COMBINED MOTION AND EDGE ANALYSIS FOR A LAYER-BASED REPRESENTATION OF IMAGE SEQUENCES

Federico Pedersini, Augusto Sarti and Stefano Tubaro

Dipartimento di Elettronica e Informazione, Politecnico di Milano
Piazza L. da Vinci, 32, 20133 Milano, Italy
pedersin, sarti, tubaro@elet.polimi.it

ABSTRACT

In this paper we propose a method for the motion-congruent segmentation of image sequences based on both motion field and luminance information. In order to do so, the affine motion models are determined by analyzing the motion field through a clustering procedure, while their regions of validity are determined by an MRF-based region estimator. This last block performs a pixel-wise re-assignment of a limited number of affine models (those determined via clustering).

1. INTRODUCTION

In order to reduce the redundancy of image sequences through motion compensation, what we need to exploit is the fact that some regions of the image move coherently across time [1]. The most recent coding techniques based on motion compensation, in fact, are based on the motion-congruent segmentation of image sequences, i.e. on the partition of the image into regions of coherent motion [1, 2].

The motion of an object in a 3D scene results in a coherent motion field in the corresponding 2D image sequence. The 3D structure of the scene, however, is usually unknown, therefore the concept of coherence of motion field must be specified in terms of the characteristics of the 2D sequence. We consider a motion field to be coherent in a certain region when there exists a parametric model of a specified class that describes synthetically and accurately the motion field inside that region. The parametric models that are used in the literature are usually linear (affine), and the estimation of the region shapes and of their motion parameters is often carried out jointly [1, 2]. The motion field, however, is usually affected by errors especially at the boundaries of occluding objects and on non-textured areas. As a consequence, a method that performs a joint model/region estimation by relying entirely on motion field ends up requiring sophisticated and computationally heavy clustering procedures in order to obtain reasonably correct segmentations.

Our approach to motion segmentation is based on the fact that, when the scene is made of a reasonably small number of large regions of coherent motion, a motion model can be accurately estimated by performing linear regression on just a portion of its region of validity rather than on the whole one. In other words, we may exclude those image regions where the motion field is not reliable, and proceed with a joint model/region estimation. Giving up on those portions of the regions that are normally difficult to classify allows us to focus more on the models than on the accuracy of the regions. Accurate region estimation, in fact, can be performed later by taking the models as granted and using motion field, luminance information and morphological shape constraints for obtaining a pixel-wise re-assignment of the previously estimated affine models.

2. MOTION SEGMENTATION

The motion-congruent segmentation technique we propose in this paper differs from those of [1] and [2] in the fact that affine motion models and their regions of validity are determined in two different phases. With reference to the scheme of Fig. 1, in fact, the motion field is used for an accurate estimation of a limited number of affine motion models and an initial estimation of their region of validity. This operation is possible through clustering in the space of the affine parameters. The regions are determined afterwards by using an MRF-based region estimator which is able to operate a pixel-wise re-assignment of the affine motion models by using motion field as well as luminance information. Models are re-assigned in such a way to minimize a potential function that tends to homogenize the flow field inside object regions. In order to do so, the region of influence of the potential function tends to "shrink" along object edges. The process stops when the regions reach a stable configuration.

0-7803-3258-X/96/$5.00 © 1996 IEEE

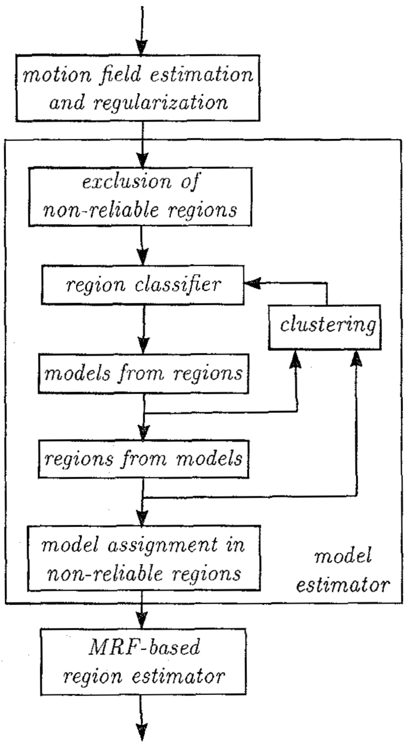

Figure 1: Overall organization of the algorithm.

It is important to emphasize the fact that, unlike other MRF-based motion segmentation techniques (see for example [2]) available in the literature, our region identification block works on a limited search space, as the selection must be done only among those affine motion models that are provided by the model identification block.

Though the region identification algorithm could start from an arbitrary configuration of regions, it is reasonable to initialize the region estimator with the regions provided by the clustering algorithm. By doing so, we speed-up the convergence of the whole process and reduce the risk of encountering relative minima of the potential function.

