Motion Segmentation by Multi-stage Affine Classification [1]

Georgi D. Borshukov, Gozde Bozdagi, Yucel Altunbasak, and A. Murat Tekalp

Department of Electrical Engineering and Center for Electronic Imaging Systems
University of Rochester, Rochester, NY 14627

Abstract

We present a multi-stage affine motion segmentation method which combines the benefits of the dominant motion and block-based affine modeling approaches. In particular, we propose two key modifications to a recent motion segmentation algorithm developed by Wang and Adelson [1]: i) the adaptive k-means clustering step is replaced by a merging step, whereby the affine parameters of a block which has the smallest representation error, rather than the respective cluster center, is used to represent each layer, and ii) we implement it in multiple stages, where pixels belonging to a single motion model are labeled at each stage. Performance improvement due to the proposed modifications is demonstrated on real video frames.

1 Introduction

Motion segmentation refers to grouping together pixels that undergo a common motion. Various methods for motion segmentation can be classified as those belonging to affine clustering approach [1], dominant motion approach [2], and simultaneous motion estimation and segmentation approach [3].

Recently, Wang and Adelson [1] proposed an affine-clustering based motion segmentation method for layered representation of video. First, a dense flow field is estimated. This flow field is, then, divided into non-overlapping rectangular blocks, and affine motion parameters are estimated for each block. Next, the affine parameters are subjected to a reliability test, based on how well the affine motion vectors within each block fit the estimated dense flow field. Those affine parameters

[1] 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.
Georgi D. Borshukov is now with the Department of Electrical Engineering and Computer Science, University of California at Berkeley. Gozde Bozdagi is now with Xerox Corporation, Webster. Yucel Altunbasak is now with HP Labs, CA.

(hypothesis) that pass the test are clustered into a small number of classes using an adaptive k-means algorithm. Finally, each flow vector is assigned to one of the resulting classes, represented by their cluster centers, using a minimum residual classifier. Motion vectors where the minimum residual is above a prespecified threshold are handled separately. Our experimentation with this method has led to the following observations:

1. The adaptive k-means clustering in essence computes an "average" affine model (hypothesis) for each layer. We show that better results can be obtained by replacing the clustering step with a "merge" step, which in effect picks the model having the smallest residual for each layer.

2. Labeling the class memberships of all motion vectors in a single stage generally produces unsatisfactory results. That is, it may result either segmentation of a single motion into multiple classes (oversegmentation), or undesired merging of two or more motions. We address this problem through a multi-stage segmentation strategy, which also automatically decides for the best number of motion classes.

The proposed multi-stage affine segmentation combines the dominant motion approach of Bergen et al. [2] with affine clustering approach of Wang and Adelson [1]. The image is divided into N rectangular blocks, and an affine motion parameter vector is estimated for each block as discussed in Section 2. We, then, identify the affine parameter vector associated with the dominant motion out of this set of seed parameters by means of a merge procedure, which is described in Section 3. Next, motion vectors which can be well-represented by this dominant motion model are identified and labeled as presented in Section 4. The procedure is then repeated excluding all motion vectors which are already labeled; that is, the affine parameter vectors of all blocks with unassigned motion vectors are merged again, this time using a different merge threshold, to find the next dominant motion parameter vector, and so on. A complete algorithmic description of the method is provided in Section 5.

2 Estimation of Seed Affine Parameters

The image is divided into N more-or-less equally spaced rectangular blocks as shown in Fig. 1. A set of affine parameter vectors are estimated, one vector for each block, by a linear least squares procedure. Assuming that a dense flow field $(u(x,y), v(x,y))$ has been estimated and divided into rectangular blocks denoted by B_i, we approximate the flow field within each block B_i by an affine model of the form

$$\begin{aligned} u_a(x,y) &= (a_{i1}-1)x + a_{i2}y + a_{i3} \\ v_a(x,y) &= a_{i4}x + (a_{i5}-1)y + a_{i6}, \end{aligned} \tag{1}$$

where (x,y) denotes the position of pixels within B_i with respect to a common origin. The parameters a_{i1}, a_{i2}, a_{i3}, a_{i4}, a_{i5}, a_{i6} can be estimated by linear least squares procedure to minimize the residual error

$$\sigma_i^2 = \frac{1}{N_i} \sum_{(x,y)\in B_i} [(u(x,y) - u_a(x,y))^2 + (v(x,y) - v_a(x,y))^2] \tag{2}$$

where N_i is the number of pixels in B_i. Blocks for which σ_i^2 is above a predetermined selection threshold, T_{stage}, are discarded, since their affine parameters are deemed inaccurate, possibly due to presence of multiple motion within the block. Affine parameters of those blocks which pass the reliability test constitute the seed parameter set.

