Robust region-merging technique for video sequences: spatiotemporal segmentation (Proceedings Paper)

Author(s): **Riccardo Leonardi; Pier A. Migliorati; Giuseppe Tofanicchio**

Date: **28 December 1998**

ISBN: **9780819431240**

Proceedings Vol. 3653

Visual Communications and Image Processing '99, Kiyoharu Aizawa; Robert L. Stevenson; Ya-Qin Zhang, Editors, pp.1258-2827

Date: **28 December 1998**
ISBN: **9780819431240**

Paper Abstract

The segmentation of video sequences into regions underlying a coherent motion is one of the most important processing in video analysis and coding. In this paper, we propose a reliability measure that indicates to what extent an affine motion model represents the motion of an image region. This reliability measure is then proposed as a criterion to coherently merge moving image regions in a Minimum Description Length (MDL) framework. To overcome the region-based motion estimation and segmentation chicken and egg problem, the motion field estimation and the segmentation task are treated separately. After a global motion compensation, a local motion field estimation is carried out starting from a translational motion model. Concurrently, a Markov Random Field model based algorithm provides for an initial static image partition. The motion estimation and segmentation problem is then formulated in the view of the MDL principle. A merging stage based on a directed weighted graph gives the final spatio-temporal segmentation. The simulation results show the effectiveness of the proposed algorithm.

DOI: 10.1117/12.334633

A Robust Region Merging Technique for Video Sequences Spatio-Temporal Segmentation

R. Leonardi[a], P. Migliorati[a], G. Tofanicchio[a]

[a]University of Brescia, Via Branze 38, Brescia, Italy - Email:leon@ing.unibs.it

ABSTRACT

The segmentation of video sequences into regions underlying a coherent motion is one of the most important processing in video analysis and coding. In this paper, we propose a reliability measure that indicates to what extent an affine motion model represents the motion of an image region. This reliability measure is then proposed as a criterion to coherently merge moving image regions in a Minimum Description Length (MDL) framework. To overcome the region-based motion estimation and segmentation chicken and egg problem, the motion field estimation and the segmentation task are treated separately. After a global motion compensation, a local motion field estimation is carried out starting from a traslational motion model. Concurrently, a Markov Random Field model based algorithm provides for an initial static image partition. The motion estimation and segmentation problem is then formulated in the view of the MDL principle. A merging stage based on a directed weighted graph gives the final spatio-temporal segmentation. The simulation results show the effectiveness of the proposed algorithm.

Keywords: Motion segmentation, affine models, merging criteria, minimum description length principle, robust clustering rules

1. INTRODUCTION

Motion is one of the most important characteristics to identify objects in a scene. Motion based segmentation is therefore very important in application such as dynamic scene analysis, time-to-collision calculation, obstacle detection and tracking of moving objects. Moreover, a motion-based segmentation can be directly used also in hybrid video coding architectures.[1] Contrary to intensity-based approaches,[2] motion based segmentation deals with few and large regions that are likely to identify real moving objects in the scene. In[3],[4] given the motion information, regions with the same affine motion are assumed as belonging to the same object. The motion parameters are extracted from the optical flow field by means of linear regression, and temporal segmentation is obtained by clustering in the parameter space; however the resulting motion based image segmentation suffers poor accuracy on object boundaries. In,[5] starting from a luminance-based segmentation, local motion estimation is performed. Then a bottom up segmentation procedure merges coherent adjacent regions.

The drawback of these methods is that the parameterisation of the motion field might not have a unique solution. In[6],[7] starting from a luminance-based segmentation, a merging procedure, based on the mean square error value of the Displaced Frame Difference (DFD) and Motion Difference (MD) respectively, provides the final spatio-temporal segmentation. The use of the DFD alone as a characterisation of motion coherence between regions may lead to a significant over-segmentation, due to the important value of the DFD in textured areas, while the use of MD alone cannot face a potential initial motion estimator failure in an estimation which uses an indirect parametric estimator. In,[8] the initial set of regions are compared on the basis of a similarity measure that integrates the motion information available in the affine parameter space and the displaced frame difference information. We believe that in order to reach satisfactory performance an approach that selects either a low DFD measure or a good correspondence of the affine model with the underlying real motion is necessary. As such, this work proposes the definition of a Minimum Description Length (MDL) approach that reflects the matching of either one of these correspondence (DFD and parametric motion model).

