Moving scene segmentation using Median Radial Basis Function Network

Adrian G. Borş and Ioannis Pitas

Department of Informatics, University of Thessaloniki
Thessaloniki 54006, Greece

Abstract— **Various approaches were suggested for simultaneous optical flow estimation and segmentation in image sequences. In this study, the moving scene is decomposed in different regions with respect to their motion, by means of a pattern recognition scheme. The inputs of the proposed scheme are the feature vectors representing still image and motion information. The classifier employed is the Median Radial Basis Function (MRBF) neural network. Each class corresponds to a moving object. An error criterion function derived from the probability estimation theory and related to the moving scene model is used as cost function. Marginal median and median of the absolute deviations estimators are employed for estimating the basis function parameters.**

I. Introduction

Motion representation and modeling is an important step towards dynamic image understanding. Many of the algorithms proposed for joint estimation of the optical flow and moving object segmentation depend on the choice of various parameters. Mainly iterative algorithms are used, employing stochastic or deterministic processing that lead to the minimization of a cost function [1].

In this study, a cost function associated with the minimization of a global criterion is proposed for simultaneous estimation of the optical flow and segmentation of the moving objects. The image is first partitioned in block sites situated on a rectangular lattice. Each block site is associated with a five dimensional feature vector describing the position, the gray level and the local motion information. The proposed method is based on the unsupervised classification of the feature vectors by considering the displaced frame difference as well. The classification is done according to a decision criterion derived from the Bayesian theory [2] and representing a metric in the parameter space.

A radial basis functions (RBF's) decomposition is known to be a good functional approximator and has been used in many applications. The first layer units implement Gaussian functions and the output units are assigned to the moving objects. The classification criterion connects the Gaussian parameters to the set of feature vectors drawn from the image sequence. Each basis function has associated a moving region. The moving regions are connected by the output units in order to model moving objects. We consider the MRBF network [3] for modeling the optical flow and moving object segmentation from the image sequence. The efficiency of the proposed algorithm when compared to the classical learning algorithm in RBF networks has been shown in [3].

II. The classification criterion

We consider the video frame partitioned in blocks situated on a rectangular grid. We associate a feature vector, describing the local image sequence properties, to each block site. This vector contains a still image feature vector $\mathbf{S}_{IJ}$ and a motion vector $\mathbf{M}_{IJ}$:

$$\mathbf{u}_{IJ} = [\mathbf{S}_{IJ}, \mathbf{M}_{IJ}]. \tag{1}$$

The still image feature vector includes the block site coordinates and its mean gray level :

$$\mathbf{S}_{IJ} = [I, J, l_{IJ}], \tag{2}$$

and the motion vector contains two components, corresponding to the local motion of the block :

$$\mathbf{M}_{IJ} = [m_{x,IJ}, m_{y,IJ}]. \tag{3}$$

They can be calculated by any optical flow algorithm, e.g., by block matching. The motion vector that minimizes the displaced frame difference is chosen [1].

A moving scene can be seen as made up of regions with different motion parameters. We assume that each frame can be segmented into L subsets, forming moving regions, denoted as $X_1, \ldots, X_L$. The moving objects are considered as compact moving entities, consisting of one or more moving regions. Each moving object is assigned to a class. Each subset X_k has associated a five-dimensional representative vector μ_k, describing the optical flow and the segmentation information of a certain moving region :

$$\mu_k = [\mathcal{S}_k, \mathcal{M}_k]. \tag{4}$$

The still image feature vector $\mathcal{S}_k$ is directly related to the segmentation label of the moving region k.

