A NEW TWO-STAGE GLOBAL/LOCAL MOTION ESTIMATION BASED ON A BACKGROUND/FOREGROUND SEGMENTATION

Fabrice Moscheni †, *Frédéric Dufaux* ‡ *and Murat Kunt* †

†Signal Processing Laboratory, Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland

‡The Media Laboratory, Massachusetts Institute of Technology
Cambridge, MA 02139, USA

ABSTRACT

In the framework of sequence coding, motion estimation and compensation has been shown to be very efficient at removing temporal redundancy. The motion existing in a scene can be mainly seen as arising from local motions superimposed to the camera motion. In this paper, a new two stage global/local motion estimation approach is presented. The global motion estimation only relies on the background information. It is based on a matching technique and the global motion model is chosen to be affine. Simulation results show significant improvements obtained with the proposed method compared to usual methods.

1. INTRODUCTION

In video coding, the reduction of temporal redundancy is the key to reach high performances. Amongst the available techniques, motion estimation and compensation has been shown to be a most efficient method. It basically tries to predict the current frame from the previous one by estimating the motion between the two frames. Hence, the motion and prediction error information, also referred to as Displaced Frame Difference (DFD), are transmitted instead of the image itself.

Most generally, the motion arising in a sequence is a combination of global and local motions. Whereas the latter results from the displacement of objects in the scene, the former is produced by the motion of the camera. The estimation of the camera motion allows to remove the global motion in the scene, resulting in a more precise subsequent local motion estimation. Furthermore, extracting the global component of the motion field leads to a reduced amount of side information to transmit the motion vectors.

For the above reasons, global motion estimation algorithms have been proposed in [1, 2, 3, 4]. These algorithms differ in the model to represent the camera motion, as well as in the technique to estimate the parameters of the chosen model. In [2], the camera motion is modelled by 4 parameters corresponding to a pan, a zoom and a rotation, whereas in [1, 3, 4] a simpler model involving only 3 parameters is used corresponding to a pan and a zoom. As far as the estimation of these parameters is concerned, the algorithms in [1, 2] are based on a differential technique, the method in [3] relies on a frame matching technique and, finally, the algorithm in [4] performs a linear regression on an estimate of the local motion.

It is obvious that the distinction between the background and foreground on one hand, and the computation of the global motion on the other hand are intimately related. In particular, the global motion parameters should be evaluated only on the background in order to prevent errors resulting from local motions. Furthermore, the delimiting of the background allows for improved performances in the context of video coding, as techniques such as background memory [5] can be efficiently applied. However, the distinction between background and foreground is not explicitly made in the algorithms proposed in [1, 2, 3, 4].

Taking into account the above considerations, this paper proposes a two-stage global/local motion estimation which explicitly separates the background and the foreground. The background/foreground mask as well as an initial guess for the global motion parameters are obtained thanks to either a clustering technique performed on a local motion field or the tracking of the previous mask and parameters. The global motion estimation is then carried out on the background. In a further stage, a global motion compensation is carried out on the whole frame in order to remove the camera motion component. A local motion estimation is then performed on the foreground regions.

In this paper, an affine global motion model (i.e. 6 parameters) is chosen. In order to estimate the latter, a matching technique is applied which only takes into account pixels belonging to the background. In the following, it is referred to as background matching. To

decrease the computational complexity and to allow a non-exhaustive search while avoiding local minima, a Gaussian pyramid structure of the input images is introduced [6]. Even though the proposed background matching technique may be time consuming, it is very robust and simulation results show that it outperforms differential techniques [1, 2] as well as methods based on a linear regression of a local motion field [4].

This paper is structured as follows. In Sec. 2, the two-stage global/local motion estimation based on a foreground/background segmentation is described. The background matching as well as alternative techniques for global motion estimation are presented in Sec. 3. Experimental results are given in Sec. 4, and Sec. 5 draws conclusions.

2. TWO-STAGE GLOBAL/LOCAL MOTION ESTIMATION

In the framework of motion compensated coding, the ability to distinguish global from local motion is highly desired. It indeed allows for a more precise estimation and representation of the motion in the scene, and permits to dramatically reduce the overhead motion information.