More specifically, a motion parameter reliability measure is proposed. It indicates to what extent an affine motion model represents the motion of an image region taking into account both the DFD and MD. This reliability measure is obtained from two indicators. The first one represents the mean square error of the residuals after compensating a region using a set of affine motion parameters and the second one represents the mean square error of the residuals between the motion vectors inside the region of interest and the motion vectors obtained

Part of the IS&T/SPIE Conference on Visual Communications and Image Processing '99 • San Jose, California • January 1999
SPIE Vol. 3653 • 0277-786X/98/$10.00

by the same affine motion parameters vector. This reliability measure is used to coherently merge neighbouring moving image regions in a Minimum Description Length (MDL)[9] framework.
In order to get the region merging algorithm started, an initial set of regions with associated motion parameter have to be obtained. An affine motion estimation and segmentation procedure is then carried out as follows. To overcome the region-based motion estimation and segmentation chicken and egg problem, the motion field estimation and the segmentation task are initially treated separately. A local motion field estimation is carried out starting from a translational motion model by means of a block matching technique.[11] Concurrently, an initial static image partition results from a Bayesian adaptive segmentation algorithm that incorporates a Markov Random Field model (MRF) to account for spatial homogeneity constraints and an adaptive mechanism to allow the local average luminance of each region to slowly vary in spatial domain.[12] A MDL estimator provides for the number of the MRF labels to be used.
The two information data streams provided by the static image partition and global/local motion estimation are then gathered together by formulating the motion estimation and segmentation problem under the MDL principle. In this way, the motion parameter estimation is carried out concurrently with the region of support identification. Given the initial set of image region with associated parameters, calculated as previously described, the regions are merged using the reliability measure to reach the final spatio-temporal segmentation.
The paper is organised as follows. Section 2 describes the motion estimation algorithm adopted whereas section 3 is devoted to the description of the luminance based image partition. Section 4 is devoted to the description of the affine motion parameter estimation whereas the proposed region merging algorithm is introduced in section 5. Results and conclusions are discussed in the final sections.

2. GLOBAL/LOCAL MOTION ESTIMATION

In natural scenes, where changes in camera position, orientation and focal length continuously occur, a global motion compensation is very important in the estimation of "physical" motion fields. Global motion parameters are therefore evaluated as described in.[10]
After global motion compensation, the local motion field is estimated by means of a block matching technique.[11] The algorithm exploits the spatial and temporal coherence characteristics of physical motion fields and provides a very smooth estimated motion field with a reduced computational complexity.

3. BAYESIAN INTENSITY-BASED IMAGE PARTITION

The initial image partition results from a generalised *k-means clustering* algorithm where spatial constraints are included by the use of an MRF model. Moreover, in order to account for local intensity variation of the image, an iterarive procedure involving averaging over a sliding window is introduced.
First, the a posteriori probability density function for the distribution of regions given the observed image is defined. The density has two component, one forces the region intensity to be close to the data and the other imposes spatial continuity. The algorithm then alternates between maximising the *a posteriori* probability density and estimating the cluster centres. The spatial adaptivity is enforced by locally estimated cluster center means, i.e., they are calculated considering pixels inside a $d \times d$ window centered at the pixel of interest; in this way, the adaptive mechanism allows the local mean of each class to vary in spatial domain. The maximization is carried out at every point in the image and the procedure is repeated until convergence is reached.
More specifically the algorithm works as follows. The initial estimate of the partition map is obtained using the classical *k-means clustering* algorithm. An MRF model is then exploited to include spatial connectivity among elements in the same class. The region labelling is obtained by the point-by-point minimization of the following energy function, given the luminance value I and the current labelling at all other points:

$$\frac{1}{2\sigma^2}[I_{\mathbf{x}} - \mu_{\mathbf{x}}^{l_{\mathbf{x}}}]^2 + \sum_{\mathbf{y}\in\eta_l(\mathbf{x})} V_l(\mathbf{x},\mathbf{y}) \tag{1}$$

$$V_l(\mathbf{x},\mathbf{y}) = \begin{cases} -\beta & \text{if } l_{\mathbf{x}} = l_{\mathbf{y}} \\ +\beta & \text{if } l_{\mathbf{x}} \neq l_{\mathbf{y}} \end{cases} \tag{2}$$

