Segmentation and Estimation of the Optical Flow

Adrian G. Borş Ioannis Pitas

Department of Informatics
University of Thessaloniki
Thessaloniki 540 06, Greece

Abstract. We propose an algorithm for simultaneous estimation and segmentation of the optical flow. The moving scene is decomposed in different regions with respect to their motion, by means of a pattern recognition scheme. The feature vectors are drawn from the image sequence and they are used to train a Radial Basis Functions (RBF) neural network. The learning algorithm for the RBF network minimizes a cost function derived from the probability estimation theory. The proposed algorithm was applied in real image sequences.

1 Introduction

Visual motion is an important aspect in the human visual system and has lately received an increasing attention. The optical flow consists of the distribution of the velocities associated with the block image elements and accounts for the existence of the motion in the image sequence [1]. Block matching algorithms, widely used in video coding, do not model the optical flow explicitly and in many situations fail to provide the optimal estimation.

Different approaches were investigated for simultaneous optical flow estimation and segmentation. In [2] the theoretical framework was derived from the maximum a posteriori probability. Simulated annealing algorithm was employed for optical flow and line field estimation. This algorithm however, does not always provide closed contours and usually requires many iterations until the convergence is achieved. Iterated Conditional Modes (ICM) is based on a deterministic local optimization approach [3]. It was used for optical flow smoothing by considering the moving objects borders [4, 5]. ICM uses a piecewise Markov Random Field model for the optical flow by associating alternatively different motion vectors to a given block site. The motion vector that decreases a local energy criterion is chosen for a given block site [5].

In this study we propose a global optimization method for the segmentation and estimation of the optical flow based on a pattern recognition scheme. The scene is classified in different moving objects according to an error criterion minimization. This cost function is expressed as a metric which takes into account different image features like the optical flow provided by the block matching, the gray level and the location of the block sites. The feature vectors are classified by means of a Radial Basis Functions network. RBF form a two-layer neural network with good approximation capabilities [6]. The learning algorithm for the RBF network, leading to the minimization of the optical flow segmentation and estimation, is robust at the presence of outliers in the data [7]. Various comparison criteria are used for the RBF network and ICM algorithm when they are applied in the "Hamburg taxi" sequence.

2 Classification of the Moving Objects

Let us denote by $(\hat{\mathcal{M}}_k, \hat{\mathcal{S}}_k)$ the estimates of the optical flow and of the moving object segmentation for the kth moving object, where the optical flow was drawn from two successive frames f_{t-1} and f_t. The classification of the moving objects in the image sequence is given by maximizing the a posteriori probability:

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

From the Bayesian rule, the above probabilities can be expressed as:

$$P(\hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j | f_t, f_{t-1}) = \frac{P(f_t | f_{t-1}, \hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j) P(\hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j | f_{t-1})}{P(f_t | f_{t-1})}. \qquad (2)$$

The probability $P(f_t|f_{t-1})$ does not depend on $(\hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j)$ and can be neglected. The probability $P(\hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j | f_{t-1})$ can be further expressed as:

$$P(\hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j | f_{t-1}) = P(\hat{\mathcal{M}}_j | \hat{\mathcal{S}}_j, f_{t-1}) P(\hat{\mathcal{S}}_j | f_{t-1}), \qquad (3)$$

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 optical flow probability depending on the segmentation and image.

Each of the above probabilities can be expressed as an energy functional depending on the image feature vector $\mathbf{X}$ [2]:

$$P(\mathbf{X}) = \frac{1}{Z} \exp\left[-\frac{E(\mathbf{X})}{\beta}\right] \qquad (4)$$

where E is an energy function, Z is a normalizing constant and β is a constant controlling the properties of the E. The problem of maximizing the probability from (1) is converted into the minimization of an energy function :

$$E_j = E(f_t | f_{t-1}, \hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j) + E(\hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j | f_{t-1}). \qquad (5)$$

The image is partitioned in blocks situated on a rectangular grid. Each block has assigned a feature vector containing the necessary information about the respective image region. The a posteriori probability of the optical flow estimation and segmentation (1) is maximized for those feature vectors drawn from the image regions corresponding to moving objects. This, eventually is expressed as the minimization of the cost function (5) depending on the feature vectors. The energy function is equivalent to the cost of vector representation by the assumed model.

3 Radial Basis Function Network

An RBF network implements a decomposition of a function in a sum of kernels:

$$P_k(\mathbf{u}) = \sum_{j=1}^{L} \lambda(k,j)\phi_j(\mathbf{u}), \tag{6}$$

where $\mathbf{u}$ is an input vector, L is the number of kernels and ϕ_j is:

$$\phi_j(\mathbf{u}) = \exp\left[-\sum_{i=1}^{N}\left(\frac{\mu_j(i)-u(i)}{r_j(i)}\right)^2 - H(\mu_j)\right], \tag{7}$$

N is the feature space dimension, μ_j, $\mathbf{r}_j$ are the centers and the variances of the Gaussian function and $H(\mu_j)$ is the linear part.

