A REGION-BASED ALGORITHM FOR IMAGE SEGMENTATION AND PARAMETRIC MOTION ESTIMATION*

Camilo C. Dorea, Montse Pardàs, Ferran Marqués

Technical University of Catalonia, Barcelona, Spain
{camilo, montse, ferran}@gps.tsc.upc.es

ABSTRACT

This paper describes an approach for integrating region-based motion estimation and region merging techniques with the purpose of obtaining precise parametric motion description and image segmentation. Segmentation is achieved with a region merging scheme based initially on color homogeneity and extended to include motion parameters in successive steps. Motion vectors are first estimated with a multiresolution gradient-based method applied directly over the input images and then refined by incorporating segmentation results into a region-based block matching scheme. The complete algorithm presents motion boundaries that coincide with color boundaries. The resulting region-based, affine motion models respect motion boundaries and provide accurate motion description even over small support regions. The method is illustrated with experimental results for an image sequence.

1. INTRODUCTION

Motion estimation and segmentation are important sources of information for many applications in multimedia and video analysis. Motion estimation is concerned with the estimate of the motion parameters of a moving object while motion segmentation attempts to identify the boundary of these objects. Both of these problems are directly related and a number of solution methods have been presented. Methods based on clustering of motion parameters, such as [1], do not take into account the spatial constraints of the image. Recently, techniques based on Binary Partition Trees [2] have provided an efficient means of segmenting objects whose boundaries coincide with the motion boundaries. Motion criteria have also been applied to merging of spatial partitions in [3] and labeling of watershed segments in [4] and [5] for video segmentation.

In numerous analysis applications, segmented regions should exhibit spatial and temporal homogeneity characteristics defining objects that have some physical meaning. Furthermore, object mask contour errors must be minimal and motion description of the image plane must be precise enough not only to aid in the spatial segmentation of the following frames but also to allow inference of semantics. In this paper we propose a novel algorithm for estimating parametric motion and segmenting motion objects based on an initial color segmentation. Accuracy in motion estimation is achieved with a region-based estimation scheme and the resulting motion models are used in motion segmentation.

*This material is based upon work partly supported by the IST program of the EU in the project IST-2000-32795 SCHEMA, by the grant TIC2001-0996 of the Spanish Government and by CAPES - Brazilian Government.

2. ALGORITHM OUTLINE

Both motion estimation and image segmentation require a priori knowledge of each other to produce successful results. The algorithm initially decouples spatial and motion information, i.e., possible boundaries for the objects are proposed based purely on spatial information and motion parameters are estimated without introducing any partition knowledge. The resulting information from each of these procedures is used to reinforce the other and refine results in subsequent segmentation and estimation steps, as depicted in the diagram of figure 1.

An initial segmentation of image $I(t)$ is achieved with color-based region merging. The result is a fine partition of the image into regions with color homogeneity where region sizes are kept small. We assume a region of uniform motion will be composed of one or more sub-regions each of which possessing uniform color, consequently, the motion boundaries will be a subset of the color boundaries determined at this stage.

Motion information will be initially represented through a dense motion vector field, i.e., estimates which best relate the position of each pixel in successive image frames. For the task at hand we adopt a gradient-based approach [6] which can quickly provide dense optical flow information with sub-pixel accuracy. This technique is implemented within a multiresolution framework, allowing estimation of a wide range of displacements. One of the drawbacks of such a scheme, however, is the rough estimates at motion boundaries due to the use of image gradients and fixed support regions.

To tackle this problem, motion detection via thresholding of the estimated motion vectors is carried out in order to define some areas of interest from the initial partition. This definition serves a dual purpose. First, the boundary points of the regions classified as moving will undergo a motion estimation refinement with a region-based block matching scheme which incorporates the segments available via the fine partition. Second, the selected regions will form a motion mask over the fine partition and color-based merging will be