Let us denote by $\hat{\mathcal{S}}_k$ the lebel estimate of the moving region k and by $\hat{\mathcal{M}}_k$ the estimate of the optical flow associated with the same moving region. A block site B_{IJ} is considered as belonging to a moving region k, $B_{IJ} \in X_k$, if it maximizes the *a posteriori* probability of the optical flow $\hat{\mathcal{M}}_k$ and moving region segmentation $\hat{\mathcal{S}}_k$ joint estimation, denoted as $P(\hat{\mathcal{M}}_k, \hat{\mathcal{S}}_k | f_{t-1}, f_t)$, when compared with the probabilities associated to the other moving regions :

$$P(\hat{\mathcal{M}}_k, \hat{\mathcal{S}}_k | f_{t-1}, f_t) > P(\hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j | f_{t-1}, f_t) \tag{5}$$

for $j = 1, \ldots, L$, $j \neq k$, where L is the number of moving regions.

The moving regions are merged based on a neighboring criterion in order to describe moving objects. Let us denote by $\hat{\mathcal{T}}_k$ the estimate of the optical flow and segmentation label associated with a moving object. We consider a neighboring measure $V(X_l, X_k)$ between two subsets representing two moving regions, X_l and X_k :

$$V(X_l, X_k) = V(X_k, X_l) = \sum_{B_{IJ} \in X_l} |\mathcal{N}_{IJ} \cap X_k| \qquad (6)$$

where $|\mathcal{N}_{IJ} \cap X_k|$ represents the number of block sites of a moving region X_k, situated in a certain neighborhood $\mathcal{N}_{IJ}$ of the block site B_{IJ}, from the moving region X_l [4]. This measure expresses the boundary length between two moving regions. If two moving regions X_k and X_l do not have any common boundary, then $V(X_l, X_k) = 0$. We define a moving region as a moving object if it contains a compact area in the image. In this case, the probability of estimating the optical flow and moving object segmentation $P(\hat{\mathcal{T}}_k | f_{t-1}, f_t)$ is :

$$\begin{aligned} V(X_k, X_k) \;&=\; \max_{i=1}^{L} V(X_k, X_i) \text{ then} \\ P(\hat{\mathcal{T}}_k | f_{t-1}, f_t) &= P(\hat{\mathcal{M}}_k, \hat{\mathcal{S}}_k | f_{t-1}, f_t). \end{aligned} \qquad (7)$$

A moving object k contains a moving region X_l, if X_l has the maximal neighborhood measure (6) with X_k :

$$\begin{aligned} V(X_l, X_k) \;&=\; \max_{i=1}^{L} \; V(X_l, X_i) \text{ then} \\ P(\hat{\mathcal{T}}_k | f_{t-1}, f_t) &= \lambda_k P(\hat{\mathcal{M}}_k, \hat{\mathcal{S}}_k | f_{t-1}, f_t) + \\ &+ \lambda_l P(\hat{\mathcal{M}}_l, \hat{\mathcal{S}}_l | f_{t-1}, f_t) \end{aligned} \qquad (8)$$

where λ_k, λ_l are the parameters weighting the contribution of each moving region probability to the moving object probability. This condition can be extended for moving objects containing many moving regions.

Let us express the *a posteriori* probabilities from (5) with respect to the features extracted from the image sequence. After applying the Bayes rule, each of the *a posteriori* distributions in (5) can be factored as follows :

$$\begin{aligned} P(\hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j | f_t, f_{t-1}) &= \frac{P(f_t | f_{t-1}, \hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j) P(\hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j | f_{t-1})}{P(f_t | f_{t-1})} = \\ &= \frac{P(f_t | f_{t-1}, \hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j) P(\hat{\mathcal{M}}_j | \hat{\mathcal{S}}_j, f_{t-1}) P(\hat{\mathcal{S}}_j | f_{t-1})}{P(f_t | f_{t-1})}. \end{aligned} \qquad (9)$$

where $P(\hat{\mathcal{S}}_j | f_{t-1})$ represents the *a priori* probability of the segmentation and $P(\hat{\mathcal{M}}_j | \hat{\mathcal{S}}_j, f_{t-1})$ is the probability of the optical flow estimation depending on the segmentation and image [2].

