A NEW APPROACH FOR REAL TIME MOTION ESTIMATION USING ROBUST STATISTICS AND MPEG DOMAIN APPLIED TO MOSAIC IMAGES CONSTRUCTION.

Lluís Barceló, Ramon L. Felip and Xavier Binefa.

Universitat Autònoma de Barcelona (UAB)
UPIIA, Dept. Informàtica
Bellaterra, (Barcelona), Spain

ABSTRACT

Dominant motion estimation in video sequence is a task that must be often be solved in Computer Vision problems but involves a high computational cost due to the overwhelming amount of data to be treated when working in image domain. In this paper we introduce a novel technique to perform motion analysis in video sequences taking advantage of the motion information of MPEG streams and its structure, using imaginary line tracking and robust statistics to overcome the noise present in compressed domain information. In order to demonstrate the reliability of our new approach, we also show the results of its application to mosaic image construction problem.

1. INTRODUCTION

Compressed domain analysis has been a widely studied topic since the emergence of the MPEG standards. Many applications have been presented along last decade, covering a significant range of problems [1]. Compressed domain analysis helps improving the performance of algorithms by using precalculed data obtained from the streams. In this paper we present a new approach to estimate dominant motion from sequences using the MPEG motion vector field and we apply it to mosaic images construction from video sequences.

Algorithms for the construction of mosaic images consist of two main steps: registration, i.e. estimating the transformations between every pair of consecutive frames of the video, and mosaic construction, i.e. the synthesis of the mosaic image from the previously estimated transformations and the frames of the video. There are many classical methods like robust optical flow [2] and parametric methods [3] used in order to obtain the homographies between every pair of consecutives frames. It must be remarked that we process fixed-camera sequences in order to avoid parallax. All these methods operate in the pixel domain: they require calculating the optical flow between each pair of consecutive frames to later retrieve their associated transformation, a step that involves high computational cost. On the other hand, the MPEG streams contain local motion information that is already calculed[4], the MPEG motion vector field. However, the motion information provided by the MPEG streams can be noisy or inaccurate due to moving objects and compression mechanisms.

Our method uses this MPEG motion information in order to avoid the loss of performance introduced by the image domain operations, applying robust statistics to overcome the noisy samples that may exist in the MPEG streams. The advantadges of using the compressed stream are twofold. In one hand, video sequence has not to be fully decompressed because only the motion information is necessary to retrieve the transformations. Secondly, avoiding the image domain operations allows to achieve online mosaic images construction algorithms. In addition to this we use the structure of the MPEG frames, presenting a simple but powerful idea based on line tracking that yields a novel way to estimate parametric descriptions of motion between consecutive frames, resulting in a robust real time aproach that can be used as a basis to be applied along with further higher level analysis like, for instance, automatic soccer analysis [5, 6, 7], achieving a considerable speed-up using our method.

2. THEORETICAL APPROACH

We have developed a new method for registering sequences based on tracking imaginary straight lines retrieved using the fixed-grid structure of the MPEG frames macroblocks and their associated motion vectors. The data obtained using our approach is affected by noise and outliers, but we apply a high breakdown point robust estimator, the vb-QMDPE [8], to ovecome this fact. Succinctly, we have an initial set of points describing lines over the reference frame and we track these points in the target frame. Using the correspondences of these lines we can extract the projective transformation that relates both frames.

We stress that our aproach uses I and P frames only in order to build the mosaic image. This is important because

This work was funded by CICYT grant TIC2003-06075.

in addition to the mentioned improvements in terms of cost, we use only about 40% of the frames of a sequence (depending on the structure of the *GOP*). When registering a video sequence we take advantadge of the motion information provided by the stream whenever available (estimating transformations from $[I, P]$ or $[P, P]$ pair of consecutives frames). Otherwise, when having a $[P, I]$ pair we interpolate the transformation using the previous and posterior calculated transformations to overcome the lack of motion information of the I-Frames.

2.1. Straight Lines Tracking

For the purpose of retrieving the homography between two frames we have taken advantadge of the MPEG structure and data as well as a theorem that states that projective transformations keep straight lines [9].

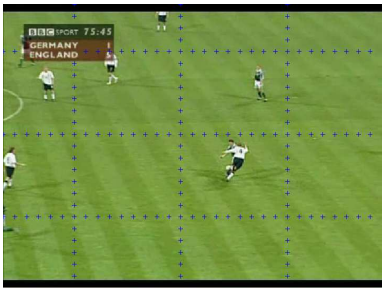
Having two frames, the processed frame and its reference frame (I_{i+1} and I_i, respectively) our aim is to obtain two sets of features that represent six imaginary straight lines in both the reference and the actual frame. The relation between the two sets of lines must be the existing motion between the pair of frames, so as to achieve that, we take advantage of the MPEG data structure, selecting as initial features f in the actual frame the top-left points corresponding to three different rows and columns of macroblocks, as seen in Figure 1a. Therefore we have six straight lines $R = [r_1, \cdots, r_6]$ called *control lines* where $r_i = (a/c, b/c, 1)^T = (t, u, 1)^T$ corresponds to a straight line of equation:

$$ax + by + c = 0 \tag{1}$$

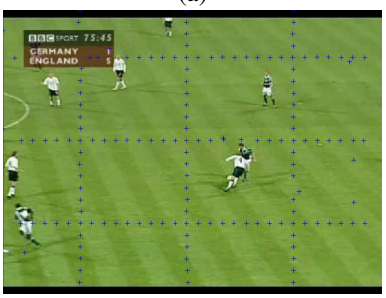
and each r_i is determined by a certain amount of image points f_{ij} (depending on the frame size) corresponding to the top-left positions of macroblocks, $r_i \subset \{f_{ij}\}$.

