Optical Flow Estimation and Moving Object Segmentation Based on Median Radial Basis Function Network

Adrian G. Borş and Ioannis Pitas

Department of Informatics, University of Thessaloniki, Thessaloniki, Greece

Abstract

Various approaches have been proposed for simultaneous optical flow estimation and segmentation in image sequences. In this study, the moving scene is decomposed into 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. Each class corresponds to a moving object. The classifier employed is the Median Radial Basis Function (MRBF) neural network. An error criterion function derived from the probability estimation theory and expressed as a function of the moving scene model is used as the cost function. Each basis function is activated by a certain image region. Marginal median and median of the absolute deviations (MAD) estimators are employed for estimating the basis function parameters. The image regions associated with the basis functions are merged by the output units in order to identify moving objects.

1 Introduction

Motion representation and modeling are important steps towards the understanding of dynamic images. From the point of view of image analysis, the algorithms used for extracting the optical flow can be classified as gradient-based and feature-based. With respect to the method employed, they can be either deterministic or stochastic, local or global. Gradient-based techniques rely on the spatio-temporal differential equation describing the motion [1]. The methods based on feature matching detect the representative features from the image sequence and try to match them [2]. Block matching algorithms, widely used in video coding [3], estimate the motion field based on the correlation between each block from one frame and a block from the reference frame.

Scene segmentation formulation from optical flow, based on *a posteriori* criterion, was introduced in [4]. A more general framework for motion and line field estimation based on Markov Random Fields (MRF) and Gibbs distributions [5] was used in [6]. Simulated annealing was employed in these approaches for estimating the optical flow and line field.

Iterated Conditional Modes (ICM) was proposed for the restoration of color images in [8] and was applied to motion field smoothing in [9, 10]. ICM is a deterministic local optimization approach used to minimize an energy function by employing constraints and can be viewed as a decision relaxation labeling algorithm [7]. This method was extended to take into account the moving region borders, which should not be smeared by the smoother [11, 12]. ICM relies on a piecewise MRF-based model for the optical flow, associating alternatively various motion vectors and segmentation labels with a given block site. The motion vector and the segmentation label are chosen in each iteration so that the displaced frame difference decreases and a smoothness constraint is fulfilled at the same time. An averaging operation is performed inside a given neighborhood, only in the regions considered to make up a moving object. ICM was modified to include the edge geometry and a gradient-based equation [13]. ICM performance depends on good initialization [12, 13]. The k-means clustering algorithm was employed for this purpose in [15].

Estimation of the optical flow based on clustering analysis was used in [16, 17]. Image intensity information was incorporated in the clustering analysis algorithm in [18]. A robust regression technique where still image segmentation and motion estimation are performed separately was analysed in [19]. Algorithms that perform simultaneous motion estimation and scene segmentation, in an iterative way, based on the minimization of an energy function, were provided in [12, 20]. The approach suggested in [12] is based on local optimization methods like ICM and Highest Confidence First and provides improvements over the segmentation on the basis of a separate estimation. A gradient-based algorithm derived from the Mean Field Theory using global optimization was introduced in [20]. Some of these algorithms have been embedded in multiresolution structures [6, 12, 13, 20].

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 moving objects. The image is 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 and representing a distance in the parameter space [21]. The moving scene is split accordingly in moving regions. We consider the MRBF network for modeling the optical flow and moving object segmentation in the image sequence. This structure is embedded in a two-layer feedforward neural network, where each output is assigned to a moving object.

A radial basis function (RBF) decomposition is known to be a good functional approximator and has been used in many applications [22]-[27]. The first-layer units implement Gaussian functions. The classification criterion connects the Gaussian parameters to the set of feature vectors drawn from the image sequence. In the second layer, the moving regions associated to the basis functions are merged in order to model the moving objects. The mixture of basis functions approximates the probabilities associated to the optical flow estimation and segmentation of the moving objects.

The learning algorithm employed for estimating the set of parameters associated to the image sequence has two stages [26, 27]. In the first stage, the basis function center and variance are found based on a clustering approach, similar to the Learning Vector Quantization (LVQ) [28]. A robust statistics-based algorithm employing the marginal median and median of absolute deviations (MAD) [29, 30] and called MRBF was proposed for training the network in [27]. The number of moving objects does not need to be specified *a priori*. It is found according to a compactness measure. The backpropagation algorithm, applied for calculating the output weights of the RBF network [22, 31], is employed in the second learning stage.

In Section 2 the derivation of the classification criterion is presented. Section 3 describes the learning algorithm. In Section 4 simulation results showing the effectiveness of the proposed method are provided. The conclusions of the present study are drawn in Section 5.

2 The classification criterion

2.1 Feature extraction

Let us consider an image sequence f_t. Each frame is partitioned in block sites B_{IJ}, for $I = 0, \ldots, n_x - 1$ and $J = 0, \ldots, n_y - 1$, situated on a rectangular grid. We associate a feature vector denoted as $\mathbf{u}_{IJ}$, describing the local image sequence properties, to each block site B_{IJ}. 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 :

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

where the mean gray level is given by:

$$\hat{l}_{IJ} = \frac{1}{D_x D_y} \sum_{h=-\frac{D_y}{2}}^{\frac{D_y}{2}} \sum_{g=-\frac{D_x}{2}}^{\frac{D_x}{2}} f(I+h, J+g) \tag{3}$$

$f(I+h, J+g)$ denotes the gray level values of the image elements and $D_x \times D_y$ is the block size. In the case of color image sequences, the color components are included in the feature vector $\mathbf{S}_{IJ}$ as well. The motion vector contains two components, corresponding to the local motion of the block :

$$\mathbf{M}_{IJ} = [m_{x,IJ}, m_{y,IJ}]. \tag{4}$$

They can be calculated by any optical flow algorithm, e.g., by block matching. Block matching methods use the correlation of a given image block of size $D_x \times D_y$ from the frame f_{t-1} with blocks of the same size, situated inside a search region $S_x \times S_y$, on the next frame f_t [3]. The motion vector that minimizes the displaced frame difference (DFD) is chosen. The DFD is given by :

$$d_{IJ}(\hat{\mathbf{M}}_{IJ}) = \sum_{k=-\frac{D_x}{2}}^{\frac{D_x}{2}} \sum_{l=-\frac{D_y}{2}}^{\frac{D_y}{2}} |f_t(I+k, J+l) - f_{t-1}(I+k+\hat{m}_{x,IJ}, J+l+\hat{m}_{y,IJ})|, \tag{5}$$

$\hat{m}_{x,IJ}$, $\hat{m}_{y,IJ}$ are the motion component estimates.

2.2 Moving object classification

A moving scene can be seen as made up of regions having 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 a class. The partition of the moving objects in moving regions depends on the type of motion they employ and their local properties.

