and \mathbf{x}_y are the horizontal and vertical components of \mathbf{x} . Ignoring the error $e_{n,n-1}(\mathbf{x})$ eq. 4 can now be expressed in the form

$$\Delta(\mathbf{x}, \mathbf{a}_0) = \mathbf{u}^T \nabla I_{n-1}(f(\mathbf{x}, \mathbf{a}_0)) \quad (6)$$

where the Displaced Frame Difference (DFD) $\Delta(\mathbf{x}, \mathbf{a}_0) = I_n(\mathbf{x}) - I_{n-1}(f(\mathbf{x}, \mathbf{a}_0))$. Equation 6 gives an equation for the unknown update vector \mathbf{u} for every background pixel in the image. Writing the system of equations in matrix form, gives

$$G\mathbf{u} = \mathbf{z} \quad (7)$$

$$\text{where } G = \begin{bmatrix} \mathbf{x}_x \frac{\partial I_{n-1}(f(\mathbf{x}_1, \mathbf{a}_0))}{\partial x} & \dots & \frac{\partial I_{n-1}(f(\mathbf{x}_1, \mathbf{a}_0))}{\partial y} \\ \vdots & \ddots & \vdots \\ \mathbf{x}_x \frac{\partial I_{n-1}(f(\mathbf{x}_N, \mathbf{a}_0))}{\partial x} & \dots & \frac{\partial I_{n-1}(f(\mathbf{x}_N, \mathbf{a}_0))}{\partial y} \end{bmatrix}$$

$$\text{and } \mathbf{z} = \begin{bmatrix} \Delta(\mathbf{x}_1, \mathbf{a}_0) \\ \vdots \\ \Delta(\mathbf{x}_N, \mathbf{a}_0) \end{bmatrix}$$

where $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ is the set of background pixels in the image. A solution for \mathbf{u} , and hence \mathbf{a} , can be found by using a pseudo-inverse approach on eq. 7 as follows

$$\mathbf{u} = (G^T G)^{-1} G^T \mathbf{z} \quad (8)$$

By setting $\mathbf{a}_0 = \mathbf{a}$ the process can now be iterated until the estimate for \mathbf{a} converges.

4. ALGORITHM SUMMARY

The new hybrid algorithm is used to generate a mosaic from an MPEG2 stream. The mosaic is expanded at each stage by recursively registering each I-frame of the stream onto the previous I-frame. The algorithm proceeds as follows,

1. In order to generate a vector field between the I-frames, the motion vectors over a complete GOP are aggregated [5, 9].

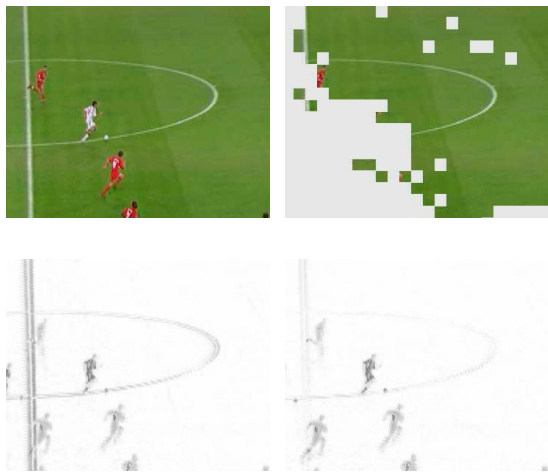


Fig. 2. Top Left: A decoded section of an I-Frame from the football sequence; Top Right: The visible parts of the frame represent the segmented background; Bottom Left: A DFD of the current I-frame and the compensated previous I-frame after the non-parametric process (dark pixels represent high differences); Bottom Right: The DFD for the hybrid algorithm.

2. The non-parametric estimation algorithm is then used to generate an estimate of the motion from the aggregated motion field and also to generate the background mask.
3. The estimate and background mask is then used to initialise the gradient based algorithm which iterates until the estimate converges.

5. EXPERIMENTAL RESULTS

Two sequences were used to test the algorithm and generate the mosaics. The first sequence comes from a football game and the second sequence from a cricket game. These sequences have been chosen to test the performance of the algorithm under different camera motions. The football sequence is characterised by a large panning while the cricket sequence exhibits a significant zooming. Mosaic for the sequences are shown in figures 4 and 5.