The first component in the expression of E_j (5) represents how well the frame f_t is represented based on the estimates $(\hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j)$:

$$E(f_t|f_{t-1}, \hat{\mathcal{M}}_j, \hat{\mathcal{S}}_j) = H(\mu_j) = \sum_{I=1}^{N_x}\sum_{J=1}^{N_y} d_{IJ}(\mu_j)^2 \tag{8}$$

where N_x and N_y are the number of blocks on x and y axes and $d_{IJ}(\mu_j)$ is the displaced difference [1] of the block site IJ for the center vector μ_j of the jth basis function.

We assign a five dimensional feature vector for each image block site. The selected features record informations about the location, gray level and motion estimate for a specific block from the image sequence. In the case of color image sequences two more features are considered, corresponding to color components. The cost function from (5) is expressed with respect to the input features as a metric in the input space:

$$E_j = \sum_{I=1}^{N_x}\sum_{J=1}^{N_y}(\mathbf{u}_{IJ}-\mu_j)(\mathbf{u}_{IJ}-\mu_j)^T + H(\mu_j) \tag{9}$$

where $\mathbf{u}_{IJ}$ are the feature vectors and $H(\mu_j)$ is the displaced difference for the chosen RBF center given by (8). In the RBF updating we use an approach based on the Learning Vector Quantization (LVQ) algorithm [8]. For a certain feature vector we update the RBF function parameters for that basis center which provides a minimum energy E_j in (9).

In the training stage, it is desirable to avoid using outlying patterns which may cause bias in the estimation of the parameters. The robust estimation for the Radial Basis Functions was analyzed in [7] and compared with that based on the classical LVQ algorithm. A fast histogram based implementation algorithm is used for the RBF robust training as it was described in [9]. The data samples assigned to a basis function (9) are marginally ordered and the center is taken as the marginal median:

$$\hat{\mu}_j = \text{med}\,\{\mathbf{X}_0, \mathbf{X}_2, \ldots, \mathbf{X}_{n-1}\} \tag{10}$$

where $\mathbf{X}_{n-1}$ is the last pattern assigned to the cluster j. For the variance parameters we use the median of the absolute deviation (MAD) estimator:

$$\hat{\sigma}_{j,h} = \frac{\text{med}\ \{|\mathbf{X}_0 - \hat{\mu}_j|, \ldots, |\mathbf{X}_{n-1} - \hat{\mu}_j|\}}{0.6745} \tag{11}$$

where 0.6745 is a scaling parameter.

We consider a geometrical interconnectivity relationship between two hidden units s and v based on the locations of the block sites assigned to them. This relationship is denoted as V_{sv} and consists of the sum of block sites assigned to the hidden unit s, neighboring the block sites assigned to the hidden unit v. The image block sites assigned to a hidden unit s are considered part of a moving object v if they make up compact regions in the image:

$$s \;=\; \arg \max_{k=1}^{L} V_{kv}. \tag{12}$$

If $s \equiv v$ then to the respective hidden unit has been assigned block sites consistent with a moving object. This procedure finds a certain number of moving objects. For each moving object is assigned an output unit. The output weights $\lambda(j,k)$ from (6) are updated using the backpropagation algorithm [6].

4 Simulation Results

A frame from the "Hamburg taxi" sequence is shown in Fig. 1. This sequence contains three principal moving objects: 1) the taxi turning around the corner, 2) a car in lower left, driving from left to right, 3) a van in the lower right driving from right to left. The optical flow obtained by using full search block matching algorithm is presented in Fig. 2. The block size is 4×4 pels.

The optical flow and the moving object segmentation by means of the RBF network are provided in the Figures 3 and 4. For each moving object only one motion vector is assigned. For comparison purposes the Iterated Conditional Modes (ICM) was tested [3, 5] in the same conditions. The results obtained after applying ICM are presented in the Figures 5 and 6. RBF provides a better optical flow estimation than ICM as well as more accurate boundaries for the segmentation. RBF network uses the interactions among all the block sites assigned to a hidden unit while ICM is based on a local smoother. The motion vectors assigned to the moving objects by means of the RBF network are very accurate.

The comparison criteria for the segmentation of the moving objects are the misclassification error expressed as the percentage of erroneously classified block sites, MSE, MAE and the number of model parameters. The results are displayed in Table 1. For both methods, the reference was considered to be the optimal optical flow and moving objects segmentation obtained in a semiautomatic way. The RBF network based representation of the motion from the image sequences provides good results in the optical flow processing and a big reduction in the necessary number of parameters. The total number of parameters required by the RBF network is $(10+M)L$. ICM algorithm requires $3N_xN_y$ parameters.

Method	Classification error (%)	MAE	MSE	Number of Parameters
RBF	3.02	0.17	0.85	210
ICM	4.07	0.33	1.15	9216

Table 1. Comparison between the RBF network and ICM.