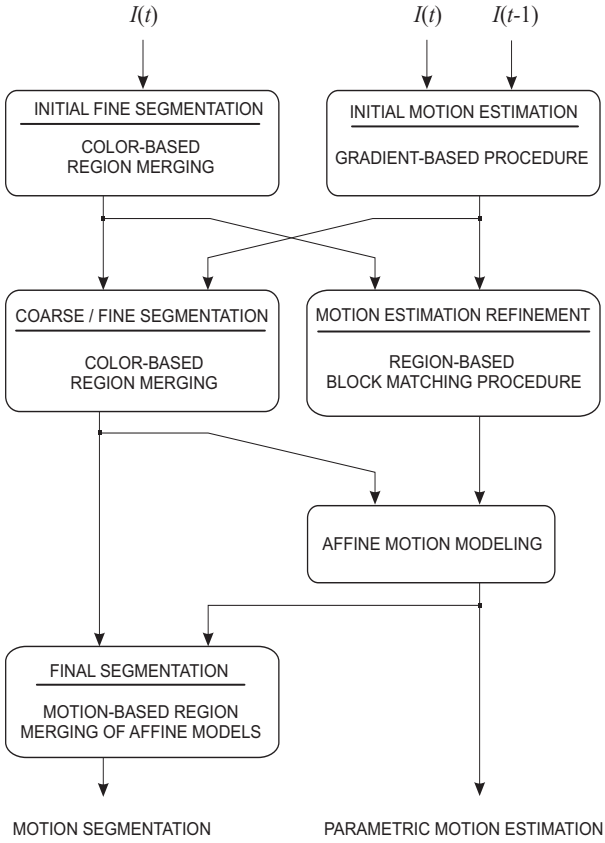

Figure 1. Block diagram of the algorithm.

allowed to continue only over regions outside of this mask, forming a coarse/fine partition. Color models are computationally less expensive to merge than motion models and the resulting larger, color-based regions will offer more stability in some of the later parametric modelings. The algorithm is still applicable in the event of global motion, such as camera movement. Currently, a version which includes global motion compensation is being developed.

From this set of dense, region-based, translational estimates we fit a parametric motion model which best describes the motion of points within the region of support given by the color-based partition results. We employ the 6-parameter affine model, chosen for its good compromise between precision and complexity. Once an affine motion model has been determined for each region of the partition, motion-based region merging of the image is carried out until a specified number of regions or an error criterion is reached.

3. MOTION ESTIMATION

3.1. Initial Motion Estimation: gradient-based procedure

The gradient-based approach by Lucas and Kanade [6] expresses the constant intensity assumption in terms of the spatio-temporal gradients I_x, I_y and I_t of the image and combines these constraints in a least squares fit as follows:

$$\sum_{\bar{x}\in N}(I_x(\bar{x},t)u + I_y(\bar{x},t)v + I_t)^2 . \quad (1)$$

Hence, the motion vectors, $\bar{u}=(u,v)^T$, may be obtained by minimizing equation (1) over some neighborhood N of n pixels $\bar{x}=(x,y)$. The solution is given by

$$(GG^T)\bar{u} = (Ge) \quad (2)$$

where $G = \begin{bmatrix} I_x(\bar{x}_1) & I_x(\bar{x}_2) & \ldots & I_x(\bar{x}_n) \\ I_y(\bar{x}_1) & I_y(\bar{x}_2) & \ldots & I_y(\bar{x}_n) \end{bmatrix}$

and $e = -\begin{bmatrix} I_t(\bar{x}_1) & I_t(\bar{x}_2) & \ldots & I_t(\bar{x}_n) \end{bmatrix}^T$.

The temporal gradient is obtained by direct subtraction on the images, i.e., $I_t(x,y)=I(x,y,t)-I(x,y,t-1)$. Spatial gradient information contained in the matrix GG^T is used to assess the reliability of estimates. Matrices which present eigenvalues below a minimum threshold are marked as unreliable. Also marked are estimates whose displaced frames present a mean absolute intensity difference above a specified threshold. The convergence range [6] of the estimates can be improved by applying pre-smoothing filters over the input images. The resulting velocities are assumed to represent the displacement of the central pixel of a square neighborhood N.

The described technique is extended to a multiresolution framework in which coarser image levels are formed by low-pass filtering and sub-sampling by 2 of the original images. The motion field is first calculated for the coarsest level, estimates are scaled by 2, bilinearly interpolated and used as initial guesses for estimating motion at subsequent finer levels.