Each subset X_k is associated to a five-dimensional representative vector μ_k, describing the optical flow and segmentation information associated with a certain moving region :

$$\mu_k = [\mathcal{S}_k, \mathcal{M}_k]. \tag{6}$$

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

Let us denote by $\hat{\mathcal{S}}_k$ and $\hat{\mathcal{M}}_k$ the estimates of the segmentation label and optical flow associated to the moving region k. 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 $\hat{\mathcal{S}}_k$ 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 :

$$P(\hat{\mathcal{M}}_k, \hat{\mathcal{S}}_k | f_{t-1}, f_t) > P(\hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j | f_{t-1}, f_t) \tag{7}$$

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

In the second stage, the moving regions are merged, in order to describe moving objects, based on a neighboring criterion. Let us denote by $\hat{\mathcal{T}}_k$ the estimate of the optical flow and of the segmentation 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 :

$$V(X_l, X_k) = V(X_k, X_l) = \sum_{B_{IJ} \in X_l} |\mathcal{N}_{IJ} \cap X_k| \tag{8}$$

where $|\mathcal{N}_{IJ} \cap X_k|$ represents the number of block sites of a moving region X_k, which are situated in a certain neighborhood $\mathcal{N}_{IJ}$ of the block site B_{IJ}, belonging to the moving region X_l [32]. This measure represents 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 object as a moving region which 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 :

$$\text{If } V(X_k, X_k) = \max_{i=1}^{L} V(X_k, X_i) \text{ then } P(\hat{\mathcal{T}}_k | f_{t-1}, f_t) = P(\hat{\mathcal{M}}_k, \hat{\mathcal{S}}_k | f_{t-1}, f_t). \tag{9}$$

A moving object k contains a moving region X_l :

$$\text{If } V(X_l, X_k) = \max_{i=1}^{L} V(X_l, X_i) \text{ then } P(\hat{\mathcal{T}}_k | f_{t-1}, f_t) = \lambda_k P(\hat{\mathcal{M}}_k, \hat{\mathcal{S}}_k | f_{t-1}, f_t) + \lambda_l P(\hat{\mathcal{M}}_l, \hat{\mathcal{S}}_l | f_{t-1}, f_t) \tag{10}$$

where λ_k, λ_l are the weights denoting 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 (7) with respect to the features extracted from the image sequence. After applying Bayes' rule, each of the *a posteriori* distributions in (7) can be factored as follows :

$$P(\hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j | f_t, f_{t-1}) = \frac{P(f_t | f_{t-1}, \hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j) P(\hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j | f_{t-1})}{P(f_t | f_{t-1})} = \frac{P(f_t | f_{t-1}, \hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j) P(\hat{\mathcal{M}}_j | \hat{\mathcal{S}}_j, f_{t-1}) P(\hat{\mathcal{S}}_j | f_{t-1})}{P(f_t | f_{t-1})}. \tag{11}$$

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 map and image [6]. The probability $P(f_t|f_{t-1})$ does not depend on $[\hat{\mathcal{S}}_j, \hat{\mathcal{M}}_j]$ and can be neglected.

Each of the above conditional probabilities can be expressed as an energy function $E(\mathbf{X})$:

$$P(\mathbf{X}) = \frac{1}{Z} \exp\left[-\frac{E(\mathbf{X})}{\beta}\right] \tag{12}$$

where Z is a normalizing constant and β is a constant controlling the properties of $E(\mathbf{X})$ [5]. Thus, the probability estimation problem (7) is converted into the minimization of an energy function, derived from (11) and (12) :

$$E_j = E(f_t|f_{t-1}, \hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j) + E(\hat{\mathcal{M}}_j|\hat{\mathcal{S}}_j, f_{t-1}) + E(\hat{\mathcal{S}}_j|f_{t-1}). \tag{13}$$

The first component of the energy function is related to the estimation of the frame f_t based on the previous frame f_{t-1}, the optical flow and the segmentation. The second component represents the energy associated to the estimated optical flow and the third the energy associated to the estimated segmentation. In order to minimize E_j, all three components should be simultaneously minimized. This corresponds to the simultaneous optical flow and frame segmentation.

2.3 The cost function

The performance criterion is related to the total squared error minimization in the feature space [33]. The energy function in (13) is expressed as a clustering metric in the feature space. This metric relates the moving region feature vectors (6) to the block site feature vectors (1).

The energy $E(f_t|f_{t-1}, \hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j)$ in (13) is represented as a weighted function of the displaced frame difference corresponding to the moving region j and denoted as $WDFD(\hat{\mathcal{M}}_j)$:

$$E(f_t|f_{t-1}, \hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j) = WDFD(\hat{\mathcal{M}}_j) = \sum_{\substack{I=0 \\ B_{IJ} \in X_j}}^{n_x-1} \sum_{J=0}^{n_y-1} [w_{IJ}(\hat{\mathcal{M}}_j) d_{IJ}(\hat{\mathcal{M}}_j)]^2, \tag{14}$$

where $d_{IJ}(\hat{\mathcal{M}}_j)$ is the DFD estimate (5) 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 the weighting factor a reliability coefficient in the output of the block matching :

$$w_{IJ}(\hat{\mathcal{M}}_j) = \frac{d_{IJ}(\hat{\mathcal{M}}_j)}{\sum_{k=-\frac{S_x}{2}}^{\frac{S_x}{2}} \sum_{l=-\frac{S_y}{2}}^{\frac{S_y}{2}} d_{I+k,J+l}(\hat{\mathcal{M}}_j)}, \tag{15}$$

where $S_x \times S_y$ is the search region for the block matching algorithm. This coefficient is small when we have good matching and large in the case of poor matching. For regions with constant gray level in the search area, this coefficient becomes :

$$w_{IJ}(\hat{\mathcal{M}}_j) = \frac{D_x\ D_y}{S_x\ S_y}. \tag{16}$$

The energy function $E(\hat{\mathcal{M}}_j|\hat{\mathcal{S}}_j, f_{t-1})$ in (13) is associated with motion vector clustering :

$$E(\hat{\mathcal{M}}_j|\hat{\mathcal{S}}_j, f_{t-1}) = \sum_{I=0}^{n_x-1} \sum_{\substack{J=0 \\ B_{IJ}\in X_j}}^{n_y-1} (\mathbf{M}_{IJ} - \hat{\mathcal{M}}_j)^T (\mathbf{M}_{IJ} - \hat{\mathcal{M}}_j), \tag{17}$$

where $\hat{\mathcal{M}}_j$ is the motion vector estimate of the moving region j and $\mathbf{M}_{IJ}$ is the motion vector associated with a block site (4) belonging to the region j.

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 :

$$E(\hat{\mathcal{S}}_j|f_{t-1}) = \sum_{I=0}^{n_x-1} \sum_{\substack{J=0 \\ B_{IJ}\in X_j}}^{n_y-1} (\mathbf{S}_{IJ} - \hat{\mathcal{S}}_j)^T (\mathbf{S}_{IJ} - \hat{\mathcal{S}}_j), \tag{18}$$

where $\hat{\mathcal{S}}_j$ and $\mathbf{S}_{IJ}$ are still image feature vectors of the moving region j and its component block sites (2), respectively. This cost function component entails that block sites assigned to a certain moving region are situated in the same image area and have similar average gray levels.

