REGION-BASED AFFINE MOTION SEGMENTATION USING COLOR INFORMATION

*P. Erhan Eren, Yucel Altunbasak, and A. Murat Tekalp**
Department of Electrical Engineering and Center for Electronic Imaging Systems
University of Rochester, Rochester, NY 14627

ABSTRACT

This paper presents region-based affine motion parameter clustering methods by motion-vector and intensity matching to provide improved robustness and alignment of motion boundaries with real object boundaries. Regions may be formed as fixed- or variable-size rectangular blocks, triangular patches, or arbitrary-shaped areas defined by color or texture uniformity. A particular combination of these affine clustering methods is then proposed to obtain the best segmentation results on a variety of image sequences. Experimental results showing the benefits of the proposed scheme are provided.

1. INTRODUCTION

Various approaches exist for motion segmentation [1], which can be classified as: dominant motion based segmentation [2], segmentation by parameter clustering [3], and simultaneous motion estimation and segmentation [4]. Bergen *et al.* [2] proposed estimation of the dominant motion and segmentation by successively extracting objects with the dominant motion. Some difficulties with this approach were reported when there's no single dominant motion. Wang and Adelson [3] suggested clustering of affine motion parameters (initialized by dividing the image into blocks) with some pre- and post-processing. However, clustering in the affine space (by affine parameter matching) is sensitive to parameter estimation errors, especially in the case of rotations and shear. Pixel-based motion segmentation methods, e.g., [2, 3], suffer from the drawback that the resulting segmentation map may contain isolated labels. Spatial continuity constraints in the form of Gibbs random field models have been introduced to overcome this problem [5]. Yet another approach is the simultaneous Bayesian motion estimation and segmentation approach of Chang *et al.* [4]. However, the computation cost of the resulting Bayesian motion segmentation algorithms limits their practical use. Furthermore, none of these techniques possess the desired property that the estimated motion boundaries coincide with the object boundaries (color or gray level edges). This problem was partially addressed in [6] by using a morphological segmentation approach.

This paper proposes a new method for enforcing parametric model-based motion boundaries to correspond to spatial object boundaries (edges) as much as possible. The proposed approach, based on the observation that motion edges are generally a subset of spatial edges, successfully reduces segmentation errors in the vicinity of motion boundaries, which are mainly due to motion estimation errors concentrated around motion edges.

*This work is supported in part by a National Science Foundation IUCRC grant and a New York State Science and Technology Foundation grant to the Center for Electronic Imaging Systems at the University of Rochester, and a grant by Eastman Kodak Company.

2. PARAMETRIC MOTION MODELING

Parametric motion models can be used on a global or local basis for a more compact representation of the motion field [1]. Global models are suitable to describe camera motion, while local models can capture multiple motions and/or mild deformations. In this paper, we employ a six-parameter affine model given by

$$\tilde{x}'_{i,j} = a_{i,1}x_{i,j} + a_{i,2}y_{i,j} + a_{i,3} \tag{1}$$

$$\tilde{y}'_{i,j} = a_{i,4}x_{i,j} + a_{i,5}y_{i,j} + a_{i,6} \tag{2}$$

where $a_{i,1}, \ldots, a_{i,6}$ denote the affine-motion parameters for the i^{th} patch, $(x_{i,j}, y_{i,j})$ is the coordinates of the j^{th} pixel within the i^{th} patch, and $(tildex'_{i,j}, tildey'_{i,j})$ is the corresponding pixel location in the next frame as predicted by the model. Other parametric forms, such eight-parameter perspective or bilinear models, may be more suitable in some applications depending on the geometry of the projection [1].

3. REGION-BASED AFFINE PARAMETER CLUSTERING

In this section, we propose a region-based approach, where the image is first segmented into predefined small regions, and then each predefined region is assigned a single motion label, to obtain spatially contiguous segmentation maps which are closely related to actual object boundaries, without a heavy computational burden.

3.1. Region Definition

The predefined regions should be such that each region has a single motion. It is generally true that motion boundaries coincide with color segment boundaries; i.e., color boundaries are almost always a superset of motion boundaries. Therefore, one can first perform a color segmentation to obtain a set of candidate motion segments. Then, those segments which have the same motion can be merged to obtain the final motion segmentation map. Other approaches to region definition include mesh-based partitioning of the scene [7], macro pixels (N × N blocks) to improve the robustness of the pixel-based affine motion segmentation. Although, the notation and equations that follow are valid for any region definition; in this paper, we employ color segmentation since it provides both smoothness and alignment of boundaries with real object edges.