where $I_{\mathbf{x}}$ is the luminance value at pixel $\mathbf{x}$, $\mu_{\mathbf{x}}^{l_{\mathbf{x}}}$ is the mean of region l calculated inside the sliding window centred at pixel $\mathbf{x}$, σ^2 is the noise variance, $\eta_l(\mathbf{x})$ is the first-order neighbourhood (i.e., *4-connected* set of points) of pixel $\mathbf{x}$, and $l_{\mathbf{x}}, l_{\mathbf{y}}$ are the labels at pixels $\mathbf{x}$ and $\mathbf{y}$. The parameter β is positive, so that two neighbouring pixels are more likely to belong to the same class than two different classes. The first term of Eq.(1) measures the fitting of the mean gray value of region l to the observed luminance value, while the second one accounts for the region spatial connectivity. The Iterated Conditional Modes (ICM) algorithm provides the maximum *a posteriori* (MAP) estimation of the intensity-based image partition.[14]
This procedure provides a set $\mathcal{R}$ of N_R^{max} regions characterized by slowly varying intensity functions.

4. AFFINE MOTION PARAMETER ESTIMATION

As motion estimation and image segmentation are interdependent tasks, they should be carried out jointly. The MDL priciple allows to solve the problem as motion parameter (real valued system parameters) and number of regions (real valued system parameters) can be estimated without separate hypothesis testing. The maximum number of regions inside the scene and the regions shape are given by the intial static segmentation. The initial motion information provides for the affine motion parameter initialization. A MDL-based motion estimation and segmentaton formulation can be given if the scene is meant to be composed by objects moving in front of the camera with an affine model. The initial static segmentation regions are combined in order to estimate the set of motion parameter vectors that provides the minimum description length of the initial local motion field.
In the following, the observed data, i.e., the motion vectors given by the initial motion estimation, are supposed to be given by several underlying affine models. If $\mathcal{R}$ is a set of disjoint regions that partition the image, every region is generated by a parametric model θ_R. If the prior probability distribution for region R is P_{θ_R}, then the combined segmentation and estimation problem under MDL is

$$\min_{\{\theta_{R_n}, N_R\}} \{\sum_{n=1}^{N_R} [-\log_2 P_{\theta_{R_n}}(\mathbf{d}(\mathbf{x})/\mathbf{x} \in R_n)] + \frac{1}{2} N' \log_2(N_p)\} \tag{3}$$

where N_R is the number of regions, $d(\mathbf{x})$ is the set of motion vectors inside region R, N_p is the total number of motion vectors, N'_R is the free parameter number assuming N_R model classes. In this application, $N'_R = hN_R - 1$, where h is an integer number proportional to the model complexity description. If the initial motion field is considered as a corrupted realization of the underlying process, the motion difference can be used as an indicator of the deviation between the estimated and the underlying segmentation. If the motion vectors components errors are assumed to be statistically indipendent with a Gaussian distribution, the first term is the coding length of the modelling error while the second term is proportional to the ideal coding of the motion parameters.
Minimization procedure - The minimization process is inherently a combination of two problems: parameter estimation and hypothesis testing. The exhaustive searching in the parameter space is computationally infeasible. If the motion discontinuity are assumed to be a subset of the luminance edges, the objects in the scene should be composed by a suitable set of adjacent regions provided by the intensity-based image partition. In this way, the number of objects hypothesis are reduced. A bottom up procedure can then be implemented to minimize the objective function (3) to obtain the suboptimal MDL estimator solution.
Initial Segmentation - The initial segmentation reduces the searching space of the available segmentation. If the initial static segmentation algorithm is assumed not to have failed in the estimation of the regions with homogeneous luminance, the maximum region number in the final motion segmentation is determined by the number of regions in the initial static segmentation. Initially, the motion segmentation is assumed to be represented with N_R^{max} motion models as the initial elementary regions.
Adjacency Graph-based Minimization - Given the initial static segmentation, an adjacency graph is constructed, where every node represent a region while each arc connecting two neighbouring regions (*4-connected* set of points) has a weight which is the ideal coding cost reduction, in bits, if the two regions are to be merged. Starting from the initial adjacency graph, the MDL-based merging process operates as follows:

1. find the arc with the largest coding cost reduction among all arcs in the graph,

2. merge two adjacent nodes connected by the arc,

3. recompute coding cost reduction of the arc associated with the new grap,

4. go back to step 1 and repeat until the coding cost reduction of all the arcs are negative.