By replacing (14), (17) and (18) in (13) we obtain the jth moving object energy :

$$E_j(\mathbf{u}_{IJ}) = \sum_{I=0}^{n_x-1} \sum_{\substack{J=0 \\ B_{IJ}\in X_j}}^{n_y-1} (\mathbf{u}_{IJ} - \hat{\mu}_j)^T (\mathbf{u}_{IJ} - \hat{\mu}_j) + WDFD(\hat{\mathcal{M}}_j), \tag{19}$$

where $\hat{\mu}_j$ is the jth moving region feature estimate (6), $\mathbf{u}_{IJ}$ is a block site feature vector (1) and $WDFD(\hat{\mathcal{M}}_j)$ is provided in (14). The first part of this expression is associated to clustering in the feature space.

3 Median Radial Basis Function learning algorithm

3.1 The estimation of the hidden unit parameters

The cost function (19) corresponds to image partition in moving regions. If we take into account the covariance matrix and we express (19) as an unnormalized probability (12), we obtain a so called radial basis function :

$$\phi_j(\mathbf{u}_{IJ}) = \exp\left[-(\mathbf{u}_{IJ} - \hat{\mu}_j)^T \hat{\mathbf{\Sigma}}_j^{-1} (\mathbf{u}_{IJ} - \hat{\mu}_j) - WDFD(\hat{\mathcal{M}}_j)\right], \tag{20}$$

where $\hat{\mu}_j$ is the estimated center vector and $\hat{\mathbf{\Sigma}}_j$ is the estimate of 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 RBF network consists of a two-layer topology where each hidden unit implements a radial-activated region function (20).

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

$$Y_k(\mathbf{u}_{IJ}) = \frac{1}{1 + \exp\left[-\sum_{j=1}^{L} \lambda_{kj} \phi_j(\mathbf{u}_{IJ})\right]} \tag{21}$$

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 (20). The *a posteriori* probability of optical flow estimation and segmentation associated to each moving object, is modeled by the output unit:

$$P(\hat{\mathcal{T}}_k | f_t, f_{t-1}) = \sum_{I=0}^{n_x-1} \sum_{J=0}^{n_y-1} \sum_{j=1}^{L} \lambda_{kj} \phi_j(\mathbf{u}_{IJ}). \tag{22}$$

RBF networks have good functional approximation capabilities and they have been employed in many applications [22]-[27]. The diagram of the proposed system, as applied to moving object segmentation, is shown in Figure 1. The RBF network entries correspond to still image features (2) and block matching motion estimates (4). Each basis function (20) consists of a Gaussian function and a constraint representing the weighted displaced frame difference (14). In the context of regularization theory, the RBF networks were shown to approximate smooth functions arbitrary well [25].

The network parameters are estimated based on a learning algorithm. A well-known RBF training algorithm [26] employs in the first training stage the LVQ algorithm [28]. For a given feature vector $\mathbf{u}_{IJ}$, we update only the parameters of that basis function which provides a minimal energy in (19) :

$$\text{If } \min_{k=1}^{L} E_k(\mathbf{u}_{IJ}) = E_j(\mathbf{u}_{IJ}) \text{ then update } \hat{\mu}_j, \hat{\mathbf{\Sigma}}_j. \tag{23}$$

A robust statistics-based training algorithm called Median Radial Basis Function (MRBF) is proposed in [27] for estimating the RBF network parameters. According to the results provided by the theoretical analysis, the algorithms based on robust statistics give a better accuracy in estimating the parameters of overlapping Gaussian mixtures [27]. The basis function center is randomly initialized in the input space range. Its updating is based on the marginal median LVQ algorithm [34]. This algorithm orders the data samples, on each dimension separately, and takes their median :

$$\hat{\mu}_k = \text{med}\,\{\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_{p-1}\} \tag{24}$$

where $\mathbf{u}_{p-1}$ is the last data sample assigned to that basis function. The median of the absolute deviations (MAD) estimator is used as a robust dispersion estimator :

$$\hat{\mathbf{r}}_k = \frac{\text{med}\,\{|\mathbf{u}_0 - \hat{\mu}_k|, \ldots, |\mathbf{u}_{p-1} - \hat{\mu}_k|\}}{0.6745} \tag{25}$$

where $\hat{\mathbf{r}}_k$ denotes the diagonal vector of the covariance matrix $\hat{\mathbf{\Sigma}}_k$ (20) and 0.6745 is the scale parameter chosen in order to make the estimator Fisher consistent for the normal distribution [29]. The model parameter estimation based on robust algorithms like marginal median and MAD enables the rejection of a certain amount of data not belonging to the given model. A fast implementation algorithm for (24, 25), based on data sample histogram modeling, is provided in [27].

3.2 The estimation of the output layer parameter

By estimating the component radial basis functions, the image is split in moving regions. A given block site is assigned to that moving region which corresponds to the most activated radial basis function :

$$\text{If } \phi_k(\mathbf{u}_{IJ}) = \max_{j=1}^{L} \phi_j(\mathbf{u}_{IJ}) \text{ then } B_{IJ} \in X_k. \tag{26}$$

If an image region X_k has not been assigned sufficient data samples after the first learning stage, according to (26), the respective hidden unit is discarded. Evidently, that hidden unit does not contribute significantly to the optical flow estimation and segmentation probability. Afterwards, the training is reiterated with a smaller number of hidden units. The number of hidden units has to be increased in the case when new moving regions appear in scene. For each two moving regions, we evaluate their boundary measure (8), assuming a four-block site neighborhood system. The moving regions which have a high interconnectivity among their component block sites, are considered as moving objects (9), and the other regions are merged with the existing moving objects according to (10), (22). An output unit is assigned to each moving object. The block sites assigned to the moving regions are labeled according to the decision (26) and (9) or (10) :

$$\text{If } B_{IJ} \in X_k \text{ then } F_k(\mathbf{u}_{IJ}) = 1,\ F_l(\mathbf{u}_{IJ}) = 0 \text{ for } l = 1, \ldots, N,\ l \neq k. \tag{27}$$

The second learning stage estimates the parameters λ_{kj} associated with the connections between hidden and output layers (22). In order to find the weights λ_{kj} we use a gradient based learning algorithm [22] :

$$\lambda_{kj} = \sum_{I=1}^{n_x} \sum_{J=1}^{n_y} [F_k(\mathbf{u}_{IJ}) - Y_k(\mathbf{u}_{IJ})] Y_k(\mathbf{u}_{IJ})[1 - Y_k(\mathbf{u}_{IJ})]\phi_j(\mathbf{u}_{IJ}) \tag{28}$$

where $Y_k(\mathbf{u}_{IJ})$ are the outputs of the network (21) and $F_k(\mathbf{u}_{IJ})$ are the target labels assigned to the block sites according to (27).

After the training stage is completed, when entering a feature vector (1), the maximum activated output unit will show the corresponding moving object. This procedure leads to the division of the image sequence into moving objects. The trained network can be applied in other frames of the same image sequence, if their optical flow estimation and segmentation probabilities are consistent with those of the frames used in the training stage. The consistency is given by the optical flow and moving object similarity measures and can be expressed as the energy function from (13).