The background is defined as the ensemble of pixels which are not subject to local motion, namely which only undergo the camera motion. Therefore, the global motion parameters should only be evaluated on those pixels in order to prevent errors arising from local motions. The distinction between the background and foreground on the one hand, and the computation of the global motion on the other hand are therefore intimately related.

This paper proposes a new approach to motion estimation and compensation. Not only a two-stage global/local motion estimation is introduced, but also the foreground is distinguished from the background. Therefore, the technique allows for a more precise estimation of the camera motion. Figure 1 shows a block diagram of the proposed algorithm.

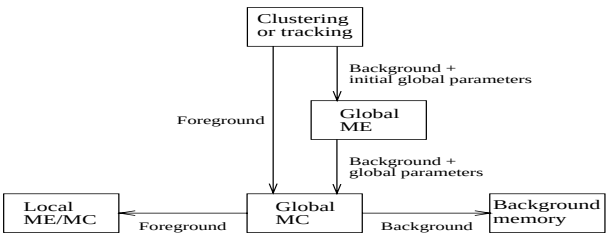

Figure 1: Block diagram of the proposed two-stage global/local motion estimation.

In order to define the background/foreground mask, the tracking of the mask obtained for the previous frame is put into competition with a clustering technique. The clustering is applied to the motion field obtained by a local motion estimation technique (e.g. block matching). In our implementation, the cluster analysis is performed by the k-means algorithm [7], the number of clusters being set to two. In this way, the motion vectors are separated into two groups based on the cluster to which they belong. The larger of these two groups is then identified as background and the second one as foreground, thus defining the background/foreground mask. In addition, the centroid of the cluster corresponding to the background provides an initial estimate for the global motion parameters. The complementary approach is to track the parameters obtained from the previous frame. The approach providing the best initial estimate in terms of Mean Square Error (MSE) is thus selected.

Different global motion models have been considered in the literature. In [1, 3, 4], a model involving only a zoom and a pan is chosen, whereas a rotation parameters is added in [2]. The latter are however quite inefficient when the global motion occurs on a scene with changing depth. Due to the large dimension of the background, a more complex model has to be used. Modelling the background as being a plane and assuming perspective projection, an eight-parameter quadratic motion model reveals itself as necessary [8]. Under the further hypothesis of distant shallow plane, the perspective projection is closely approximated by the orthographic projection. With the latter projection, the motion of a plane is represented by an affine model.

The global motion estimation is then carried out only on the portion of the sequence corresponding to the background. Thus it prevents bias due to local motions and results in more accurate camera motion parameters. Furthermore, the initial estimates provided by the clustering/tracking technique has proved to be very useful in order to improve the global motion estimation. The global motion existing in the scene is then removed through global motion compensation. The latter compensation is performed on the whole image so as to remove the camera motion component in the whole scene. Due to its robustness and ability to obtain the true motion parameters, a global motion estimation based on background matching has been chosen (see Sec. 3).

After the global motion compensation, the local motion estimation is finally performed on the foreground regions. Any classical local motion estimation technique (e.g. block matching) can be used to that end. The global motion being removed, the motion in the foreground is only due to the displacements of the objects in the scene.

3. GLOBAL MOTION ESTIMATION

In this section, the estimation of the global motion parameters is addressed in more details. As described is

Sec. 2, the global motion model is chosen to be affine and the estimation is only performed on the background. Amongst the available techniques for global motion estimation, three main approaches can be distinguished. Namely, there are the matching approach, the differential approach and the approach based on a regression of a local motion field. Implementations of the three global motion estimation approaches are presented hereafter. A comparison of their characteristics is then carried out. From such an analysis, the matching approach clearly confirms itself as the most appropriate for global motion estimation in the framework of video coding. This will be confirmed by simulations results in Sec. 4.