Motion Parameter Estimation - The parameters are obtained by setting the partial derivatives of the description length to zero with respect to modelling error variance and affine parameters within the interested two regions.

Robust Affine Parameter Estimation - As the initial motion estimation is poorly accurate at motion boundaries, the initial motion field is corrupted by outlier motion vectors that can damage the final motion vector estimation and segmentation. Given the set of regions provided by the motion segmentation, the affine motion parameters for each region are refined by means of a robust regression method.

Let $d(\mathbf{x}) = (d_x(\mathbf{x}), d_y(\mathbf{x}))$ be the estimated motion vector at pixel $\mathbf{x} = (x, y)$ and $d_\theta(\mathbf{x}) = (d_{x,\theta}(\mathbf{x}), d_{y,\theta}(\mathbf{x}))$ be the motion vector generated at pixel $\mathbf{x}$ by the affine motion parameter vector $\theta = (a_{x0}, a_{xx}, a_{xy}, a_{y0}, a_{yx}, a_{yy})$ where

$$\begin{aligned} d_{x,\theta}(\mathbf{x}) &= a_{x0} + a_{xx}x + a_{xy}y \\ d_{y,\theta}(\mathbf{x}) &= a_{y0} + a_{yx}y + a_{yy}y. \end{aligned} \tag{4}$$

To identify the affine motion parameters from a given set of motion vectors, a "Weighted Least Squares" method is adopted.[13] For every object $R \in \mathcal{R}$ provided by the MDL based motion estimation, an affine motion parameter vector θ_R^0 is estimated by means of a least square procedure. The residual error between the actual motion vector $d(\mathbf{x})$ and the displacement $d_{\theta_R^0}(\mathbf{x})$ given by the estimated motion parameters θ_R^0 is calculated for every pixel $\mathbf{x}$ in the region R as

$$\begin{aligned} \epsilon_{x,\theta_R^0}(\mathbf{x}) &= d_x(\mathbf{x}) - d_{x,\theta_R^0}(\mathbf{x}), \\ \epsilon_{y,\theta_R^0}(\mathbf{x}) &= d_y(\mathbf{x}) - d_{y,\theta_R^0}(\mathbf{x}). \end{aligned}$$

Hence, a *robust* estimation of these residual errors standard deviation σ_R is carried out.[13]

Each considered pixel is then allocated a weight which is inversely proportional to the residual error

$$w_k(\mathbf{x}) = [1 - (E_{\theta_R^{k-1}}(\mathbf{x}))/(\sigma_R)^2]^2$$

where

$$E_{\theta_R^k}(\mathbf{x}) = \|d(\mathbf{x}) - d_{\theta_R^k}(\mathbf{x})\|^2. \tag{5}$$

At the kth iteration the new set of affine parameters for the *inlier* pixels of object R is obtained by weighted least squares. The process "weighting coefficients calculation - weighted least squares estimation" is then iterated until the number of motion vectors inside the region reaches an asymptotic value. The resulting motion parameter vector θ_R is considered as the robust estimation of parameters vector of region R.

5. THE PROPOSED ROBUST MERGING CRITERION

Some region merging techniques for spatio-temporal segmentation use the temporal information available only in the parameter space. The method proposed in,[34] defines as region similarity measure the distances in the motion parameter space. The merging decision is based on a clustering procedure and regions assigned to the same cluster are merged into a single moving object. This method is obviously sensitive to errors in motion estimation and to the distance measure used in the clustering process. Also, depending on the scene and the motion model chosen, similar optical flow fields may be represented by very different sets of motion parameters. In other words, the parameterisation of the motion field needs not have a unique solution and the hypothesis that the motion parameters represent the entire motion information may indeed be wrong. This implies that two regions, which

are moving in a similar way, may turn out to have very different motion parameters and thus will not be merged by the clustering procedure.
The proposed algorithm overcomes this problem introducing a reliability measure that indicates to what extent an affine motion parameter represents the motion of one image region; it takes into account both the displaced frame difference (DFD) and the motion difference (MD), i.e., the motion error between the estimated motion vector and the motion vector generated by the parametric motion model of the region. The reliability measure associated to a motion parameter θ_j when applied to an image region R_i is