The MRBF network records in its weights the motion, luminance and localization of the moving objects. The network can be used in a multiresolution (hierarchical) representation of the image. Let us consider that the MRBF network was trained on a certain image partition and afterwards we input feature vectors from a different block size image partition. If the block size is large, then the segmentation will provide rough boundaries but the training time will be short. The network can be trained with features corresponding to big blocks and, afterwards be applied on an image partition in blocks of smaller size. The location features must be scaled according to the ratio between the block sizes used in training and testing. In order to save time, the network can be applied at a higher resolution only for the block sites situated at the moving object boundaries.

4 Simulation results

The MRBF network was tested on various image sequences. In this study we provide two representative results. In Figure 2 a frame from the "Hamburg Taxi" sequence is shown. This sequence contains three important moving objects: 1) a taxi near the center turning around the corner, 2) a car in the lower left, driving from left to right, 3) a van in the lower right, driving from right to left. The optical flow provided by the full search block matching algorithm [3], is shown in Figure 3. The feature vectors are drawn from the first and third frames of the "Hamburg Taxi" sequence. The learning algorithm described in Section 3 is applied to the given data and the optical flow smoothed by the MRBF network is provided in Figure 4. The moving objects segmented by means of the MRBF network are shown in Figure 5.

The optical flow and moving object segmentation provided by ICM [8]-[15] for "Hamburg Taxi" sequence are shown in Figures 6 and 7, respectively. We have used the full search block matching and k-means clustering algorithm for ICM initialization [15]. Each block site has associated a label and a motion vector which are iteratively modified based on an optimization method, employing local constraints. This algorithm takes into account the borders of the moving objects, but smoothes the optical flow only locally, inside a given neighborhood. Certain discontinuities can be observed in the optical flow smoothed by the ICM algorithm as shown in Figure 6.

The MRBF network processes entire image regions assigned to the same moving object and exploits the interdependency among their block sites. Starting with various numbers of hidden units, the network always detected four moving objects in the "Hamburg Taxi" sequence and provided accurate optical flows. The moving object segmentation is quite accurate despite the poor quality of the object edges. The optimal optical flow and moving object segmentation, evaluated semiautomatically by averaging the displacements of various clear features and manually segmenting the image, are shown in Figures 8 and 9. The results provided by the MRBF network are better than those given by the ICM and they are quite close to the optimal results.

The MRBF network trained on a 4×4 pixel block partition, whose results are presented in Figures 4 and 5, is applied on a pixel-level partition of the same frame, in a hierarchical approach. The segmentation of the moving objects is provided in Figure 10. The "taxi" and "van" moving objects are better segmented, as can be seen from Figures 5 and 10. The segmentation results obtained when applying the MRBF network trained using the first and third frames, on the 8th frame are shown in Figure 11. This result shows the network's capability to embed in its weights the parameters associated with the moving objects.

A frame from "Trevor White" image sequence and its corresponding optical flow provided by the block matching algorithm are shown in Figures 12 and 13. The optical flow smoothed by the MRBF network and by ICM are displayed in Figures 14 and 15. The movement segmentation when using the MRBF network and ICM algorithm is illustrated in Figures 16, 17. These results show that MRBF network provides a good moving object segmentation and smooth optical flow even in the case when the movement is complex as it is with the "Trevor White" sequence.

Numerical comparisons when applying the MRBF network and ICM algorithm in "Hamburg Taxi" sequence are provided in Table 1. The comparison criteria when using the training frames are the optical flow mean absolute error (MAE), mean squared error (MSE), the misclassification error, the number of parameters required by each algorithm in order to estimate the optical flow and segment the motion, and the necessary number of iterations in order to achieve the convergence. The optical flow MAE and MSE are :

$$\text{MAE} = \frac{1}{n_x n_y} \sum_{I=0}^{n_x-1} \sum_{J=0}^{n_y-1} (|\hat{m}_{x,IJ} - m_{x,IJ}| + |\hat{m}_{y,IJ} - m_{y,IJ}|) \tag{29}$$

$$\text{MSE} = \frac{1}{n_x n_y} \sum_{I=0}^{n_x-1} \sum_{J=0}^{n_y-1} (\hat{\mathbf{M}}_{IJ} - \mathbf{M}_{IJ})^T (\hat{\mathbf{M}}_{IJ} - \mathbf{M}_{IJ}) \tag{30}$$

where $\hat{\mathbf{M}}_{IJ}$ is the block motion estimate. The misclassification error represents the percentage of erroneous decisions when assigning each block site to a moving object. The optical flow used as reference $\mathbf{M}_{IJ}$, and the optimal segmentation are provided in Figures 8 and 9. 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 total number of parameters required by the ICM algorithm is $3n_x n_y$. The MRBF network trained using data samples drawn from the first and third frames has been applied on the frames 2 to 9 of the "Hamburg Taxi" sequence. The misclassification error and the total processing time corresponding to the nine frames (including the feature computation and the training time for the MRBF) evaluated on a Silicon Graphics Indy Workstation are provided in Table 1. The time spent in the MRBF training stage, for this example, is 27.9 seconds.

It can be seen from Table 1, as well as from Figures 4-17, that the MRBF network outperforms the ICM algorithm. The average time per frame for the MRBF network is larger when compared with the ICM algorithm but MRBF network requires less iterations in training. The reduction in the number of parameters, provided by the MRBF network is very high, when compared to the ICM.

5 Conclusions

The MRBF neural network when used for optical flow estimation and moving object segmentation is analysed in this study. The considered model consists of 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. The weights of the network are updated by means of an unsupervised learning algorithm. A robust training algorithm is employed for estimating the hidden unit parameters. The mixture of hidden units is 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. After the training stage, the network records in its weights the moving object information. The number of parameters required by the MRBF network to model the optical flow and moving object segmentation is small. The information contained in the network parameters can be further used for image sequence analysis and region-based coding.

References

[1] J. K. Kearney, W. B. Thompson, D. L. Boley, "Optical flow estimation: an error analysis of gradient based methods with local optimization," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 9, No. 2, pp. 229-244, Mar. 1987.

[2] J. Weng, N. Ahuja, T. S. Huang, "Matching two perspective views," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 14, No. 8, pp. 806-825, Aug. 1992.

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

[4] D. W. Murray, B. F. Buxton, "Scene segmentation from visual motion using global optimization," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 9, No. 2, pp. 220-228, Mar. 1987.

[5] S. Geman, D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, " *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 6, No. 6,pp. 721-741, Nov. 1984.

[6] 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.

[7] H. M. Kalayeh, G. Horgan, "Adaptive relaxation labeling," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 6, pp. 369-372, 1984.

[8] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy Statist. Soc.,* vol. 48, ser. B, No. 3, pp. 259-302, 1986.

[9] T. Aach, A. Kaup, R. Mester, "Combined displacement estimation and segmentation of stereo image pairs based on Gibbs random fields," *Proc. IEEE Int. Conf. Acoust. Speech and Signal Proc.,* Albuquerque, NM, pp. 2301-2304, Apr. 1990.