Figure 2 illustrates the improved estimation of the global motion parameters over a GOP for the hybrid algorithm over the non-parametric technique. The bottom left image displays the global motion compensated DFD using the parameters obtained from the non-parametric algorithm. A small error in the estimated parameters is visible in the DFD as the halfway line of the pitch has not been precisely registered from the previous frame onto the current frame. This results in the dark lines (high difference) that can be seen in the background region of the image. The bottom right image shows the DFD generated from the parameters of the hybrid approach. In this DFD, the intensity in the area of the halfway line has been reduced since the hybrid parameters more accurately describe

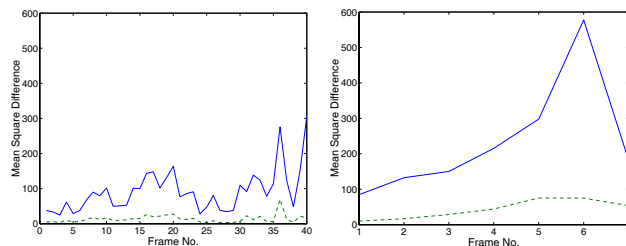


Fig. 3. Showing Mean Squared DFD of the background pixels vs frame for the football sequence (left) and cricket sequence (right). The values for the hybrid algorithm(- -) are substantially less than that of the non-parametric algorithm(-).

the global motion between the frames. Although the estimate obtained from the non-parametric algorithm is not precise, it is sufficiently accurate to allow the gradient based technique to estimate the parameters accurately when combined with the background segmentation.

Examining the results over the 2 test sequences the refinement of the hybrid algorithm significantly reduces the mean square value of the DFD in the background region. This is illustrated by the graphs in Figure 3 which show that the mean square background DFD of the two approaches for each frame of the test sequences. For most frames a reduction in value of up to 80% is obtained.

It should be pointed out that difference in the two parameter estimates is generally small in proportion to the size of the measurement. For example, for the frame shown in Figure 2 the difference between the estimated horizontal translation (t_x) for the two methods over the GOP is 1.3 pel compared to the estimate of 120.5 pel for the hybrid algorithm. Although the errors are small, they are significant for mosaicking where errors in the parameters accumulate in the mosaic over time.

6. CONCLUDING REMARKS

This paper proposed a new robust global motion estimation algorithm. It uses the non-parametric technique to generate an initial estimate of the motion and a background segmentation which are sufficient to allow the gradient based refinement to estimate the global motion over 12 frames accurately. The algorithm was applied to estimate global motion from MPEG streams in order to form mosaics. However it is important to point out that this approach can be used to generate global motion from any motion vector field. Further research in this area will focus on discrimination of motion vectors caused by erratic motions and also on the use of feature based matching so as to improve the accuracy of the algorithm.

7. REFERENCES

- [1] H.Nicholas, "New methods for dynamic mosaicing," *IEEE Transactions on Image Processing*, vol. 10, pp. 1239–1251, August 2001.



Fig. 4. Mosaics of the football sequence. Top mosaic: using non-parametric method only. In this mosaic as the camera to the right the top sideline tends to curve upwards in the mosaic; Bottom Mosaic: using the hybrid algorithm.



Fig. 5. Mosaic of the cricket sequence (far left). Middle Left to Far Right: Initially the camera shows a wide shot of the pitch. As the sequence progresses the camera zooms tightly, eventually showing a small area surrounding the batsman.

- [2] H. Sawhney and S. Ayer, "Compact representation of video through dominant and multiple motion estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, August 1996.
- [3] F. Dufaux and J. Konrad, "Efficient, robust and fast global motion estimation for video coding," *IEEE Transactions on Image Processing*, vol. 9, 2000.
- [4] J.-M. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *Journal of Visual Communication and Image Representation*, vol. 6, 1995.
- [5] R. Coudray and B. Besserer, "Global motion estimation for MPEG-encoded streams," in *Proc. of IEEE Int. Conf. on Image Processing*, Singapore, 2004.
- [6] A. Kokaram and P. Delacourt, "A new global motion estimation algorithm and its application to retrieval in sports events," in *IEEE International Workshop on Multimedia Signal Processing*, Cannes, France, 2001.
- [7] R. Coudray and B. Besserer, "Motion based segmentation using mpeg streams and watershed method," in *Int. Symposium on Visual Computing - Lecture Notes in Computer Sciences*, 2005.
- [8] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 583–598, 1991.
- [9] R. Coudray and B. Besserer, "Agregation, selection et utilisation de l'information de mouvement issue d'un flux MPEG," in *GRETSI*, 2005.