We want to compute the homography that relates the transformation between the frame I_{i+1} and the frame I_i. Following the straight lines theorem mentioned previously, features $\{f_{ij}\}$ for each straight line r_i in the frame I_{i+1} should also represent a straight line r'_i in the reference frame I_i. We first find the transformed features of frame I_{i+1} in I_i by means of the MPEG data and its structure. The corresponding features $\{f'_{ij}\}$ in I_i are the positions of the points $\{f_{ij}\}$ with their associated MPEG motion vector added. As result, we have a set of features $\{f'_{ij}\}$ that describe 6 straight lines $r'_i = [r'_1, \cdots, r'_6]$, where each $r'_i = (t', u', 1)$ is the transformed straight line r_i. In Figure 1 we can see the set of features F and its corresponding features F' after obtaining the corresponding motion vectors.

Finally, we must find the six straight lines that best fit F' in I_i. We must use not all the features of F', not only two (two features f'_a and f'_b would define a line, but one of them or both could be outliers). However, as seen in Figure 2b, F' might contain a high percentage of outliers, often more than 50% due to the moving objects and noise introduced by the MPEG compression, making traditional Least Squares approach unappliable. For this reason, we use the *variable bandwidth QMDPE*[8], a high breakpoint estimator, retrieving correct fits when having up to 80% of outliers. Estimation results using the vb-QMDPE are also shown in Figure 2b.



(a)



(b)

Fig. 1. a) Initial features F in frame I_{i+1}. b) The corresponding features of a), F', in frame I_i using the extracted motion vectors: we can see that the estimations are affected by the moving objects (the players) and for the superimposed scoreboard.

2.2. Homography Estimation

Once the six transformed lines $R' = [r'_1, \cdots, r'_6]$ in the frame I_i are obtained, the mentioned homography theorem [9] is applied in the following way:

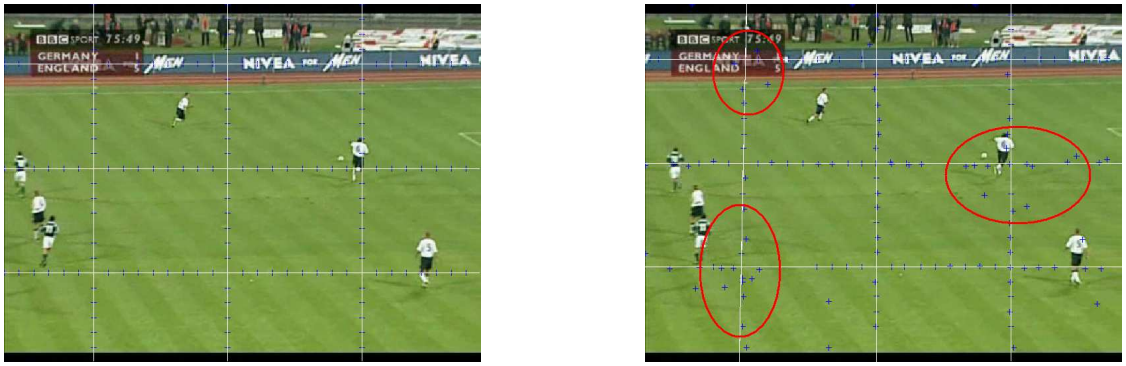
$$r'_i = (H^{-1})^T r_i \tag{2}$$

where $r_i = (t, u, 1)^T$ and H is the homography represented by a non-singular 3×3 matrix:

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \tag{3}$$

Switching left with right in Equation (2) we obtain:

$$r_i = H^T r'_i \tag{4}$$



(a) (b)

Fig. 2. Image (a) shows the initial features F and straight lines retrieved from the MPEG structure and (b) contain the features plus their motion vector F' and the line estimation results using vb-QMDPE . The outliers that make a Least Square approach unfeasible are highlighted in (b).