3.2. Region-Based Motion-Vector Matching

We want to find the motion segmentation map $\mathbf{L}$ and the corresponding affine parameter sets $\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_K$ which best fit the motion-vector field, such that

$$\sum_{i,j} \| \mathbf{V}(i,j) - \mathbf{P}(\mathbf{A}_{L(i,j)};(i,j)) \| \quad (3)$$

is minimized, where $\mathbf{L}(i,j)$ is the motion label at the pixel location (i,j) in the image to be segmented and has one of the values $1,2,...,K$; $\mathbf{P}$ is a projection operator defined by

$$\mathbf{P}(\mathbf{A}_k;(i,j)) = \begin{bmatrix} (a_{k,1}-1)i + a_{k,2}j + a_{k,3} \\ a_{k,4}i + (a_{k,5}-1)j + a_{k,6} \end{bmatrix} \quad (4)$$

and $\mathbf{V}(i,j)$ is the motion vector at the pixel location (i,j)

$$\mathbf{V}(i,j) = \begin{bmatrix} v_x(i,j) \\ v_y(i,j) \end{bmatrix} \quad (5)$$

with v_x, v_y being the horizontal and vertical components of the motion vector, respectively.

This minimization is carried out under the smoothness constraint provided by the regions. Let $\mathbf{C}$ represent the region map containing M regions, formed by disconnected regions with uniform color. The motion segmentation map, $\mathbf{L}$, is initialized and then updated by the following iterations:

1. Update the affine parameter sets $\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_K$, such that

$$\mathbf{A_s} = Arg \;\; min_{\mathbf{A}} \sum_{(i,j)\in \mathbf{L}_s} \| \mathbf{V}(i,j) - \mathbf{P}(\mathbf{A};(i,j)) \| \quad (6)$$

where $\mathbf{L_s}$ is the set of pixels (i,j) with the motion label $\mathbf{L}(i,j)$=s.

This minimization can be written in linear matrix equation form as:

$$\begin{bmatrix} i & j & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & i & j & 1 \end{bmatrix} \begin{bmatrix} a_{s,1} \\ a_{s,2} \\ a_{s,3} \\ a_{s,4} \\ a_{s,5} \\ a_{s,6} \end{bmatrix} = \begin{bmatrix} v_x(i,j) \\ v_y(i,j) \end{bmatrix} \quad (7)$$

for all (i,j) in $\mathbf{L}_s$.

2. Assign a motion label to each region $\mathbf{C_m}$, $m = 1,2,...,M$, such that

$$\mathbf{L}(\mathbf{C}_m) = Arg\; min_s(\sum_{(i,j)\in \mathbf{C}_m} \| \mathbf{V}(i,j) - \mathbf{P}(\mathbf{A}_s;(i,j)) \|) \quad (8)$$

where $s = 1,2,...K$, and $\mathbf{C_m}$ is the set of pixels (i,j) with the region label $\mathbf{C}(i,j)$=m.

In the above equations, if each region $\mathbf{C}_m$ consists of a single pixel, the iterations are effectively carried over the individual pixels, and the motion label assignment is performed at each pixel independently. This case corresponds to affine parameter clustering with pixel-based motion-vector matching. Therefore, the above formulation covers both the pixel-based and the region-based framework.

The pixel-based motion matching criteria has been used for motion labeling of pixels by Wang and Adelson [3]. However, they perform the update of the affine parameter sets by affine parameter matching instead of motion-vector matching, and region-based assignment is not utilized in their method. Similarly, the Bayesian formulation proposed by Murray and Buxton [5] employs the concept of pixel-based affine clustering in the motion vector space. The label assignment equation is essentially the same, except that they include a smoothness term through the Gibbs random field models, to enforce spatially contiguous region formation and use simulated annealing algorithm for minimization.

3.3. Region-Based Intensity Matching

The formulation in the previous section incorporates the intensity information into the motion segmentation process indirectly, through the use of color segmentation. Below we present a similar scheme, where the affine parameter clustering is effectively performed through intensity matching and therefore the intensity information is used directly.