$$C_m(i,j) = (\frac{N_i}{2}\log_2(\frac{DFD(i,j)}{N_i}), N_i \log_2(\frac{MD(i,j)}{2N_i})) \tag{6}$$

where N_i is the pixel number of region R_i and

$$DFD(i,j) = \sum_{\mathbf{x}\in R_i} [I_t(\mathbf{x}) - I_{t-1}(\mathbf{x} - d_{\theta_j}\mathbf{x})]^2 \tag{7}$$

and

$$MD(i,j) = \sum_{\mathbf{x}\in R_i} ||d(\mathbf{x}) - d_{\theta_j}(\mathbf{x})||^2 \tag{8}$$

where I_t is the luminance of the frame at time t, $d(\mathbf{x}) = (d_x(\mathbf{x}), d_y(\mathbf{x}))$ is the estimated motion vector at pixel $\mathbf{x} = (x,y)$ and $d_{\theta_j}(\mathbf{x}) = (d_{x,\theta_j}(\mathbf{x}), d_{y,\theta_j}(\mathbf{x}))$ is the motion vector generated at pixel $\mathbf{x}$ by the affine motion parameter vector θ_j (Eq. 4). In the view of the MDL principle, this reliability measure is proportional to the number of digits it takes to write down the luminance value of the actual frame given the previous frame (DFD) and the motion field (MD) inside region R_i given the parameter vector of the affine motion model. This reliability measure can be exploited as a criterion for a motion based region merging.
Given an initial set of image regions with associated motion parameters, the reliability measure is used to construct an adjacency directed graph where the nodes represent the regions and directed arcs are weighted with the set of description length reduction where the motion parameter vector associated to the node the arc starts from is applied to the region the arc is directed towards. The clustering strategy employs two rules, referred to as "hard" rule and "soft" rule, respectively.
The "hard" rule merges two adjacent regions only if the motion parameter vector of one region is representative of the motion of the other *and* viceversa. Here two hypothesis H_1 and H_2 are tested. H_1 means that R_1, R_2 are to be merged and so the motion parameter vector of one region is representative of the motion of the other and viceversa; H_2 means that R_1, R_2 are separated and each affine parameter vector represents its region motion alone.

The MDL-test is based on the affine parameter reliability masure as follows. Given the two arcs connecting each pair of nodes, a term by term sum of the description length values of (DFD, MD) is carried out

$$C_H = \{C_m(1,2) - C_m(1,1)\} + \{C_m(2,1) - C_m(2,2)\} - \frac{1}{2}h\log_2(N_p) \tag{9}$$

where h is proportional to the model complexity (i.e., number of regions) and N_p is the number of pixels inside the image. Under the MDL principle, Eq. (9) is equivalent to calculate the data description length reduction when two region are merged. $C_m(i,j) - C_m(i,i)$ is the coding cost reduction that quantifies how much the motion parameter of region j represents the motion of region i. Ideally, if the regions were coherently moving, then the coding cost of at least the luminance or motion vectors of the region should be similar. Moreover, if the initial motion field estimation has totally failed, the affine parameters associated to the region are not reliable

as they are calculated from motion vectors and do not describe well the motion of the region. If the coding cost of the proposed motion parameter vector is lower than the one associated with the region parameters, then the parameter vector indentifies the underlying motion trasformation that has provided the observed data (in this case the frame luminance values). The last term is proportional to the decription length reduction when the number of model that describes the data is lessen by one.
More specifically, the lowest indicator is then chosen as the coding cost reduction which determines the level of motion coherence between the two regions; if it is lower than zero, the regions are merged:

$$\min_{\{DFD,MD\}} C_H = \begin{cases} < 0 & H_1 \text{ is verified} \\ > 0 & H_2 \text{ is verified} \end{cases} \tag{10}$$