[10] M. M. Chang, A. M. Tekalp, M. I. Sezan, "Motion-field segmentation using an adaptive MAP criterion," *Proc. IEEE Int. Conf. Acoust. Speech and Signal Proc.,* Minneapolis, MN, pp. V-33 - V-36, Apr. 1993.

[11] C. Stiller, "Motion-Estimation for coding of moving video at 8 kbit/s with Gibbs modeled vector field smoothing," *Proc. SPIE Visual Communications and Image Processing,* vol. 1360, pp. 468-476, 1990.

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

[13] 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.

[14] C. Stiller, "Object-oriented coding employing dense motion fields," *Proc. IEEE Int. Conf. Acoust. Speech and Signal Proc.,* Adelaide, Australia, pp. V-273 - V-276, May 1994.

[15] T. N. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Trans. Signal Processing,* vol. 40, no. 4, pp. 901-914, Apr. 1992.

[16] B. G. Schunck, "Image flow segmentation and estimation by constraint line clustering," *IEEE Trans. on Pattern Anal. and Machine Intel.,* vol. 11, no. 10, pp. 1010-1027, Oct. 1989.

[17] J. Y. A. Wang, E. H. Adelson, "Layered representation for motion analysis," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition,* New York, NY, pp. 361-366, June 1993.

[18] D. P. Kottke, Y. Sun, "Motion estimation via clustering matching," *IEEE Trans. on Pattern Anal. and Machine Intel.,* vol. 16, no. 11, pp. 1128-1132, Nov. 1994.

[19] B. Duc, P. Schroeter, J. Bigun, " Spatio-Temporal robust motion estimation and segmentation," *Proc. Conf. on Computer Analysis of Images and Patterns,* Prague, Czech Republic, pp. 238-245, 6-8 Sep. 1995.

[20] T. Y. Tian, M. Shah, "Motion segmentation and estimation," *Proc. IEEE Conf. on Image Processing,* vol II, Austin, TX, pp. 785-789, Nov. 1994.

[21] A. G. Borş, I. Pitas, "Segmentation and estimation of the optical flow," *Proc. of Int. Conf. of Computer Analysis of Images and Patterns,* Prague, Czech Rep., pp. 680-685, Sep. 1995.

[22] A. G. Borş, M. Gabbouj, "Minimal topology for a radial basis functions neural network for pattern classification," *Digital Signal Processing: A review journal*, vol. 4, no. 3, pp. 173-188, July 1994.

[23] E. J. Hartman, J. D. Keeler, J. M. Kowalski, "Layered neural networks with Gaussian hidden units as universal approximations," *Neural Computation,* vol. 2, pp. 210-215, 1990

[24] J. Park, J. W. Sandberg, "Universal approximation using radial basis functions network," *Neural Computation,* vol. 3, pp. 246-257, 1991.

[25] T. Poggio, F. Girosi, "Networks for approximation and learning," *Proc. of the IEEE,* vol. 78, no. 9, pp. 1481-1497, Sep. 1990.

[26] L. Xu, A. Krzyzak, E. Oja, "Rival penalized competitive learning for clustering analysis, RBF net, and curve detection," *IEEE Trans. on Neural Networks,* vol. 4, no. 4, pp. 636-649, July 1993.

[27] 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.

[28] T. K. Kohonen, *Self-organization and associative memory.* Berlin: Springer-Verlag, 1989.

[29] G. Seber, *Multivariate Observations.* John Wiley, 1984.

[30] I. Pitas, A. N. Venetsanopoulos, *Nonlinear Digital Filters: principles and applications.* Norwell, MA: Kluwer Academic, 1990.

[31] S. Haykin, *Neural Networks: A Comprehensive Foundation.* NY: Macmillan, 1994.

[32] 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.

[33] 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.

[34] I. Pitas, C. Kotropoulos, N. Nikolaidis, R. Yang, M. Gabbouj, "Order statistics learning vector quantizer," *IEEE Trans. on Image Processing,* vol. 5, no. 6, pp. 1048-1053, June 1996.

Table 1: COMPARISON BETWEEN MRBF NETWORK AND ICM WHEN APPLIED IN THE "HAMBURG TAXI" IMAGE SEQUENCE.

Algorithm	Training Frame					Image Sequence (9 Frames)	
	Classification Error (%)	MAE	MSE	Number of Parameters	Number of Iterations	Classification Error (%)	Total Processing Time (s)
MRBF	3.02	0.17	0.85	210	13	3.55	74.2
ICM	4.07	0.33	1.15	9216	23	8.70	57.4

List of Tables

List of Figures

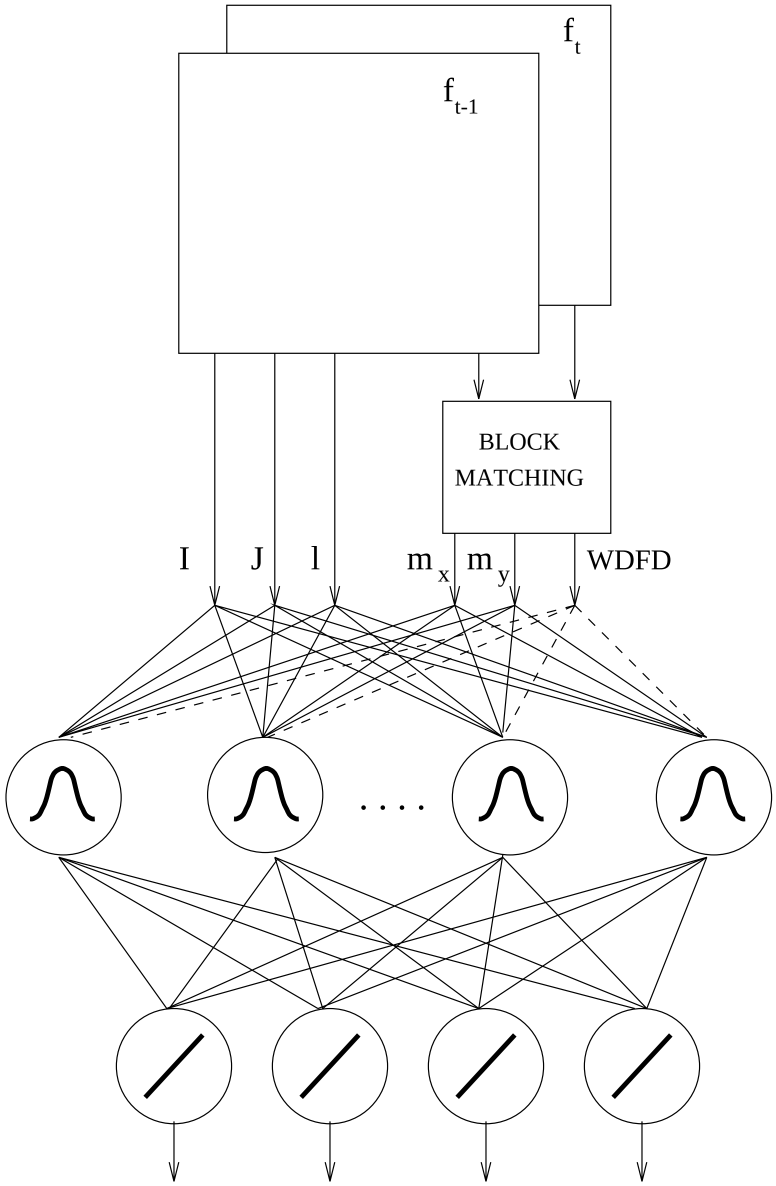

Figure 1: The MRBF structure for optical flow estimation and segmentation: I and J are the location coordinate features assigned to a block site, l is the gray level and m_x, m_y are the motion vector components provided by the block matching algorithm.