3 Dominant Motion Selection by a Merge Procedure

Different blocks that belong to the same moving object may have somewhat different affine parameter estimates due to inaccuracies in the initial dense motion estimation or due to the fact that all objects can not be represented by planar patches. To group all blocks that are part of the same motion together, an adaptive k-means procedure has been used by Wang and Adelson [1]. A cluster center represents an “average” affine model, which we observed is usually “less adequate" to represent the motion of the object than for example the model that has the least residual error over a seed block on that object. The proposed multi-stage, multi-pass "merge" procedure attempts to address this problem.

At each stage, we identify the affine parameters corresponding to the dominant motion in the scene by merging those unmarked seed blocks whose affine parameters are "close" to each other according to some distance measure in the 6-D affine parameter space. At the first pass of the first stage all seed blocks are unmarked. An example of the distance function that can be employed is that used by Wang and Adelson [1]. However, after each merge decision, we keep the affine parameters of the seed which has the smaller residual error over the merged blocks, thus preserving the better of the two models instead of taking their average. Each merged block is marked. In order the result to be independent of the starting block position, the merge process requires multiple passes over the seed blocks at each stage, where seed blocks that are within T_m (merge threshold) of the first unmarked seed block are merged in each pass. Furthermore, to minimize the impact of the merge threshold T_m on the result, we start with a small T_m at the first stage, and progressively increase it at each stage (T_m remains fixed for all passes at a given stage). Finally, we select the affine parameters of the class which has the highest population of assigned blocks as the dominant motion at that stage.

Each stage of the merge procedure is followed by a labeling procedure which is described in the next section. The marked seed blocks are reset after the labeling procedure for the next stage of the merge procedure as also discussed in the next section.

4 Labeling Procedure

Segmentation, which refers to assigning each motion vector a class label, is implemented in such a way that the motion vectors that best fit a single affine model (the dominant model of that stage) is labeled in each stage. Stage 1 of the labeling procedure considers the assignment of the motion vectors to the seed with the highest population after stage 1 of the merge procedure given in Section 3. The criterion for label assignment is similar to that of Wang and Adelson [1]. A motion vector is labeled if

$$[(u(x,y) - u_a(x,y))^2 + (v(x,y) - v_a(x,y))^2] < T_{assign}, \tag{3}$$

where $(u_a(x,y), v_a(x,y))$ denotes the parametric motion vector predicted by the dominant affine model, and T_{assign} is the assignment threshold. The assignment threshold is initialized at a small value and is raised from stage to stage to accommodate lower accuracy seed models at later stages of the algorithm. Note that the models become increasingly less accurate towards the final stages, since we keep increasing the merge threshold T_m as the stages progress.

Motion vectors that are labeled at any stage are not used at subsequent stages. To this effect, the labeled image is post-processed to remove isolated patches of labeled and unlabeled pixels. Then, only the seed blocks with unlabeled pixels are considered in the next stage of the algorithm for the merge procedure with an increased value of T_m as discussed. The stages, consisting of merge and labeling procedures, continue until almost all motion vectors are labeled. The proposed multistage procedure provides increased accuracy and robustness.

5 The Complete Multi-Stage Segmentation Algorithm

A summary of the proposed multi-stage segmentation algorithm is as follows:

1. Estimate a dense optical flow field; e.g., by using the algorithm in [4].

2. Divide the motion field into rectangular blocks.

3. For each block, estimate the affine parameters by the method of linear least squares (as described in Section 2).

4. Threshold the motion residual by T_{stage} to determine reliable blocks.

5. Set the value of T_m and apply the merge procedure described in Section 3 to find the dominant affine model for motion vector labeling.

6. Set the value of T_{assign}. Label those motion vectors satisfying the velocity check criterion given by (3).

7. Remove all labeled pixels, eliminate isolated regions, and identify the seed blocks to be considered at the next stage.

8. If all motion vectors are assigned or no seed blocks remain for consideration at the next stage, then stop, otherwise go to step 4.