The "soft" rule relaxes the "hard" condition for merging regions as it merges regions if at least the motion parameter vector of one region is representative of the motion of the other.
Again, two hypothesis H_3 and H_4 are tested. H_3 means that R_1, R_2 are to be merged and so the motion parameter vector of one region is representative of the motion of the other *or* viceversa; H_4 means that R_1, R_2 are separated and each affine parameters represent its region motion alone:

$$C_S^{12} = C_m(1,2) - C_m(1,1) - \frac{1}{2}h\log_2(N_p) \tag{11}$$

$$C_S^{21} = C_m(2,1) - C_m(2,2) - \frac{1}{2}h\log_2(N_p) \tag{12}$$

For both arcs, the greater element in the coding cost reduction pair is chosen as the motion coherence level. If at least one of the two is lower than zero, the two regions are considered as coherently moving:

$$\begin{cases} \text{if } \min_{\{DFD,MD\}} C_S^{12} \vee \min_{\{DFD,MD\}} C_S^{21} < 0 & H_3 \text{ is verified} \\ \text{if } \min_{\{DFD,MD\}} C_S^{12} \wedge \min_{\{DFD,MD\}} C_S^{21} > 0 & H_4 \text{ is verified} \end{cases} \tag{13}$$

The graph-based region clustering,[8 9] is first carried out by applying the "hard" rule. The two connected nodes which show the greatest motion coherence reliability are merged. The graph weights are then updated: the motion parameter vector of the new formed region is calculated among the motion vectors belonging to it and the arc weights connecting the region to the other ones are calcultated as in Eq.(9). This procedure is iterated until all the description lengths are lower than zero. The remaining regions are then merged according to the "soft" rule. Again, the best motion coherence reliability value is searched for. After two region merging, the resulting new one is not considered but the most negative weight is searched among all the other arcs not connected to it. When all the weights are positive, the graph is updated as for the "hard rule", determing the arcs weight of the new regions when they are to be merged with the old ones. The clustering process is then carried out on the new graph; it stops when the arc weights of the new graph are all positive.

6. SIMULATION RESULTS

The proposed region merging technique for spatio-temporal segmentation of video sequences has been tested on the CIF sequences "Flower Garden", "Foreman" and "Table Tennis".
Fig. 1 shows one frame of the "Flower Garden" sequence. As mentioned before, the proposed algorithm expects an initial set of regions as input. The initial image partition, shown in Fig. 2, is carried out with $K = 8$ MRF labels and it results 1069 regions. The affine motion parameter estimation and region segmentation estimates the initial motion models with associated regions of support as shown in Fig. 3. The proposed method merges these regions to form meaningful objects. Fig. 4 shows the segmentation produced by the "hard" rule clustering strategy. The segmentation is furtherly improved by applying the "soft" rule as shown in Fig. 5. Seven objects

have been obtained, where the tree, the flower bed and the background houses can be identified. Figs 6, 9 show original frame for "Table tennis" and "Foreman", Figs 7, 10 the intensity-based image partition, while Figs 8, 11 show the final spatio-temporal segmentation maps. While the spatio-temporal merging technique provides a "Table Tennis" frame segmentation in which the arm, the ball and the background are identified, the "Foreman" frame segmentation is over-segmented. This event could be explained by the fact that the affine model is perhaps insufficient to describe the motion of the sequence. A more complex motion model should better suite the motion trasformation.

7. CONCLUSIONS

In this work the problem of spatio-temporal segmentation of video sequences is addressed to identify moving objects in a scene. An affine motion model reliability measure is proposed as a criterion to coherently merge moving image regions in a Minimum Description Length framework. The initial motion field estimation and static image segmentation are carried out as separate tasks; then a formulation of the motion estimation and segmentation tasks is proposed in the view of the MDL principle. A merging stage which uses the proposed reliability measure is finally carried out, based on a weighted, directed graph, in order to decide the opportunity to merge the image regions. The simulation results show the effectiveness of the proposed region merging technique. We are now developing an extension of the merging technique in order to identify and cluster coherently moving regions that are not spatially connected.