Figure 2: Frame from the "Hamburg Taxi" sequence.

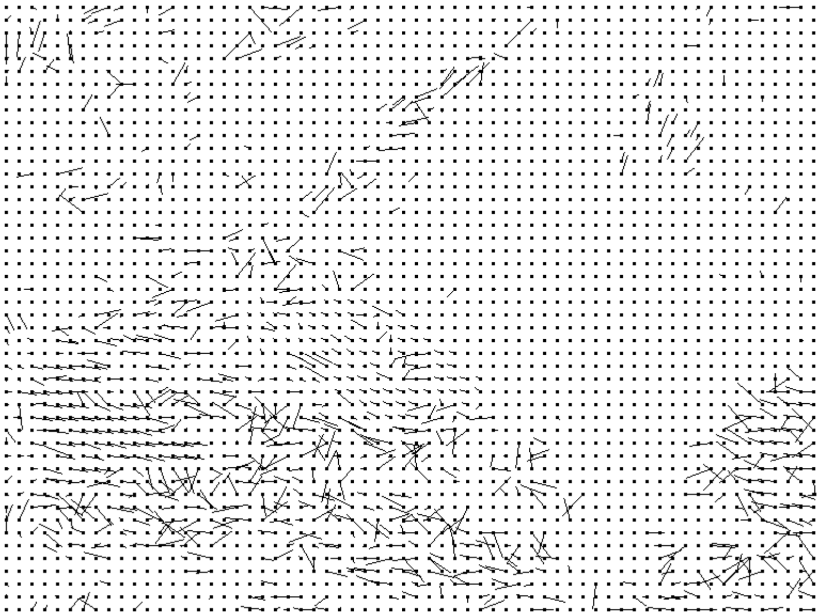

Figure 3: The optical flow produced by the full search block matching algorithm.

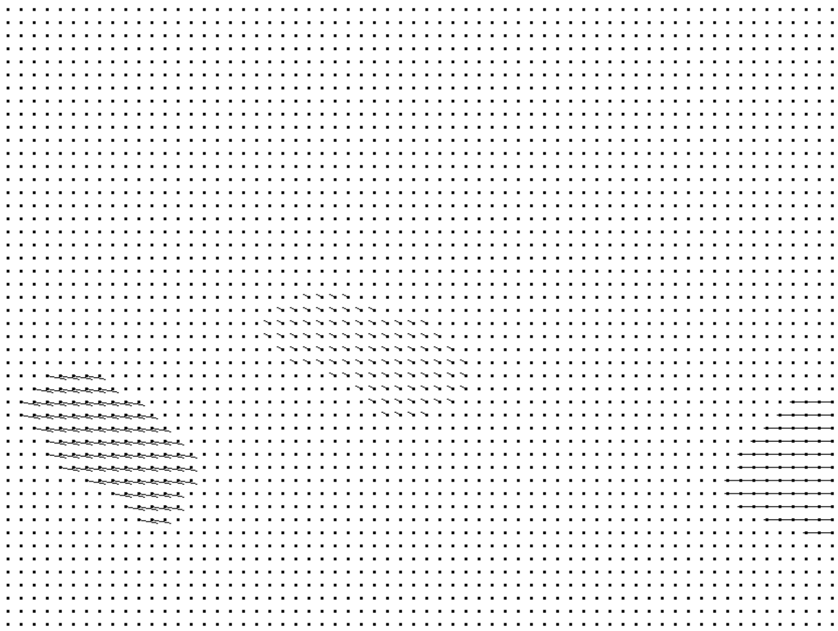

Figure 4: The optical flow smoothed by the MRBF network.

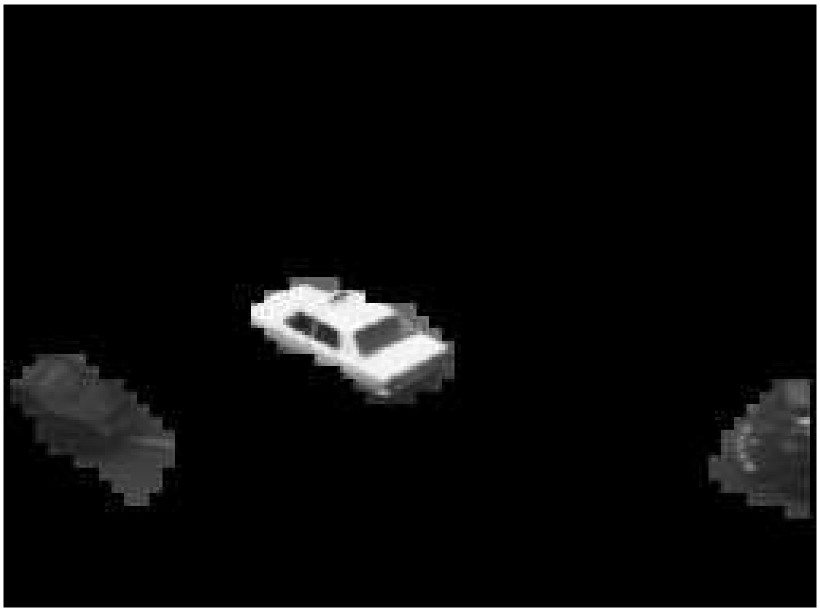

Figure 5: The segmentation of the moving objects based on the MRBF network.

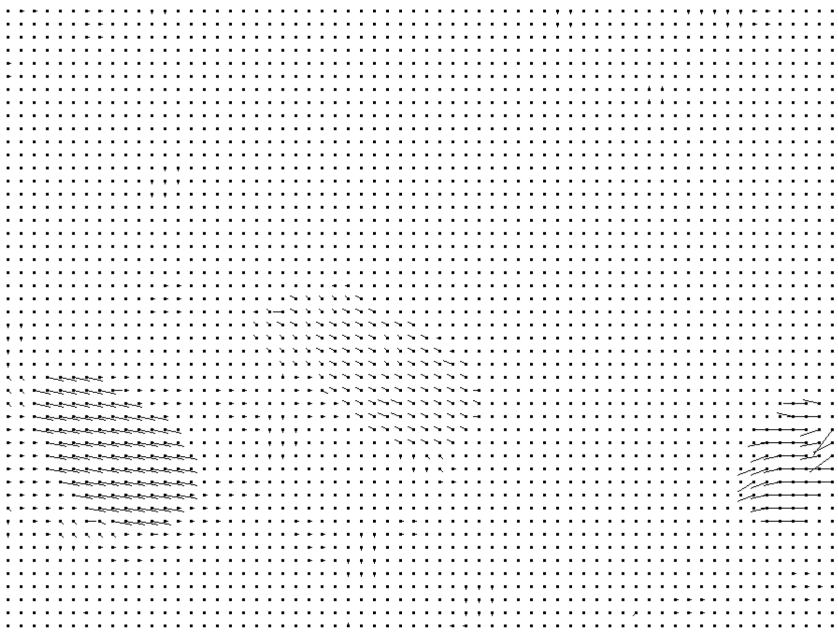

Figure 6: The optical flow smoothed by ICM.