3.2. Motion estimation refinement: region-based block matching procedure

Matching or correlation approaches adopt the same assumptions as gradient-based approaches but use instead a matching strategy in the search for motion vectors. The estimated motion vectors will be the arguments that minimize the error function:

$$\sum_{\bar{x}\in N}\left|I(x,y,t) - I(x-u,y-v,t-1)\right| . \quad (3)$$

The assumption of spatial coherency among the pixels of the square neighborhood N, or block, is violated in the case of a motion boundary. Unlike gradient-based approaches, matching techniques work directly over the image intensities and can be straightforwardly integrated with the color-based fine partition information from $I(t)$ to emphasize data which belong to the same motion region. In addition, this region-based approach does not require large support area. A weight matrix is thus introduced into the matching scheme emphasizing the pixels from the neighborhood N which belong to the same region as the central pixel. Pixels from

other regions, which could possibly be subject to other motions, are assigned a smaller weight.

Due to the computational cost of these estimates, region-based block matching will only be performed over a mask which is formed from the contours of color-based regions which are considered to have sufficient motion. The resulting estimates will replace the gradient-based values over these contour areas.

3.3. Affine modeling

The affine model can represent translations, rotations about the viewing axis and isotropic expansion of the motion field. It is defined by six-parameters, $a_1,...,a_6$, such that

$$\begin{pmatrix} d_x \\ d_y \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ a_4 & a_5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a_3 \\ a_6 \end{pmatrix} \tag{4}$$

for all pixels (x,y) within a region of support. Given a set of dense motion vector estimates, an affine model can be determined for each region of the image partition through a least squares fit in which the parameters are selected to minimize the sum of squared differences between the estimated motion field and the modeled displacement vectors, $\bar{d} = (d_x, d_y)$. Any estimates previously marked as unreliable are excluded from the computation.

4. REGION MERGING AND SEGMENTATION

Both color and motion segmentation are based on the general merging algotithm presented in [7]. The algorithm constructs a Region Adjacency Graph (RAG) composed of a set of nodes representing the regions and a set of links defining the connectivity between the regions. The merging algorithm removes some of the links and merges the corrresponding nodes.

In the case of color-based region merging, the regions are initially assumed to be individual pixels. Each region is modeled by a vector M of 3 elements containing the means of the YUV components. The order $O(R_i,R_j)$ in which two connected regions R_i and R_j are merged is based on a similarity measure such as:

$$O(R_i,R_j) = S_i \left\| M_{R_i} - M_R \right\|_2 + S_j \left\| M_{R_j} - M_R \right\|_2 \tag{5}$$

where S_i and S_j denote the size of the regions and R = $R_i \cup R_j$. The model of the union is assumed to be the model of the largest of its regions. The algorithm keeps merging the regions which present the highest similarity measure while updating the models until a termination criterion, such as a final number of regions, is reached.

Motion-based region merging is also accomplished with the general merging algorithm of [7]. In the motion-based case the model M is the set of affine parameters previously determined for each of the color-based partition regions. The merging order determines which of the affine models can best represent the motion of the union of regions in terms of a displaced frame difference measure. The merging order is given by

$$O(R_i,R_j) = \min_{K \in \{R_i,R_j\}} \left(\frac{\sum_{\bar{x} \in R} \left| I(\bar{x},t) - I(\bar{x} - \bar{d}_K(\bar{x}),t-1) \right|}{\left(\sum_{\bar{x} \in R} \left| \nabla I(\bar{x},t) \right| \right) / S} \right) \tag{6}$$

where $\bar{d}_K$ are the modeled vectors associated with region of support K.

Likewise, the merging criterion states that similar regions are merged until a specified number of regions is reached. The resulting regions are characterized by color and motion homogeneity and will define a video object segmentation.

5. RESULTS

Results are presented for the "Stefan" sequence with frame pairs 24 and 25 where the camera is static while the tennis player presents movement. Figure 2 depicts a partition of frame 25 into 750 regions obtained with the color-based region merging algorithm described in [7]. This segmentation defines the fine partition in which we may observe the generally accurate preservation of contour information. Frames 24 and 25 are used in gradient-based motion estimation via a multiresolution implementation of the Lucas and Kanade method with 3 resolution levels, 9x9 neighborhood windows and Gaussian pre-smoothing filters (σ=0.7). Regions are then detected as moving if at least 20% of the successfully estimated vectors within the region present magnitude greater than 0.5 pixels. Contours of these moving regions are presented in figure 3. Contour areas are assumed to be motion boundaries and subject to motion estimation refinement with the region-based matching scheme where same region pixels are emphasized with 2x weight.