REFERENCES

1. R. Leonardi, "Region Based Image and Video Compression", in *Proc. of the 1995 International Conference on Digital Signal Processing*, Cyprus, Jun. 1995.
2. P. Salembier, M. Pardas, "Hierarchical Morphological Segmentation for Image Sequence Coding", *IEEE Transactions on Image Processing*, Vol. 3, No. 5, pp. 639-651, Sept. 1994.
3. J. Y. Wang, E. H. Adelson, "Representing Moving Images with Layers", *IEEE Transactions on Image Processing*, Vol. 3, No. 5, pp. 625-638, Sept. 1994.
4. F. Pedersini, A. Sarti, S. Tubaro, "Combined Motion and Edge Analysis for a Layered Based Representation of Image Sequences", in *Proc. IEEE International Conference on Image Processing*, pp. 921-924, Lousanne, Switzerland, Sept. 1996.
5. F. Dufaux, F. Moscheni, Andrew Lippman, "Spatio-Temporal Segmentation Based on Motion and Static Segmentation", in *Proc. IEEE International Conference on Image Processing*, pp. 306-309, Washington D.C., Oct. 1995.
6. J. G. Choi, et al., "Spatio-Temporal Video Segmentation Using a Joint Similarity Measure", IEEE Trans. on Circuits and System for Video Technology, Vol. 7, No. 2, pp. 279-286, Apr. 1997.
7. F. Morier, et al., "Robust Segmentation of Moving Image Sequences", in Proc. IEEE ICIP-97, pp. 719-722, Santa Barbara, California, Oct. 1997.
8. F. Moscheni, et al., "Robust region Merging for Spatio-Temporal Segmentation", in Proc. IEEE ICIP-96, pp. 501-504, Lausanne, Switzerland, Sept. 1996.
9. H. Zheng, S. D. Blostein, "Motion-Based Object Segmentation and Estimation Using the MDL Principle", *IEEE Transactions on Image Processing*, Vol. 4, No. 9, pp. 1223-1235, Sept. 1995.
10. P. Migliorati, S. Tubaro, "Multistage Motion Estimation for Image Interpolation", *Signal Processing: Image Communication*, Vol. 7, No. 3, pp. 187-199, 1995.
11. G. de Haan, H. Huijgen, "New Algorithm for Motion Estimation", in *Proc. Third International Workshop on HDTV*, Torino, Italy, 1989.
12. T. N. Pappas, "An Adaptive Clustering Algorithm for Image Segmentation", *IEEE Transactions on Signal Processing*, Vol. 40, No. 4, pp. 901-914, Apr. 1992.
13. P. Meer, D. Mintz, Azriel Rosenfeld, "Robust Regression Methods for Computer Vision: A Review". *International Journal of Computer Vision*, Vol. 6, No. 1, pp. 59-70, 1991.
14. J. Besag, "On the Statistical Analysis of Dirty Pictures", *Journal of Royal Statistical Society* B 48, No. 3, pp. 259-302, 1986.

Figure 1. "Flower Garden" initial frame.

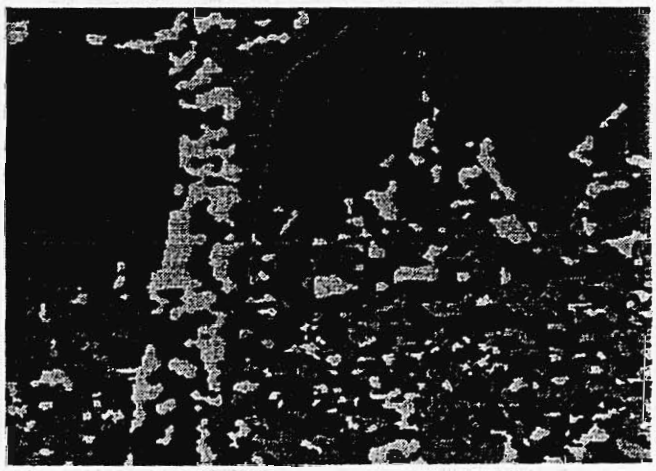

Figure 2. "Flower Garden" static segmentation ($K = 8, \sigma = 32, N_R^{max} = 1069$).

Figure 3. "Flower Garden" affine motion estimation output map.