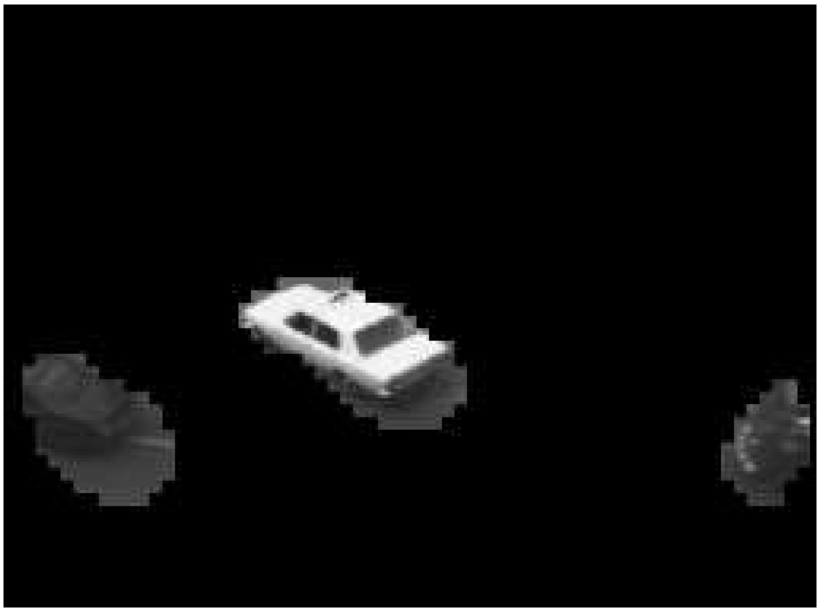

Figure 7: Segmentation of the moving objects based on ICM.

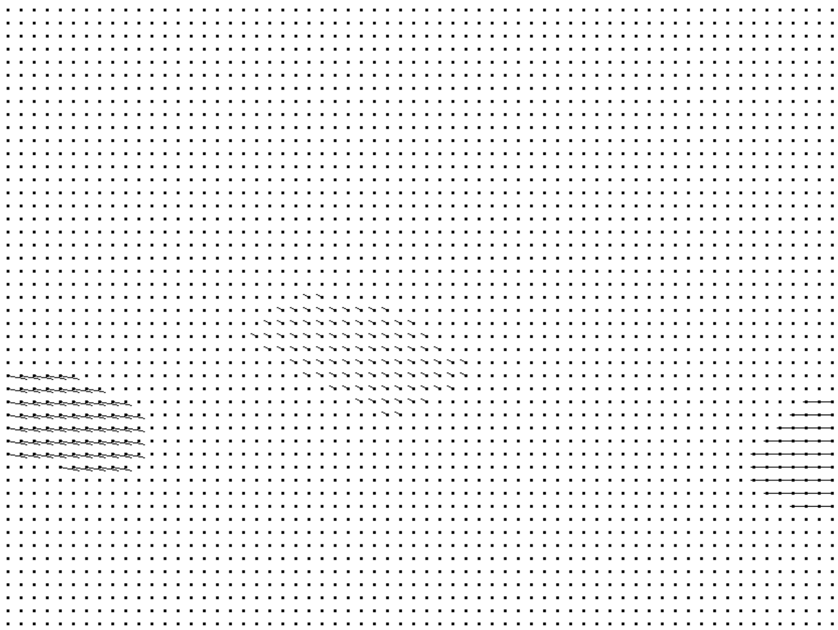

Figure 8: The reference optical flow for the "Hamburg Taxi" sequence.

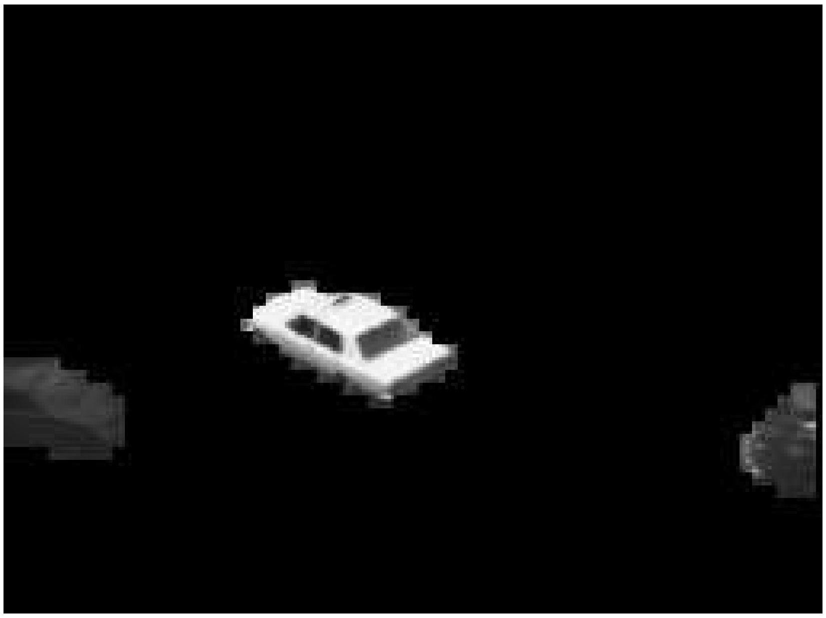

Figure 9: The reference moving object segmentation for the "Hamburg Taxi" sequence.

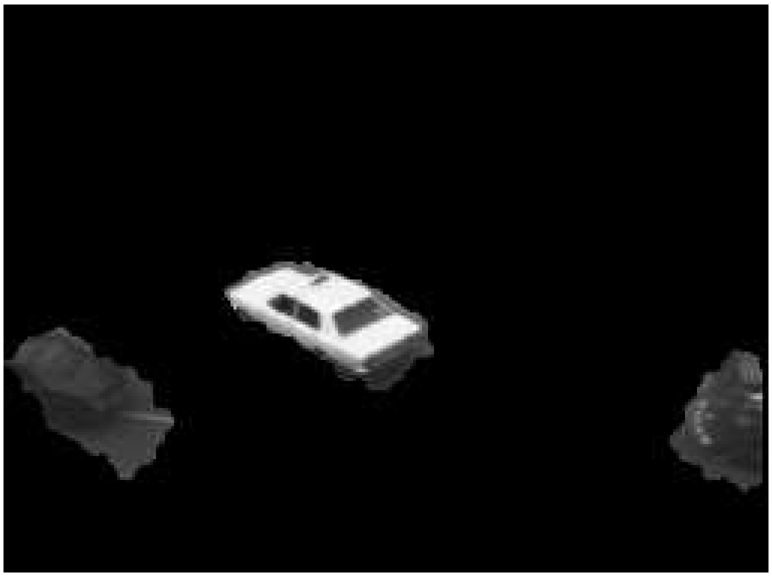

Figure 10: The MRBF network-based segmentation when applied at pixel level partition in the “Hamburg Taxi” sequence.