5 Conclusions

In this study we propose a new approach for the optical flow estimation and segmentation. The proposed algorithm is derived from a classification criterion for the moving objects and it employs the Radial Basis Functions networks. The classification criterion is used for the RBF network training and takes into account the motion information provided by the block matching, the gray level, the location of the block sites and it is based on a minimal displaced frame difference. After a robust and efficient learning stage, the algorithm assigns one motion vector for each moving region.

References

1. A. N. Netravali, B. G. Haskell, *Digital Pictures, Representation and compression,* (Applications of Communications Theory, series editor: R. W. Lucky), Plenum Press, 1988.
2. J. Konrad, E. Dubois, "Bayesian estimation of motion vector fields" *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. PAMI-14, No. 9, pp. 910-927, Sep. 1992.
3. J. Besag, "On the statistical analysis of dirty pictures," *J. Roy Statist. Soc.,* vol. 48, ser. B, No. 3, pp. 259-302, 1986.
4. F. Heitz, P. Bouthemy, "Multimodal estimation of discontinuous optical flow using Markov Random Fields," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. PAMI-15, No. 12, Dec. 1993.
5. M. M. Chan, 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.
6. A. G. Borş, M. Gabbouj, "Minimal topology for a radial basis functions neural network for pattern classification," *Digital Signal Processing,* vol. 4, no. 3, pp. 173-188, July 1994.
7. A. G. Borş, I. Pitas, "Robust estimation of radial basis functions," *Proc. of IEEE Workshop on Neural Networks for Signal Processing IV,* Ermioni, Greece, pp. 105-114, 6-8 Sep. 1994.
8. R. J. Schalkoff, *Pattern Recognition: Statistical, Structural and Neural Approaches,* John Wiley, 1992.
9. A. G. Borş, I. Pitas, "Median radial basis functions network for optical flow processing," *Proc. of IEEE Workshop on Nonlinear Signal and Image Processing,* Neos Marmaras, Greece, pp. 702-705, 20-22 June 1995.

Fig. 1. The first frame from the "Hamburg taxi" sequence.

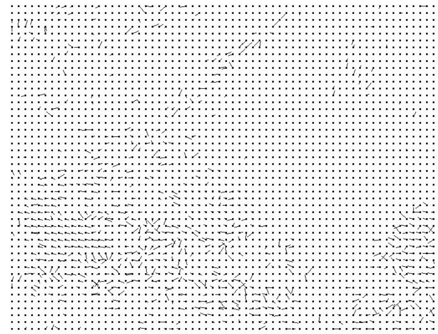

Fig. 2. Optical flow provided by the block matching estimation.

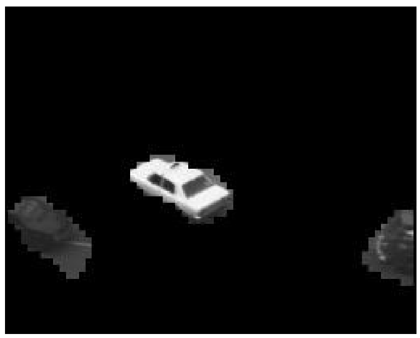

Fig. 3. RBF based segmentation.

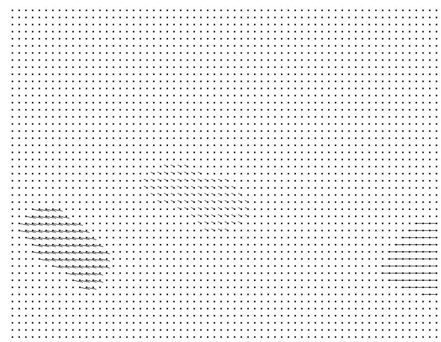

Fig. 4. The optical flow processed by RBF.

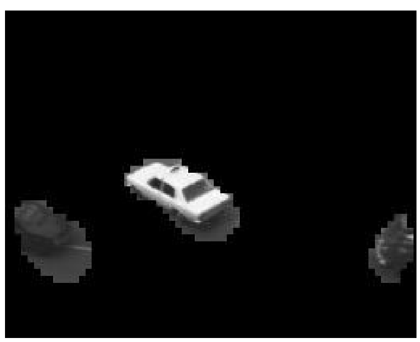

Fig. 5. ICM based segmentation.

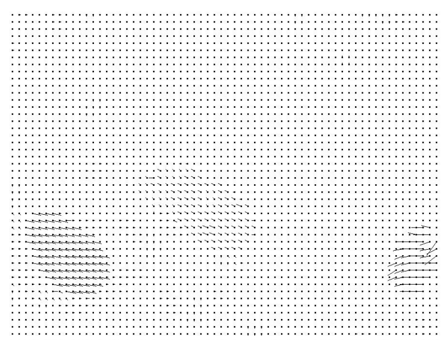

Fig. 6. The optical flow processed by ICM.