Each of the above conditional probabilities can be expressed as an energy functional :

$$P(\mathbf{X}) = \frac{1}{Z} \exp\left[-\frac{E(\mathbf{X})}{\beta}\right] \qquad (10)$$

where Z is a normalizing constant and β is a constant controlling the properties of $E(X)$. The probability estimation problem (5) is converted into the minimization of an energy functional :

$$E_j = E(f_t | f_{t-1}, \hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j) + E(\hat{\mathcal{M}}_j | \hat{\mathcal{S}}_j, f_{t-1}) + E(\hat{\mathcal{S}}_j | f_{t-1}). \qquad (11)$$

In order to minimize E_j, all three components should be simultaneously minimized. This corresponds to the optical flow and image sequence segmentation map simultaneous processing.

A performance criterion is related to the total squared error minimization in the feature space [5]. The energy functional in (11) can be expressed as a clustering metric in the feature space. This metric relates the moving region feature vectors (4) to the block site feature vectors (1).

The energy $E(f_t | f_{t-1}, \hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j)$ in (11) is represented as a weighted function of the displaced frame difference, denoted as $WDFD(\hat{\mathcal{M}}_j)$ and corresponding to the moving object j:

$$WDFD(\hat{\mathcal{M}}_j) = \sum_{I=0}^{n_x-1} \sum_{\substack{J=0 \\ B_{IJ} \in X_j}}^{n_y-1} [w_{IJ}(\hat{\mathcal{M}}_j) d_{IJ}(\hat{\mathcal{M}}_j)]^2 \qquad (12)$$

where $d_{IJ}(\hat{\mathcal{M}}_j)$ is the DFD estimate for the motion vector $\hat{\mathcal{M}}_j$ and $w_{IJ}(\hat{\mathcal{M}}_j)$ is a weighting factor corresponding to the block B_{IJ} and depending on the motion vector $\hat{\mathcal{M}}_j$. We consider as weighting factor a reliability coefficient which measures the confidence on the output of the block matching result :

$$w_{IJ}(\hat{\mathcal{M}}_j) = \frac{d_{IJ}(\hat{\mathcal{M}}_j)}{\sum_{k=-\frac{S_x}{2}}^{\frac{S_x}{2}} \sum_{l=-\frac{S_y}{2}}^{\frac{S_y}{2}} d_{I+k,J+l}(\hat{\mathcal{M}}_j)} \qquad (13)$$

where $S_x \times S_y$ is the search region for the block matching algorithm. This coefficient is small when we have a good matching and large in the case of poor matching.

The energy functional $E(\hat{\mathcal{M}}_j | \hat{\mathcal{S}}_j, f_{t-1})$ in (11) is associated with motion vector clustering :

$$\begin{aligned} E(\hat{\mathcal{M}}_j | \hat{\mathcal{S}}_j, f_{t-1}) = \sum_{I=0}^{n_x-1} \sum_{\substack{J=0 \\ B_{IJ} \in X_j}}^{n_y-1} (\mathbf{M}_{IJ} - \hat{\mathcal{M}}_j)^T (\mathbf{M}_{IJ} - \\ - \hat{\mathcal{M}}_j) \quad . \end{aligned} \qquad (14)$$

The cost function associated to the moving region segmentation $E(\hat{\mathcal{S}}_j | f_{t-1})$ is related to vector clustering with respect to their gray level, and their geometrical proximity :

$$E(\hat{\mathcal{S}}_j | f_{t-1}) = \sum_{I=0}^{n_x-1} \sum_{\substack{J=0 \\ B_{IJ} \in X_j}}^{n_y-1} (\mathbf{S}_{IJ} - \hat{\mathcal{S}}_j)^T (\mathbf{S}_{IJ} - \hat{\mathcal{S}}_j). \qquad (15)$$

By replacing the expressions (12), (14) and (15) in (11) we obtain the energy associated with the moving region j :

$$\begin{aligned} E_j(\mathbf{u}_{IJ}) = \sum_{I=0}^{n_x-1} \sum_{\substack{J=0 \\ B_{IJ} \in X_j}}^{n_y-1} (\mathbf{u}_{IJ} - \hat{\mu}_j)^T (\mathbf{u}_{IJ} - \hat{\mu}_j) + \\ + WDFD(\hat{\mathcal{M}}_j) \end{aligned} \qquad (16)$$

where $WDFD(\hat{\mathcal{M}}_j)$ is provided in (12).