Therefore, each line correspondence in the plane provides two equations for each one of the 8 unknown entries of H:

$$\begin{aligned} t(h_{13}t' + h_{23}u' + 1) &= h_{11}t' + h_{21}u' + h_{31} \\ u(h_{13}t' + h_{23}u' + 1) &= h_{12}t' + h_{22}u' + h_{32} \end{aligned} \quad (5)$$

It is necessary to find at least four line correspondences to define a unique projective transformation matrix, up to a scale factor. However, six lines have been used in order to make the estimation more robust because video sequences are very likely to contain multiple variable size moving objects. The equations of (5) can be rearranged in matrix form, obtaining the following equation system:

$$\begin{bmatrix} t'_i & 0 & -t_i t'_i & u'_i & 0 & -t_i u'_i & 1 & 0 \\ 0 & t'_i & -u_i t'_i & 0 & u'_i & -u_i u'_i & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} = B$$

where B is equal to:

$$B = \begin{bmatrix} t_i \\ u_i \\ \vdots \end{bmatrix}$$

Finally, the solution of the above system equation, found using a robust method (*Least Median Squares*[10]), yields the homography that relates the transformation between the frames I_i and I_{i+1}. This procedure is performed for each pair of related I or P frames of the MPEG sequence until we process the whole sequence.

3. EXPERIMENTAL RESULTS : MOSAIC IMAGE SYNTHESIS

For test purposes, we have applied our approach to mosaic images construction. This problem is perfect to test the robustness and accuracy of our algorithm because a single misestimation involves an error that is propagued along the sequence. The procedure followed is next explained and results of mosaic images of video sequences are shown in Figure 3, and in Figure 4 we can see some frames of the used sequences. Once the whole sequence has been processed all the transformations between related frames are available. However, in order to build the mosaic image all the frames must reference the same initial frame. For this reason, we calculate firstly the cumulative transformation of each frame with respect to the reference frame, in our case, the first frame of each sequence. This task is carried out by multiplying the transformation matrices leftwise:

$$\begin{aligned} H_{11} &= I_{3\times 3} \\ H_{12} &= H_{11}H_{12} \\ H_{13} &= H_{11}H_{12}H_{23} = H_{12}H_{23} \\ &\vdots \\ H_{1n} &= H_{11}H_{12}H_{23}\cdots H_{n-1n} = H_{1n-1}H_{n-1n} \end{aligned} \quad (6)$$

where H_{ij} is the homography between the frames I_j and I_i (the *cumulative transformation* between the frames). In order to construct the final mosaic we transform each frame using its corresponding cumulative transformation and we apply a mean or median operator in order to obtain the mosaic using the whole transformed frames. Online real time mosaics are obtained using the actual frame to build the mosaic instead of using the median or mean operator.

Fig. 3. Results of the construction of mosaic images in compressed domain using our approach.

Fig. 4. Some frames from the sequences used to test the construction of mosaic images in compressed domain using our approach.

4. CONCLUSIONS

We have developed a new method to register video sequences and obtain parametric descriptions of the existing motion using motion data available from the MPEG streams. The contributions of our method with respect to the traditional approaches are many. In one hand we achieve a significant improvement in terms of computational cost when comparing our approach with existing pixel domain techniques. On the other hand, the use of imaginary lines from the MPEG motion vector field structure and robust estimators gives our approach plenty of reliability, overcoming the inaccuracies and noise that exist in MPEG data. In addition to this, the use of compressed domain data allows us to estimate motion in sequences with frames of large dimensions, because the number of motion vectors to handle is far lower than the amount of pixels in the traditional image domain techniques (1:64).

We emphasize our dedication to the robustness of the method. We have taken many measures and used several robust estimators (vbQMPDE and LMedS) in critical parts in order to make our approach reliable although we use information from the compressed domain. As a result, our method has been applied to construct mosaic images with excellent results, proving that is a reliable technique to work as a basis for further analysis in problems where dominant motion must be estimated and yields the possibility to develop online real time systems thanks to the saving of computational cost of our approach.

5. REFERENCES

[1] H. Wang et al., "Survey of compressed-domain features used in audio-visual indexing and analysis," *Journal of Visual Communication and Image Representation*, vol. 14, pp. 150, June 2003.

[2] Michael J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding: CVIU*, vol. 63, no. 1, pp. 75–104, 1996.

[3] Michael J. Black and P. Anandan, "A framework for the robust estimation of optical flow," in *Fourth International Conf. on Computer Vision*, 1993, pp. 231–236.

[4] *ISO/IEC IS 13818-2, MPEG-2 Video*, 2000.

[5] J. Assfalg, M. Bertini, A. Del Bimbo, W. Nunziati, and P. Pala, "Soccer highlights detection and recognition using hmms," *in Proc. IEEE Int. Conf. Multimedia and Expo*, 2002.

[6] Ahmet Ekin and A. Murat Tekalp, "Automatic soccer video analysis and summarization," *Symp. Electronic Imaging: Science and Technology: Storage and Retrieval for Image and Video Databases IV*, 2003.

[7] D. Yow, B.-L. Yeo, M. Yeung, and B. Liu, "Analysis and presentation of soccer highlights from digital video," *in Proc. Asian Conf. Computer Vision*, 1995.

[8] Hanzi Wang and David Suter, "Mdpe: A very robust estimator for model fitting and range image segmentation," *International Journal of Computer Vision (IJCV)*, 2004.

[9] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[10] Peter J. Rosseeuw and Annick M. Leroy, *Robust Regression and Outlier Detection*, Wiley-Interscience, 1987.