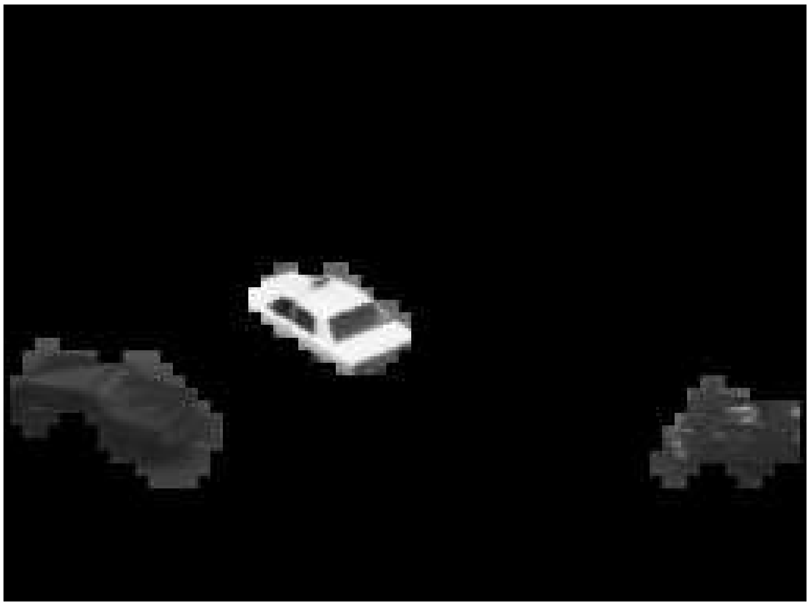

Figure 11: Moving object segmentation in the 8th frame obtained after applying the network trained with data samples drawn from the first and third frames.

Figure 12: Frame from the "Trevor White" image sequence.

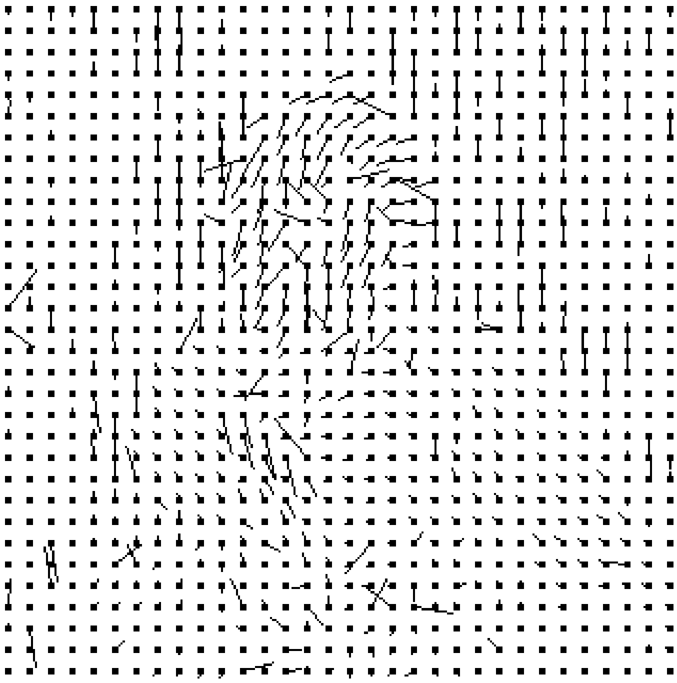

Figure 13: Optical flow in the "Trevor White" image sequence provided by the full search block matching algorithm.

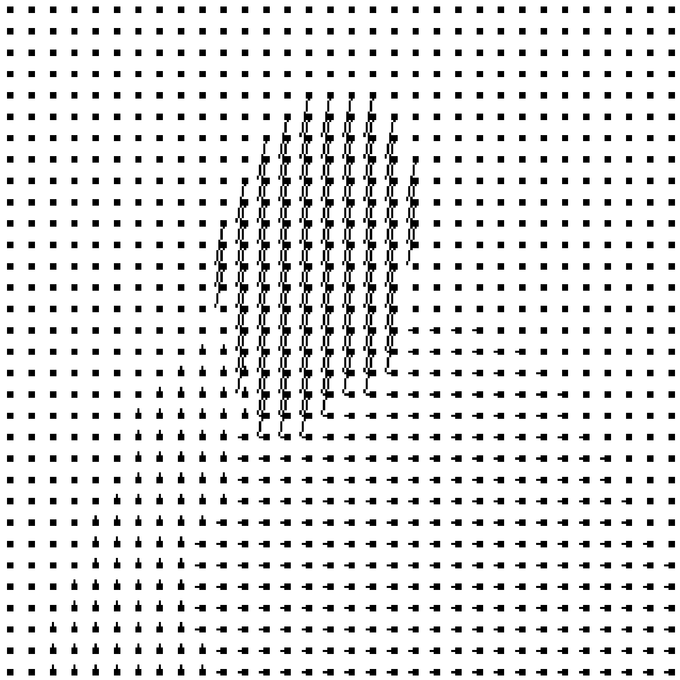

Figure 14: Optical flow smoothed by the MRBF network.

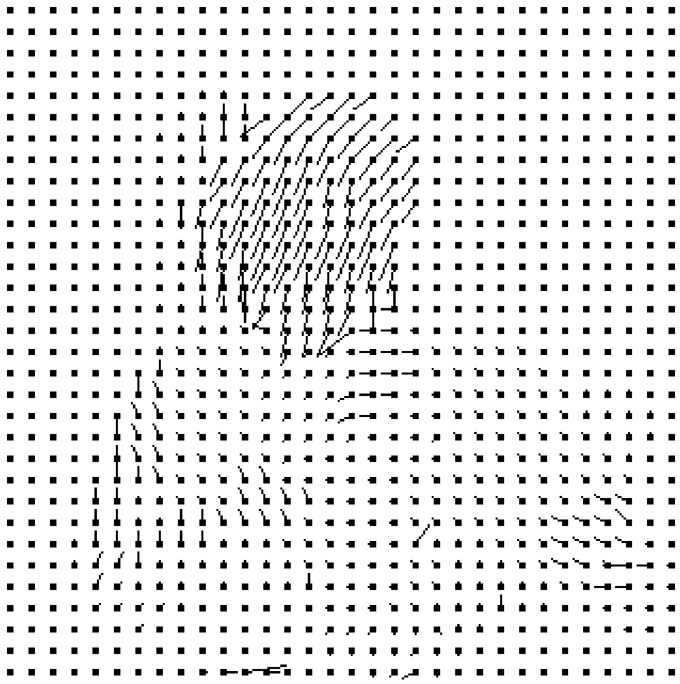

Figure 15: Optical flow smoothed by the ICM.

Figure 16: Movement segmentation provided by the MRBF network.

Figure 17: Movement segmentation provided by the ICM.