Let $\mathbf{I}_k$ and $\mathbf{I}_{k-1}$ be the two frames and the motion-vectors are estimated for each pixel in the reference frame $\mathbf{I}_k$. In this formulation, we find the motion segmentation map $\mathbf{L}$ and the corresponding affine parameter sets $\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_K$ such that

$$\sum_{i,j} \| \mathbf{I}_k(i,j) - \mathbf{Q}(\mathbf{I}_{k-1};\mathbf{A}_{L(i,j)};(i,j)) \| \quad (9)$$

is minimized, where $\mathbf{Q}$ is a motion compensation operator defined as

$$\mathbf{Q}(\mathbf{I}_{k-1};\mathbf{A}_{L(i,j)};(i,j)) = \mathbf{I}_{k-1}(i',j') \quad (10)$$

and

$$[i',j']^T = [i,j]^T + \mathbf{P}(\mathbf{A}_{L(C(i,j))};(i,j)) \quad (11)$$

Again, in this case, the minimization is carried out under the smoothness constraint provided by the regions. $\mathbf{C}$ is the region map containing M regions. The motion segmentation map, $\mathbf{L}$, is initialized and then updated by the following iterations:

1. Update the affine parameter sets $\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_K$, such that

$$\mathbf{A_s} = Arg \;\; min_{\mathbf{A}} \sum_{(i,j)\in \mathbf{L}_s} \| \mathbf{V}(i,j) - \mathbf{P}(\mathbf{A};(i,j)) \| \quad (12)$$

where $\mathbf{L_s}$ is the set of pixels with the with motion label s in $\mathbf{L}$.

Similarly, this minimization can be written as in eqn 6.

2. Assign a motion label to each region $\mathbf{C_m}$, $m = 1,2,...,M$, such that

$$\mathbf{L}(\mathbf{C}_m) = Arg\; min_s(\sum_{(i,j)\in \mathbf{C}_m} \| \mathbf{I}(i,j) - \mathbf{Q}(\mathbf{I}_{k-1};\mathbf{A}_s;(i,j)) \|) \quad (13)$$

where $s = 1,2,...K$, and $\mathbf{C_m}$ is the set of pixels (i,j) with the region label $\mathbf{C}(i,j)$=m.

4. ALGORITHM

A particular combination of the clustering methods formulated above is utilized to obtain the motion segmentation results depicted in the "Results" section. The proposed algorithm is as follows:

1. Estimate the motion-vector field between the two frames, and eliminate "inaccurate" motion vectors by thresholding DFD(displaced frame difference)
2. Perform spatial color/gray-level segmentation on the reference frame and split the color segments with disconnected regions
3. Initialize the motion segmentation map
4. Cluster affine parameters by pixel-based motion-vector matching, i.e. each region is formed of a pixel:
   (a) Update the affine parameters for each motion class, using the least squares method as in eqn. 6
   (b) Assign each pixel with "accurate" motion vectors to the class which yields the minimum motion-vector residue for that pixel, as in eqn. 8

(c) Iterate until the number of pixels changing motion classes in the iteration is below a threshold

5. Cluster affine parameters by region-based motion-vector matching:
   (a) Update the affine parameters for each motion class, using the least squares method as in eqn. 6
   (b) Assign each region to the motion class which yields the minimum motion residue for that region, as in eqn. 8
   (c) Iterate until the number of pixels changing motion classes in the iteration is below a threshold
6. Cluster affine parameters by region-based intensity matching:
   (a) Update the affine parameters for each motion class, using the least squares method as in eqn. 6
   (b) Assign each region to the motion class which yields the minimum intensity residue, as in eqn. 13
   (c) Iterate until the number of pixels changing motion classes in the iteration is below a threshold

Several observations are in order in the design of this algorithm: Affine parameter clustering by pixel-based motion-vector matching suffers from errors in the motion-vector field; and therefore, gives a crude segmentation map, but serves as a good initialization for the next step. This segmentation map is further enhanced by the smoothness imposed during the region-based assignment, boundary alignment is achieved and a fairly decent motion segmentation map is obtained at this part. Next, this result is checked against the DFD criteria using the intensity-based matching.