Figure 4 shows a final motion segmentation using gradient-based motion estimation only, in contrast to figure 5, which illustrates a final motion segmentation (indicated as an output in figure 1) with the complete algorithm, including the region-based block matching estimation step on contour areas. The benefits of the region-based matching refinement scheme over motion boundaries are appreciable in the player's right arm region. The greater precision of motion estimation over such boundaries reduces occlusion problems. The segmented object presented in figure 5, as well as in figure 4, is one of the two final regions obtained as a result of motion-based merging, the other region being the complement or background. Figure 6 represents the second of the outputs depicted in figure 1, i.e., the resulting motion vector field after affine modeling of the motion estimates with the coarse / fine partition. Segmentation results are also presented for the "Akiyo" sequence, frame 70, with the same algorithm parameters. Figure 7 contains the color-based partition with 200 regions (a) and depicts the final motion segmentation (b). The algorithm obtains a good segmentation of the head which presents movement.

Figure 2. Fine partition of Stefan (frame 25) into 750 regions by color-based merging.

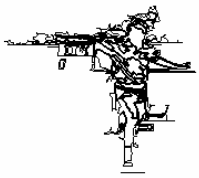

Figure 3. Contours of fine partition regions (figure 2) detected as moving.

Figure 4. Segmentation of Stefan (frame 25) obtained with gradient-based motion estimation only (no region-based block matching estimation refinement).

Figure 5. Segmentation of Stefan (frame 25) obtained with complete algorithm.

6. CONCLUSIONS

We have presented in this paper an algorithm for image segmentation and parametric motion estimation which successively integrates spatial, color-based segmentation with region-based estimation reinforcing results at each step. As illustrated in our test results, motion boundaries can be correctly detected and motion estimates over such boundaries improved.

We plan to extend our approach by incorporating information from multiple frames, obtaining more stable segmentation and estimation across successive frames. Another promising application is the characterization of objects based on complex motion models, resulting from the association of the various affine parameters already determined.

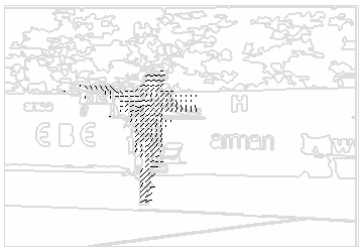

Figure 6. Affine motion vectors results for Stefan (frames 24 and 25).

(a) (b)

Figure 7. (a) Fine partition of Akiyo (frame 70) into 200 regions by color-based merging, (b) segmentation of Akiyo obtained with complete algorithm.

7. REFERENCES

[1] J. Y. A. Wang and E. H. Adelson, "Representing moving images with layers", *IEEE Trans. Image Processing*, vol. 3, pp. 625-638, 1994.

[2] P. Salembier and L. Garrido, "Binary Partition Tree as an Efficient Rpresentation for Image Processing, Segmentation and Information Retrieval", *IEEE Trans. Image Processing*, vol. 9, pp. 561-576, 2000.

[3] F. Moscheni, S. Bhattacharjee and M. Kunt, "Spatialtemporal segmentation based on region merging", *IEEE Trans. Patt. Anal. Mach. Intell*, vol. 20, pp. 897-915, 1998.

[4] I. Patras, E.A. Hendriks and R.L. Lagendijk, "Video segmentation by MAP labeling of watershed segments", *IEEE Trans. Patt. Anal. Mach. Intell,* vol. 23, pp. 326-332, 2001.

[5] Y. Tsaig and A. Averbuch, "Automatic segmentation of moving objects in video sequences: a region labeling approach", *IEEE Trans. Circuits Sys. Video Tech.*, vol. 12, pp. 597-612, 2002

[6] B. D. Lucas and T. Kanade, "An iterative image registration technique with application to stereo vision", in *Proc.of Image Understanding Workshop*, pp. 121-130, 1981.

[7] L. Garrido and P. Salembier, "Region based analysis of video sequences with a general merging algorithm", *IX European Signal Processing Conf., EUSIPCO'98*, vol. III, pp. 1693-1696, Rhodes, Greece, 1998.