2.1. Motion Field Estimation

The motion estimator we employ is a multi-scale coarse-to-fine algorithm based on a gradient approach [3]. In order to improve the reliability of the motion field estimation, both backward and forward flow fields are determined and compared. Scales go progressively from a block-size of 32×32 pixel down to 2×2 pixel. For each scale, we apply a least-squares technique to each block by solving the following equation

$$I_t(x,y) = I_{t-1}(x - \hat{v}_x - \Delta v_x, y - \hat{v}_y - \Delta v_y) ,$$

where I_t and I_{t-1} represent the current and the previous images, and $(\hat{v}_x, \hat{v}_y)$ is an estimate of the block displacement obtained in the previous step of the algorithm. $(\Delta v_x, \Delta v_y)$ is the update on the motion field relative to the new resolution level.

Motion field will be used for determining a limited set of affine motion models that are suitable for describing the evolution of the objects of the imaged scene. We assume that the scene contains a limited number of large regions of homogeneous motion therefore regularizing the motion field through median filtering can be an inexpensive way of improving the estimation quality while reducing the convergence speed of the clustering procedure.

2.2. Model identification

In order to identify a limited set of motion models, we do not need motion vectors to be assigned to all image points, as long as they include the most significant information about the models to be estimated. This fact allows us to discard unreliable portions of the motion field after having identified them through a thresholding procedure. More precisely, all motion vectors that cause the Motion Compensated Luminance Difference (MCLD) to be above a certain threshold, are declared unreliable, therefore they will not be used for model identification.

As stated above, we need to find an approximation $\hat{\mathbf{v}}$ of a given motion field $\mathbf{v} = [v_x \ v_y]^T$, of the form

$$\hat{\mathbf{v}}(x,y) = \sum_{k=1}^{N} I_k(x,y)\mathbf{v}^{(k)}(x,y) \ , \ I_k = \begin{cases} 1 & (x,y) \in R_k \\ 0 & \text{elsewhere,} \end{cases}$$

R_k being the region of validity of the k-th affine model

$$\mathbf{v}^{(k)}(x,y) = \begin{bmatrix} v_x \\ v_y \end{bmatrix}^{(k)} = \begin{bmatrix} a_{xx} & a_{xy} \\ a_{yx} & a_{yy} \end{bmatrix}^{(k)} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix}^{(k)}$$

The task of the model identification block is that of determining a limited set of affine parameters $\mathbf{a}^{(k)} = [a_{x0} \ a_{xx} \ a_{xy} \ a_{y0} \ a_{yx} \ a_{yy}]^{(k)}$ and their approximate regions of validity R_k. This can be done iteratively as shown in Fig. 1.

Models from regions – The affine model that best approximates the motion field $\mathbf{v}$ in a given region R_k, as discussed in [1], can be computed with a standard linear regression approach

$$\mathbf{a}^{(k)} = \left(\sum_{R_k} \phi^T \phi \right)^{-1} \sum_{R_k} \phi^T \ [v_x(x,y) \ v_y(x,y)] ,$$

where $\phi = [1\ x\ y]$ is the regressor. The corresponding residual error

$$\sigma_k^2 = \frac{1}{N_k} \sum_{R_k} \left| \mathbf{v}(x,y) - \mathbf{v}^{(k)}(x,y) \right|^2$$

can be used as a measure of the model reliability.

Regions from models – Once a new collection of affine models is available, we determine their region of validity by selecting, for each image point the affine model that minimizes the MCLD. The selection is made among the affine models of the available collection.

Clustering – We combine affine parameters that are close to each other (i.e. that are likely to describe the motion of the same object) through clustering (a modification of Forgy's method [4]) in the parameter space. This operation allows us to reduce the number of models in the scene through model merging.

Region classifier – The clustering block is only able to merge models while, depending on the scene evolution, we might need to split some of them when the parameters drift too much apart from their centroid. In order to do so, the region classifier performs some checks on the shape of the available regions ad decides whether to split them according to a few morphological constraints. In particular, the classifier will take into account their size and mutual connection. The region classifier is allowed to declare some region fragments as "unclassified" as an accurate and complete region segmentation will be performed afterwards.

The algorithm stops when the model configuration becomes stable. The initial segmentation is provided by the motion-compensated segmentation of the previous image. For the first image of the sequence we can choose an arbitrary segmentation. Notice that the number of regions may decrease as well as increase during the evolution of the imaged scene.

Last step of the model identification block consists of assigning a model to all regions that had been previously ruled out as non-reliable or labeled as "non classified". This operation is performed, once again, by choosing among the available affine models the one that minimizes the MCLD.