III. Median Radial Basis Function Network

The cost function (16) corresponds to image partition in moving regions. If we take into account the covariance matrix and we express (16) as an unnormalized probability (10), we obtain a so called radial basis function (RBF) :

$$\phi_j(\mathbf{u}) = \exp\left[-(\mathbf{u}-\hat{\mu}_j)^T\hat{\mathbf{\Sigma}}_j^{-1}(\mathbf{u}-\hat{\mu}_j) - WDFD(\hat{\mathcal{M}}_j)\right], \tag{17}$$

where $\hat{\mu}_j$ is the center vector and $\hat{\mathbf{\Sigma}}_j$ is the covariance matrix. Each basis function must be defined such that it maximizes the probability of the optical flow estimation and segmentation of a certain moving region $P(\hat{\mathcal{M}}_k, \hat{\mathcal{S}}_k | f_t, f_{t-1})$.

The output layer implements a weighted sum of hidden unit outputs, scaled to the interval $(0,1)$ by a sigmoidal function :

$$Y_k(\mathbf{u}_{IJ}) = \frac{1}{1 + \exp\left[-\sum_{j=1}^{L}\lambda_{kj}\phi_j(\mathbf{u}_{IJ})\right]} \tag{18}$$

for $k = 1, \ldots, N$, where λ_{kj} is the parameter associated to the connection between the hidden unit j and the output unit k, L is the number of basis functions, N is the number of moving objects and $\phi_j(\mathbf{u}_{IJ})$ is provided in (17). The *a posteriori* probability of optical flow estimation and segmentation associated to each moving object $\hat{\mathcal{T}}_k$, is modeled by the output unit. The moving object probabilities estimation leads to finding the optical flow and segmentation map of the image sequence.

The structure of the network used for modeling the motion is represented in Fig. 1. A robust statistics-based training algorithm called Median Radial Basis Function (MRBF) is proposed in [3] for estimating the RBF network parameters. The basis function center updating is based on the marginal median LVQ algorithm. The marginal median LVQ algorithm orders the data samples associated to a center, on each dimension separately, and takes the median of the data as the new estimate for the center :

$$\hat{\mu}_k = \text{med}\ \{\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{p-1}\} \tag{19}$$

where $\mathbf{u}_{p-1}$ is the last data sample assigned to the basis function k, based on a minimal function E_j in (16). The median of the absolute deviations (MAD) estimator is used as a robust estimator for calculating the dispersion parameter :

$$\hat{\mathbf{r}}_k = \frac{\text{med}\ \{|\mathbf{u}_0 - \hat{\mu}_k|, \ldots, |\mathbf{u}_{p-1} - \hat{\mu}_k|\}}{0.6745} \tag{20}$$

where $\hat{\mathbf{r}}_k$ denotes the diagonal vector of the covariance matrix $\hat{\mathbf{\Sigma}}_k$ and 0.6745 is the scale parameter in order to make the estimator Fisher consistent for the normal distribution. A fast implementation algorithm for (19, 20) based on data sample histogram modeling is provided in [3].