All the global motion estimation techniques described in this section are embedded in a multidimensional framework. To that end, a Gaussian pyramid structure of the input images is used [6]. The final motion parameters at one level propagate as initial parameters to the next level. At the top level, the initial estimates are provided by the clustering/tracking technique. Finally, in the case of the differential approach as well as in the case of the approach based on the regression of the local motion field, refinements of the estimates are carried out at each level by motion compensated iteration [2]. This is equivalent to optimizing the estimates through a Gauss-Newton minimization algorithm.

Both the matching and differential techniques can be seen as least squares minimizations of the motion parameters estimation problem. As the minimization is carried out on the signal, both can also be considered as be direct techniques. The first approach to global motion estimation is the proposed background matching technique. It is a generalization of block matching motion estimation to the whole background. Similar to frame matching [3], its specificity lies in the exclusive use of the background information. Without making any assumption on the luminance signal, it implicitly derives the motion parameters by directly minimizing the reconstruction error.

For the least squares minimization turns out to be non-linear, the differential technique of global motion estimation uses a model of the luminance signal. Relying either on a first order luminance approximation [8] or a second order approximation [1, 2], it is thus able to derive an explicit minimization problem. The minimization process is then carried out only on the pixels belonging to the background.

Finally, the last technique carries out the estimation process in two steps [4]. In a first stage, a local motion field is computed. The global motion parameters are then computed by regression on the latter local displacement vectors. Only the motion vectors of background pixels are taken into account in the regression. This technique can be seen as indirect as it does not compute the motion parameters from the luminance signal itself.

Although the three techniques rely on a similar hierarchical framework, the motivations are different. In the case of the differential technique as well as for the technique performing a regression on a local motion field, the multiresolution structure aims at making the estimates more robust and accurate. In the case of the matching technique, the purpose is however to reduce the computational load by allowing a non-exhaustive search while avoiding local minima.

Moreover, the differential technique may be put into trouble due to its model of the luminance signal. The consistence of such an approximation significantly influences the performance of the estimation process. In particular, noise and large motions may be problematic. Regarding the technique performing a regression on a local motion field, inaccuracies in its computation may alter the estimation process, leading to spoilt motion parameters. Conversely to the two other approaches, the background matching technique assumes no model of the luminance signal and carries out the minimization process on the latter. It is thus characterized by its resilience to noise as well as being able to find the true camera motion.

4. EXPERIMENTAL RESULTS

In a first step, the global motion estimation techniques described here above are compared. The motion model is chosen to be affine. This latter choice will be corroborated by the simulations comparing global motion models with respectively 2, 4 and 6 parameters. Finally, the performances of the new two-stage global/local motion estimation based on a background/foreground segmentation are compared with a classical local motion estimation. In all the simulations, the local motion field is computed by multigrid motion estimation [9]. Simulations have been carried out on the luminance component of the sequence "Table Tennis" in CIF format.

Figure 2 compares the proposed background matching technique with the other two techniques. The Mean Square Error on the background is reported for each frame. The background matching performs significantly better than the two other approaches, in particular on the last frames which correspond to an increased camera motion. In case the DFD was coded, the noise introduced in such manner would further widen the gap of performance. The background matching technique reveals itself as being stable and able to robustly estimate the global motion parameters.

As far as the global motion models are concerned, Fig. 3 reports the performances obtained with motion models which have respectively 2, 4 and 6 parameters. The background matching technique was used for the global motion estimation. The affine model is clearly shown to outperform the others. Its use is therefore

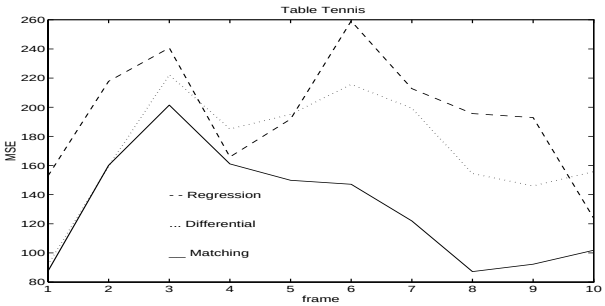

Figure 2: Comparison of the three approaches to global motion estimation: matching technique, differential technique and regression on a local motion field.

necessary in order to cope with complex camera motion. Such a result confirms the approximation of an orthographic projection of a plane.