2.3. Region identification

As we have already pointed out earlier, the regions detected in the previous step are only accurate enough to make the estimate of the relative affine models reliable. The actual region identification is achieved through an MRF-based regularization approach [5, 2].

The process consists of a point-wise minimization of a global energy function of the form

$$W^{(k)}(x,y) = \alpha U^{(k)}(x,y) + V^{(k)}(x,y) , \qquad (1)$$

with respect to the models $\mathbf{a}^{(k)} \in \mathcal{A}$, $\mathcal{A}$ being the *finite* collection of affine parameters determined by the model identification block. The first term of eq. (1) is the MCLD corresponding to the affine parameters $\mathbf{a}$

$$U^{(k)}(x,y) = \left(I(x,y) - I(x - v_x^{(k)}, y - v_y^{(k)}) \right)^2 ,$$

while the second term is a measure of the motion field "smoothness" with reference to a neighborhood that tends to "squeeze" against object contours:

$$U^{(k)}(x,y) = \sum_{\{p,q\} \in \mathcal{R}'} ||\mathbf{v}^{(k)}(x,y) - \mathbf{v}(p,q)||^2 ,$$

where $\mathcal{R}'$ is obtained by excluding from a given second-order neighborhood $\mathcal{R}$ of (x,y) those points that lie "beyond" significant edges that pass through $\mathcal{R}$ (with respect to (x,y)). With this choice, the points that are likely to correspond to different objects, will tend to have independent motion vectors. In order to be able to deal with the lack of motion field information in those image areas that are uncovered with respect to the previous frame, we take into account the "forward" motion field as well.

3. SIMULATION RESULTS

The proposed segmentation technique has been tested on the CIF sequences "Flower Garden" (Fig. 2) and "Table Tennis" (Fig. 4). The quality of the resulting segmentation results as being particularly good near edges that constitute the object boundary. The fact that regions tend to approximate the shape of individually moving objects emphasized by Fig. 3, where the region corresponding to a tree has been removed. The background is now completely visible thanks to an accumulation mechanism of the type described in [1].

When the global motion (such as zoom and/or pan) is a dominant portion of the whole motion field, it might be difficult to perform a correct segmentation. In the "Table Tennis" sequence, for example, the strong zooming effect tends to "mask" the local motion. In order to overcome this difficulty, it is necessary to preliminarily estimate the global motion and then compensate for it. The results of such an operation are visible in Fig. 4.

4. CONCLUSIONS

In this paper we proposed and tested a new technique for performing a motion-congruent segmentation of im-

(a) (b) (c)

Figure 2: Segmentation of the "Flower Garden" sequence. *a)* Regions obtained through clustering (model estimator), *b)* regions produced by the MRF-based region estimator, *c)* layer-based frame prediction.

Figure 3: "Flower Garden" sequence after removal of the layer corresponding to the tree.

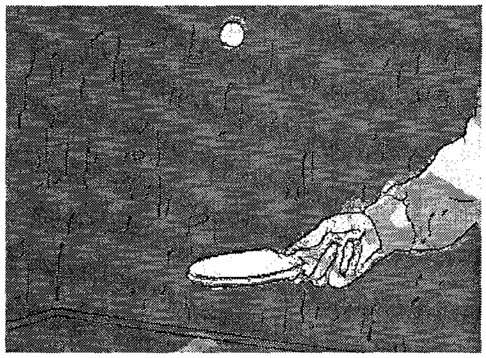

Figure 4: Segmentation of the "Table Tennis" sequence after global motion compensation.

age sequences based on motion field and luminance. The method can provide high-quality segmentations, especially in those cases in which the scene objects have visible borders and are textured enough. We are currently working on improvements for the global motion estimation, for the layer accumulation and the statistical determination of the segmentation parameters.

5. REFERENCES

[1] J.Y.A. Wang, E.H. Adelson: "Representing moving images with layers." *IEEE Trans. on Image Processing*, Vol. 3, No. 5, Sept. 1994.

[2] J.M. Odobez, P. Beaulieu: "MRF-based motion segmentation exploiting a 2D motion model robust estimation" IEEE - ICIP '95. Washington D.C., USA, Oct. 23-26, 1995. Vol. III, pp. 628-631.

[3] J.R. Bergen, P.J. Burt: "A three-frame algorithm for estimating two-component image motion." *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 9, Sept. 1992, pp. 886-895.

[4] R. Dubes, A.K. Jain: "Clustering techniques: the user's dilemma", *Pattern Recognition*, Vol. 8, pp. 247-260.

[5] S. Geman, D. Geman: "Stochastic relaxation, Gibbs distribution, and the bayesian restoration of images." *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 11, Nov. 1984, pp. 721-741.