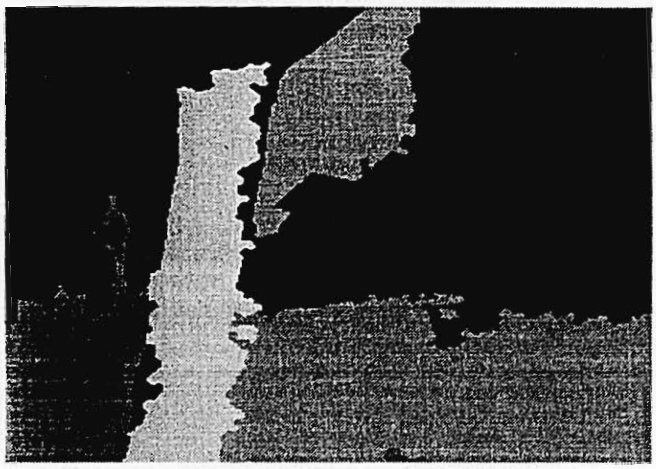

Figure 4. "Flower Garden" "hard" rule clustering output map.

Figure 5. "Flower Garden" "soft" rule clustering output map ($N_R = 7$).

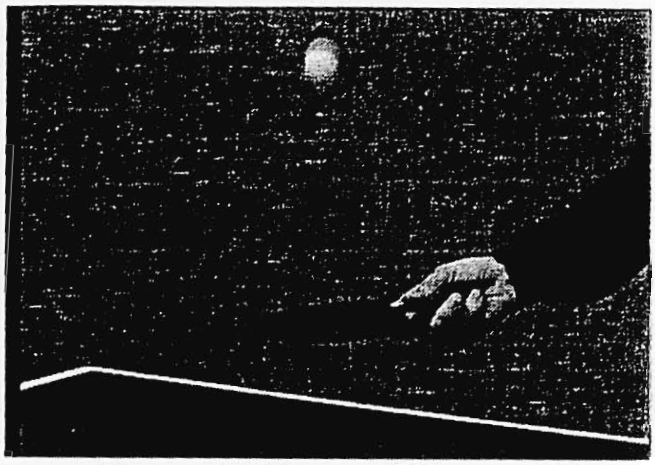

Figure 6. "Table tennis" initial frame.

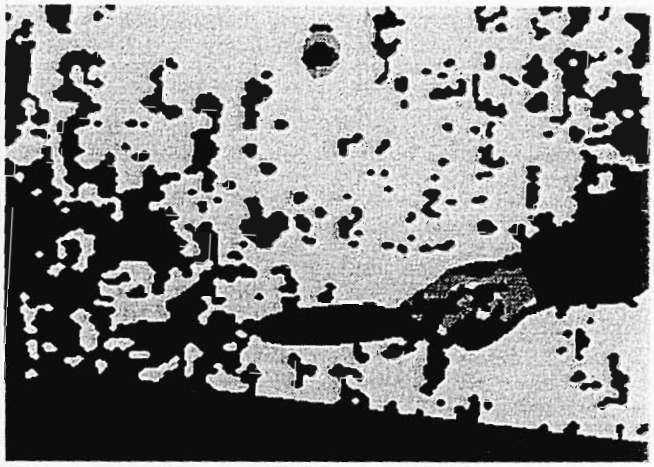

Figure 7. "Table tennis" static segmentation ($K = 6, \sigma = 32, N_R^{max} = 192$).

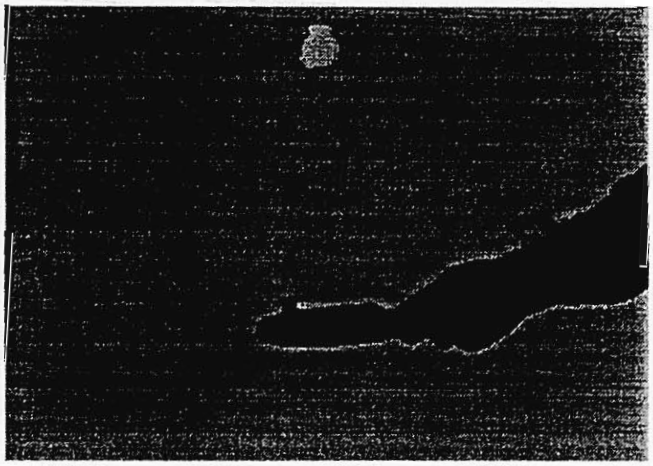

Figure 8. "Table tennis" "soft" rule clustering output map ($N_R = 4$).

Figure 9. "Foreman" initial frame.

Figure 10. "Foreman" static segmentation ($K = 13, \sigma, N_R^{max} = 397$).

Figure 11. "Foreman" "soft" rule clustering output map ($N_R = 34$).