6 Results

We have implemented the proposed algorithm in C++ on a Sparc 20 workstation. Fig. 2a and 2b show the seventh and eight frames of the MPEG test sequence *Mobile and Calendar*. The estimated motion field between these two frames are shown in Fig. 3. There are 5 independently moving objects in the scene: The ball and the toy pendulum are rotating independently, the train is translating, the calendar is translating in upward direction with respect to the wall paper, and the background is translating due to the camera pan. The segmentation results obtained by the proposed algorithm and by the method of Wang and Adelson are shown in Fig. 4a and 4b, respectively. In the method of Wang and Adelson, if parameters are chosen to describe the rotating ball as a single object, other objects with close motions (e.g., the calendar and background) are also inadvertently merged. However, we are able to distinguish between the calendar and train, while keeping the ball as a single object by performing merging and assignment in multiple stages. We can also see the effect of averaging on the train. Choosing the most reliable block as the seed block enables us to eliminate the effect of less reliable blocks (the border blocks) on the assignment.

Figs. 5a and 5b; and 8a and 8b show the first and second frames of the *Flower Garden* sequence and the sixth and ninth frames of the *Salesman* sequence, respectively. The estimated motion vectors and the segmentation results are shown in Figs. 6, 9, 7 and 10, respectively. For comparison purposes, the results obtained by the method of Wang and Adelson are also given. These results also clearly demonstrate how we can capture each individually moving object using the proposed multi-stage assignment approach.

7 Conclusion

In this paper, we presented an algorithm which exploits the benefits of the dominant motion and block-based affine modeling approaches. The proposed method can handle motion vectors suffering from noise in a more robust way than previous approaches. We demonstrated the performance and robustness of the proposed algorithm on real video frames.

8 References

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

[2] J. R. Berger, 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] A. M. Tekalp, *Digital Video Processing,* Prentice-Hall, 1995.

[4] 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.

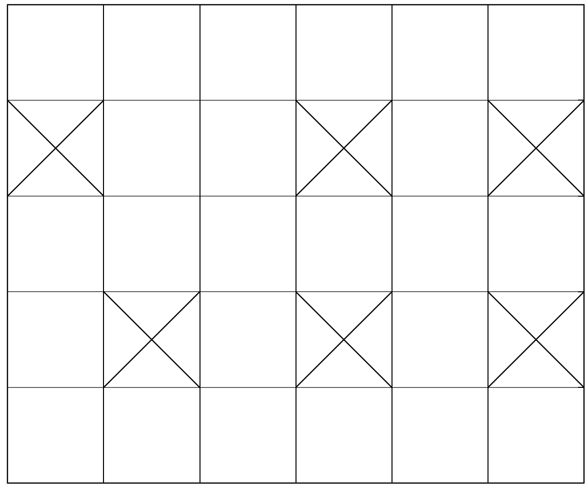

Figure 1: Initial seed blocks.

Figure 2: The *Mobile and Calendar* sequence: a) the 7th frame; b) the 8th frame.

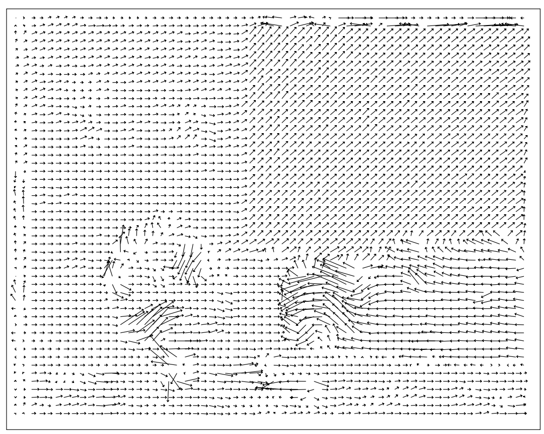

Figure 3: Estimated motion field.