5. RESULTS

The proposed algorithm is tested on the MPEG-4 test sequences "mother&daughter"(frames 1-2) and "mobile&calendar"(frames 136-137). In the two frames of the "mother&daughter", mother's head is rotating while her body, the background and the child are stationary. In the "mobile&calendar", there is a camera pan across a scene containing a rotating ball, a moving train and a vertically translating calendar against a stationary background. Figure 1(a) shows the frame 2 of the "mother&daughter" sequence, whereas the frame 137 of the "mobile&calendar" sequence is given in Figure 2(a). For the results to follow, the subsections in figures 1 and 2 correspond to the sequences "mother&daughter" and "mobile&calendar", respectively. In both sequences, color segmentation is performed on the shown reference images, using the fuzzy c-means technique [8]. To further refine the color segmentation, disconnected regions of each color segment are split into individual segments. The resulting segmentation maps are shown in figures 1(b) and 2(b), and the segments thus obtained are to be utilized in the region-based clustering steps of the algorithm. The motion field in each case is calculated by the Lucas-Kanade algorithm [9] with three levels of hierarchy and the initial segmentation of the motion field is obtained by pixel-based affine clustering in the motion-vector space. The results of this step are shown in figures 1(c) and 2(c). Note that the motion segmentation maps are not coherent due to assigments at individual pixel level. Then, utilizing the pre-computed color segmentation maps, the region-based affine clustering in the motion-vector space is performed in each case, to enforce the boundary alignment and the spatial coherence. The resulting segmentation maps are shown in figures 1(d) and 2(d). In each case, the improvements can be clearly seen. The final step is to perform the region based clustering using the intensity information. In this step, a region is merged with another if the latter's affine parameter set gives a smaller DFD than its own. And this accounts for the changes in the background and the calendar segments in figure 2(e), whereas the segmentation for the "mother&daughter" remains the same as in figure 1(d), therefore not shown.

6. CONCLUSION

The proposed region-based approach effectively provides spatially smooth motion segments and enables alignment of motion segment boundaries with actual object boundaries. The pixel-based motion segmentation methods are error-prune in the vicinity of object boundaries, since motion estimation generally fails in these areas due to occlusion. This problem can be overcome with our proposed scheme provided that an accurate color segmentation can be obtained. Although, the performance of the method depends on the definition of the initial regions, we observe that using the color based region definition, we can accept some false color segments, since they can be merged later. Experimental results demonstrate the robustness of the proposed method, which can also be viewed as integration of motion and color segmentation.

REFERENCES

[1] A. M. Tekalp, *Digital Video Processing*, Prentice-Hall, 1995.

[2] J. R. Bergen, P. J. Burt and K. Hanna, "Dynamic multiple-motion computation," in *Artificial Intelligence and Computer Vision* (Y.A. Feldman and A. Bruckstein, Eds.). Holland: Elsevier, 1992, pp. 147-156.

[3] J. Y. A. Wang and E. H. Adelson, "Representing moving images with layers," *IEEE Trans. Image Processing*, vol. 3, No. 5, pp. 625-638, 1994.

[4] M. M. Chang, M. I. Sezan, and A. M. Tekalp, "An Algorithm for Simultaneous Motion Estimation and Scene Segmentation," *Proc. IEEE Int. Conf. Acoust. Speech Sign. Proc.*, Adelaide, Australia, April 1994.

[5] D. W. Murray and B. F. Buxton, "Scene segmentation from visual motion using global optimization," *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. 9, no. 2, pp. 220–228, Mar. 1987.

[6] C. Gu, T. Ebrahimi, and M. Kunt, "Morphological moving object segmentation and tracking for content-based video coding," *Proc. Int. Symp. Multimedia Comm. and Video Coding*, New York, NY, Oct. 11-13. 1995.

[7] Y. Altunbasak and A. M. Tekalp, "Content-based mesh design for 2-D object-based video coding," *Symp. on Multimedia Comm. and Video Coding*, New York City, New York, Oct. 1995; also in *Multimedia Communications and Video Coding* (Y. Wang, S. Panwar, S.-P. Kim, and H. L. Bertoni, eds.), Plenum Press, 1996.

[8] Y. W. Lim, S. U. Lee, "On The Color Image Segmentation Algorithm Based on the Thresholding and the Fuzzy c-Means Techniques", *Pattern Recognition*, vol. 23, no. 9, pp. 935-952, 1990.