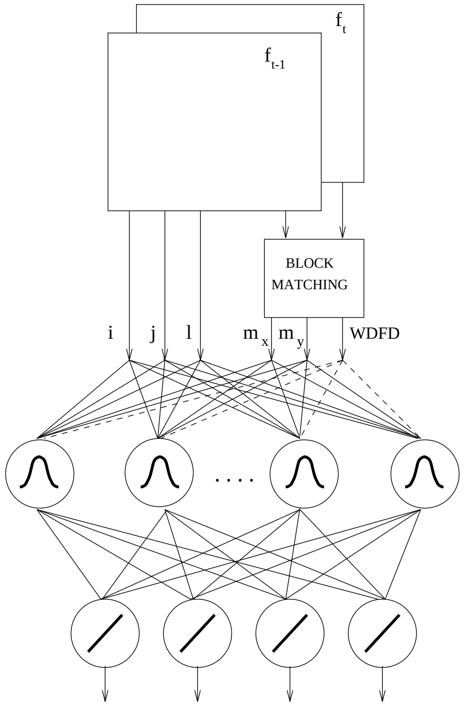

Fig. 1. The MRBF network structure for moving scene modeling : I and J are the position coordinates features assigned to a block site, l is the gray level and m_x, m_y are the motion vector components provided by the block matching algorithm.

By estimating the radial basis functions, the image is split in moving regions. A block site is assigned to that moving region corresponding to the most activated radial basis function. For each two moving regions we evaluate the boundary measure (6) between them, assuming a four block site neighborhood system. The moving regions which have a high interconectivity among their component block sites are considered as moving objects (7) and the other regions are merged according to (8) in order to represent moving objects. For each moving object is assigned an output unit. The block sites assigned to the moving regions are labeled according to (7) or (8). The labels are considered as targets for the backprogation algorithm, in order to estimate the λ_{kj} weights. These parameters smooth the connection regions among the parameter areas associated with the same moving object.

After the training stage is completed, when providing a feature vector (1), the maximum activated output will show the corresponding moving object. This procedure leads to the partition of the image sequence in moving objects. The network can be applied in other frames from the same image sequence, if their optical flow estimation and segmentation probabilities are consistent with those of the frames used in the training stage. It also can be used in a multiresolution (hierarchical) representation of the image. Let us consider that the network was trained on a certain image partition and afterwards we input feature vectors from a different block size partition. If the block size is large, the feature vector number is small and the training time will be short, but the segmentation will provide rough boundaries. The network can be trained with features corresponding to big blocks and afterwards applied on an image partition in blocks of smaller size.

IV. Simulation results

The MRBF network was tested on various image sequences [6]. In Fig. 2 (a) a frame from the "Trevor White" sequence is shown. The optical flow provided by the full search block matching algorithm when considering 8 × 8 blocks, is shown in Fig. 2 (b). The feature vectors are drawn from the second and seventh frames of this sequence. After evaluating the cost function with respect to the feature vectors according to the procedure described in Section II, the learning algorithm described in Section III is used for estimating the parameters of the MRBF network. The segmented moving objects and the optical flow smoothed by the MRBF network are shown in Figs. 2 (c) and (d). For comparison purposes, the moving object segmentation and the optical flow provided by ICM [7], [8] for the same image sequence are shown in Figs. 2 (e) and (f).

The MRBF network processes entire image regions assigned to the same moving object and exploits the interdependency among their block sites. The MRBF network provides smooth and accurate optical flows. The feature extraction and MRBF training time when using the algorithm described in [3] for this example is 23.4 seconds on a Silicon Graphics Indy Workstation. However, the trained network can be applied on frames whose optical flow and moving object probability is consistent with that obtained in the training stage and the average time per frame is lower in this case. The total number of necessary parameters for the MRBF network is $(10 + N)L$, where L is the number of hidden units and N that of outputs. The MRBF network requires 112 parameters instead of 3072 parameters used by the ICM algorithm, for representing the moving scene in the "Trevor White" frame for the assumed block partition.

V. Conclusions

This study analyses the MRBF neural network when used for modeling the optical flow estimation and moving object segmentation. For segmenting the moving scene we employ a mixture of kernel functions whose parameters are found by training. The criterion for segmenting the moving objects is derived from the *a posteriori* probability maximization criterion. Consequently, a cost function is obtained and used as a feature space metric in the learning stage. The cost function takes into account the local motion information, the gray level or color components, the geometrical proximity and considers the displaced frame difference as well. For estimating the hidden unit parameters, a robust unsupervised training algorithm is employed. The hidden units are fed into the output units, each of them associated to a moving object. The moving region areas found in the first learning stage are merged, based on a compactness measure, forming moving objects. The optical flow provided by the proposed algorithm proved to be accurate and smooth.