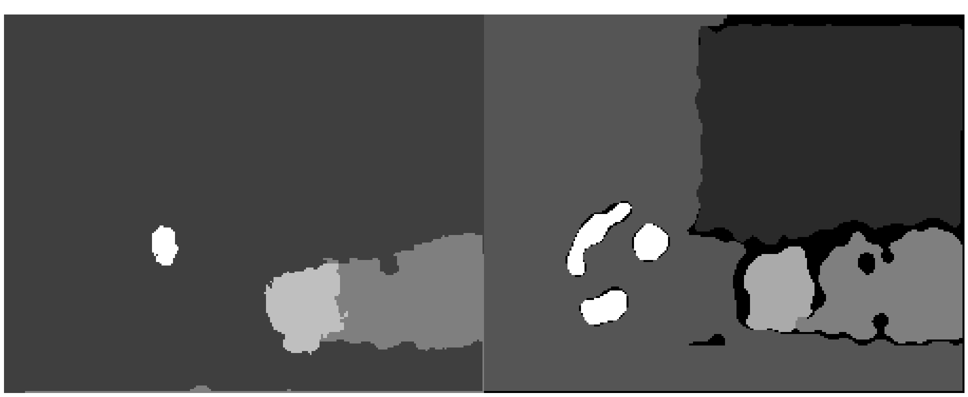

Figure 4: Segmentation obtained by a) the Wang-Adelson method; b) the proposed method.

Figure 5: The *Flower Garden* sequence: a) the 1st frame; b) the 2nd frame.

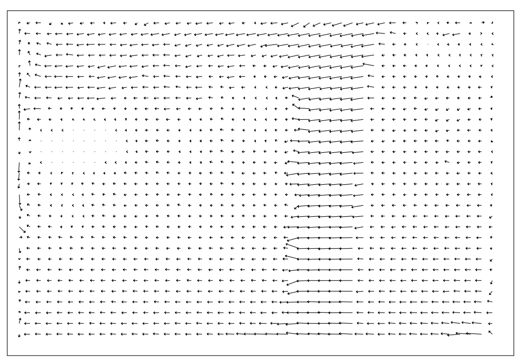

Figure 6: Estimated motion field.

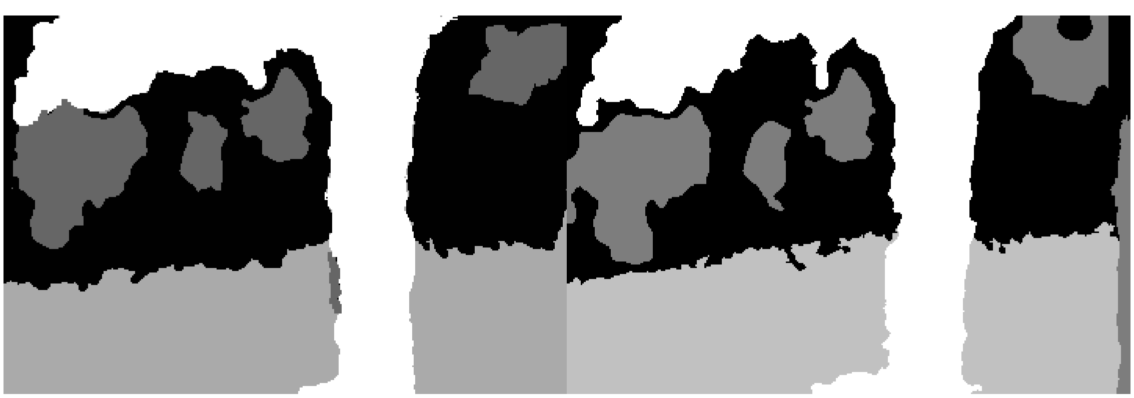

Figure 7: Segmentation obtained by a) the Wang-Adelson method; b) the proposed method.

Figure 8: The *Salesman* sequence: a) the 6th frame; b) the 9th frame.

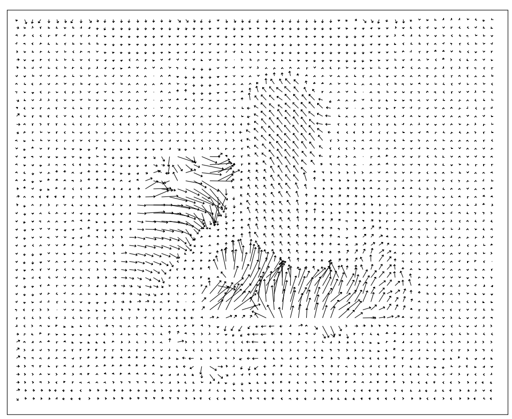

Figure 9: Estimated motion field.

Figure 10: Segmentation obtained by a) the Wang-Adelson method; b) the proposed method.