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

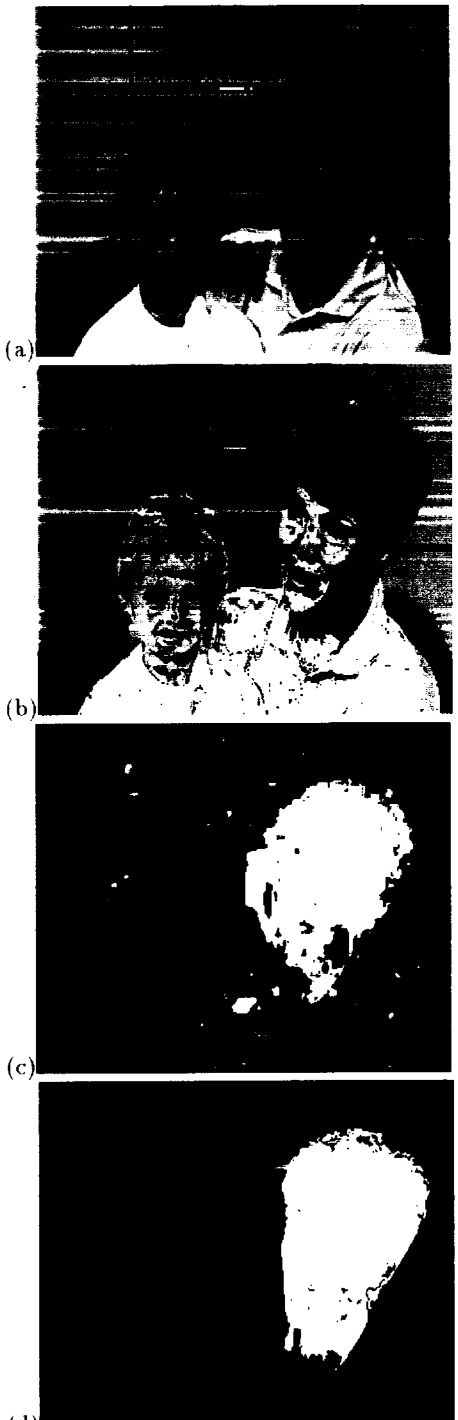

Figure 1. a) Reference frame; b) Region map obtained by color segmentation and splitting; c) Motion segmentation map after pixel-based motion-vector matching; d) Motion segmentation map after region-based motion-vector matching

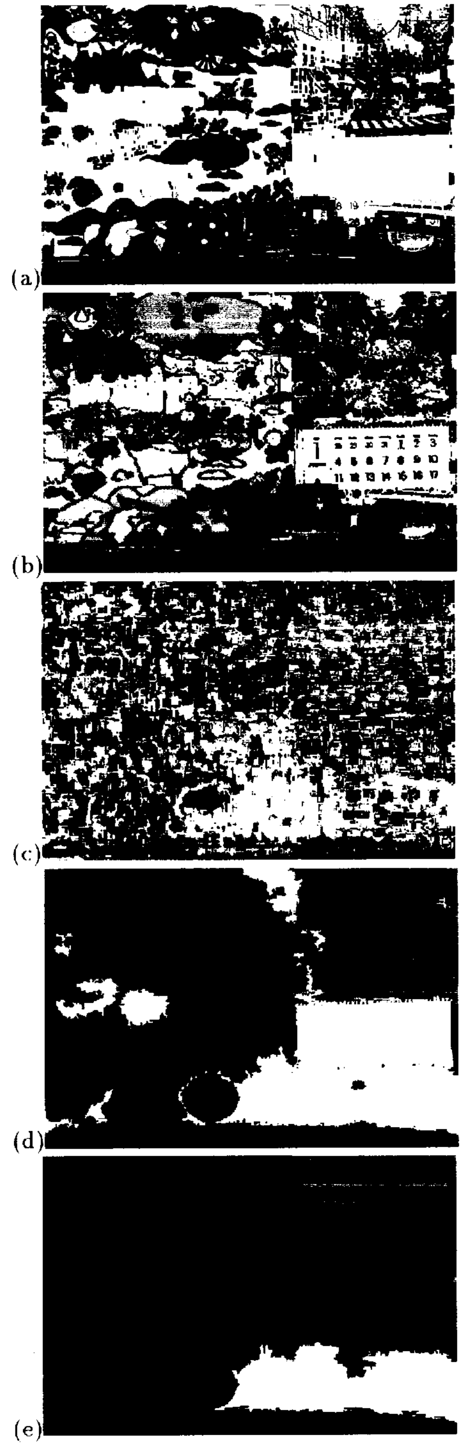

Figure 2. a) Reference frame; b) Region map obtained by color segmentation and splitting; c) Motion segmentation map after pixel-based motion-vector matching; d) Motion segmentation map after region-based motion-vector matching; e) Motion segmentation map after region-based intensity matching