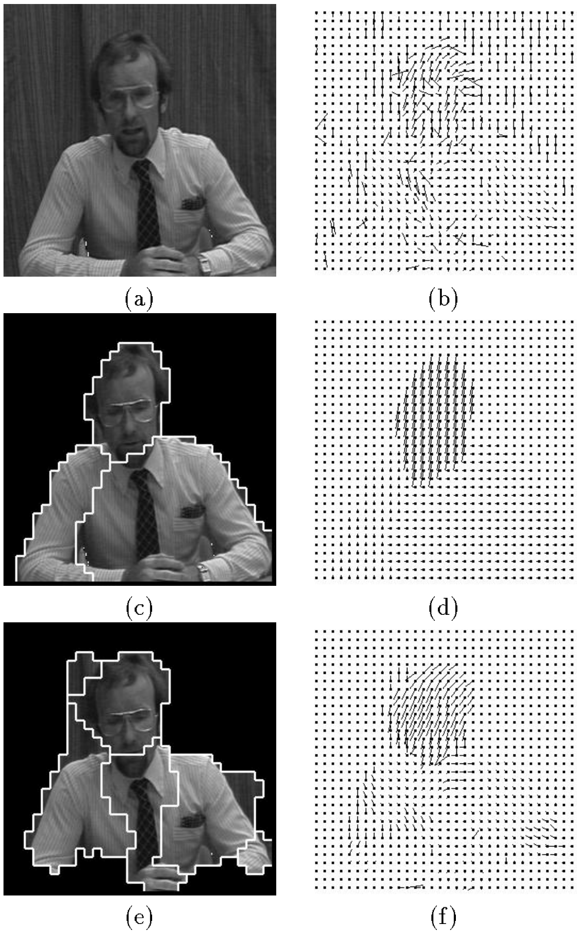

Fig. 2. (a) A frame from the "Trevor White" sequence; (b) The optical flow produced by the full search block matching algorithm; (c) The moving object segmentation provided by the MRBF; (d) The optical flow smoothed by the MRBF; (e) The moving object segmentation provided by the ICM; (f) The optical flow smoothed by the ICM.

References

[1] A. M. Tekalp, *Digital Video Processing.* Upper Saddle River, NJ: Prentice Hall, 1995.

[2] J. Konrad, E. Dubois, "Bayesian estimation of motion vector fields" *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 14, no. 9, pp. 910-927, Sep. 1992.

[3] A. G. Borş, I. Pitas, "Median radial basis function neural network," *IEEE Trans. on Neural Networks,* vol. 7, no. 6, pp. 1351-1364, Nov. 1996.

[4] I. M. Elfadel, R. W. Picard, "Gibbs random fields, cooccurrences, and texture modeling," *IEEE Trans. on Pattern Anal. and Machine Intel.,* vol. 16, no. 1, pp. 24-37, Jan. 1994.

[5] J. Marroquin, S. Mitter, T. Poggio, "Probabilistic solution of ill-posed problems in computational vision," *Jour. Amer. Stat. Assoc.,* vol. 82, no. 397, pp. 76-89, 1987.

[6] A. G. Borş, I. Pitas, "Moving object recognition based on radial basis functions networks," *Proc. of Workshop on Image and Multidimensional Signal Proc.,* Belize City, Belize, pp. 34-35, 1996.

[7] M. M. Chang, M. I. Sezan, A. M. Tekalp, "An algorithm for simultaneous motion estimation and scene segmentation," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing,* Adelaide, Australia, pp. V-221 - V-224, 1994.

[8] F. Heitz, P. Bouthemy, "Multimodal estimation of discontinuous optical flow using Markov Random Fields," *IEEE Trans. on Pattern Anal. and Machine Intel.,* vol. 15, no. 12, pp. 1217-1232, Dec. 1993.