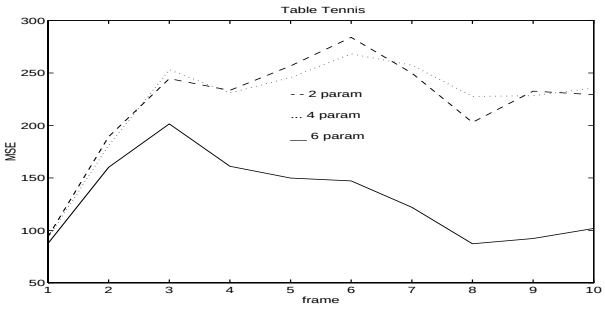

Figure 3: Comparison of the respectively 2, 4 and 6 parameters model for camera motion.

Finally, Fig. 4 compares the proposed motion estimation with a classical local motion estimation [9]. For each frame, the total bit rate (i.e. motion information and DFD) is assessed in terms of its entropy. The motion information and DFD are respectively quantized with steps of 0.5 and 20. The proposed method achieves significantly better performances than the classical local method.

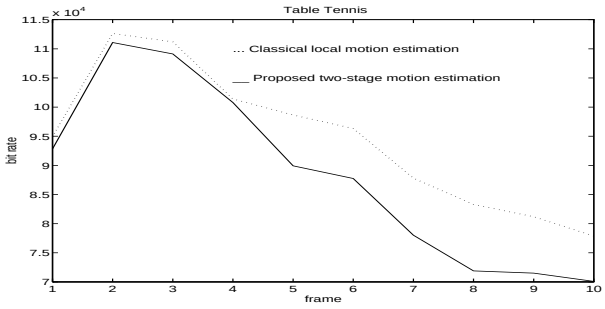

Figure 4: Comparison between the proposed two-stage motion estimation and a classical local motion estimation.

5. CONCLUSION

In this paper, a new two-stage global/local motion estimation is proposed. The global motion is represented by a six parameters affine model. The motion parameters are estimated by a background matching technique which only takes into account pixels of the background. Experimental results have shown that the matching technique outperforms differential and regression techniques, and that an affine model is required to cope with complex camera motion. Finally, it has been shown that the two-stage global/local motion estimation outperforms the classical local motion estimation.

6. REFERENCES

[1] M. Hoetter. Differential estimation of the global motion parameters zoom and pan. *Signal Processing*, vol. 16, pp. 249-265, March 1989.

[2] S.F. Wu and J. Kittler. A differential method for simultaneous estimation of rotation, change of scale and translation. *Signal Processing: Image Communication*, vol. 2, no. 1, pp. 69-80, May 1990.

[3] D. Adolph and R. Buschmann. 1.15 Mbit/s coding of video signals including global motion compensation. *Signal Processing: Image Communication*, vol. 3, nos. 2-3, pp. 259-274, June 1991.

[4] Y.T. Tse and R.L. Baker. Global zoom/pan estimation and compensation for video compression. In *IEEE Proc. ICASSP'91*, volume IV, pages 2725–2728, Toronto, Canada, May 1991.

[5] D. Hepper. Efficiency analysis and application of uncovered background prediction in a low bit rate image coder. *IEEE Trans. Commun.*, vol. COM-38, no. 9, pp. 1578-1584, September 1990.

[6] P. J. Burt and E. H. Adelson. The laplacian pyramid as a compact image code. *IEEE Trans. Commun.*, vol. COM-31, no. 4, pp. 482-540, April 1983.

[7] L. Kaufman and P.J. Rousseeuw. *Finding groups in data.* John Wiley&Sons, Inc., New York, 1990.

[8] P. Anandan, J.R. Bergen, K.J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In M.I. Sezan and R.L. Lagendijk, editors, *Motion Analysis and Image Sequence Processing*, pages 1–22. Kluwer Academic Publishers, 1993.

[9] F. Dufaux. *Multigrid Block Matching Motion Estimation for Generic Video Coding.* PhD thesis, Swiss Federal Institute of Technology, Lausanne, 1994.