Segmentation of moving objects by robust motion parameter estimation over multiple frames *

S. Ayer, P. Schroeter and J. Bigün

Swiss Federal Institute of Technology, Signal Processing Laboratory
CH-1015 Lausanne, Switzerland

Abstract. A method for detecting and segmenting accurately moving objects in monocular image sequences is proposed. It consists of two modules, namely a motion estimation and a motion segmentation module. The motion estimation problem is formulated as a time varying motion parameter estimation over multiple frames. Robust regression techniques are used to estimate these parameters. The motion parameters for the different moving objects are obtained by successive estimations on regions for which the previously estimated motion parameters are not valid. The segmentation module combines all motion parameters and the gray level information in order to obtain the motion boundaries and to improve them by using time integration. Experimental results on real image sequences with static or moving camera in the presence of multiple moving objects are reported.

1 Introduction

Detecting and segmenting moving objects in image sequences have been given a large attention in the research community. Most of the early research has been concentrated on the estimation of optical flow computed between image pairs. Such a flow field assigns to each pixel of one image a translational vector containing local motion information. The classical pixelwise classification based on the optical flow can be employed to detect multiple motions (see for example [1] and [11]). However, these methods are very sensitive to the quality of the optical flow, and as small spatial and temporal regions do not always carry sufficient motion information, the optical flow computation can be very inaccurate. Another approach is the use of parametric motion estimators, which describe the motion over a larger spatial region in terms of a parametric model. In this case, a model is used in order to constrain the flow field computation, so that the flow field does not vary in an uncontrolled fashion. However, when the region of analysis contains multiple moving objects, one is compelled to use an estimation method that can recover simultaneously the model parameters and the motion discontinuities. To achieve this goal, several ideas have been proposed, among

* This work has been supported by Thomson-CSF, Rennes,France

Lecture Notes in Computer Science, Vol. 801
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94
© Springer-Verlag Berlin Heidelberg 1994

which we can discriminate those based on line process (discontinuities detection) [5] and outlier detection [9, 7].

In this paper, we propose a new method for time-varying motion analysis and segmentation which uses both a large spatial region and a large temporal support. The motion estimation problem is formulated as one of time-varying parameter estimation over multiple frames. A robust regression technique [12] is used to estimate the motion parameters. As these techniques are resistant to a given percentage of outliers in the data, they allow us to overcome the problem of multiple moving objects inside the region of analysis. Such techniques have recently been used in computer vision [10], e.g. motion parameter estimation and segmentation in a pair of images [3, 6, 7], detection of moving objects [16].

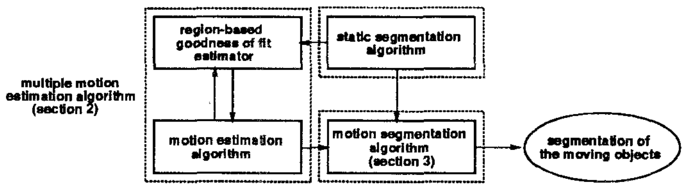

Fig. 1. Block diagram of the system

The different blocks of the method are shown in Fig. 1. Note that, in this paper, we assume that the number of moving objects does not change in the image sequence. The motion estimation algorithm is first applied to an entire region of the image (here the whole image). By means of a segmentation algorithm [13], we then obtain subregions with similar gray level and determine in which of these regions the estimated parameters are not valid. Further motion estimations are applied to these regions. Thus, by successive application of the robust motion estimation algorithm, the number of regions for which the previously estimated motion parameters could not explain the motion well is reduced to zero. Then, the motion segmentation algorithm combines the motion parameters coming from the multiple estimations and the gray level information in order to obtain the motion boundaries and to improve them by using time integration. The pixels within a statically segmented subregion are constrained to follow the same motion allowing us to classify correctly pixels with low gradient information.

Experimental results indicate that the proposed scheme is robust to the presence of different moving objects and is also general enough to deal with scenes with moving or static cameras, with objects close to or far from the camera, and in a stationary or non-stationary environment.

The time-varying robust motion estimation algorithm is presented in the next section. Section 3 describes how moving objects are segmented by using multiple frames. Finally, experimental results are reported in Sect. 4 and concluding remarks are given in Sect. 5.

2 Model-Based Time-Varying Motion Estimation

The motion estimation problem is here formulated as one of time-varying parameter estimation. In Fig. 2, we give an overview of the different basic operations which are involved in the implementation of the algorithm.

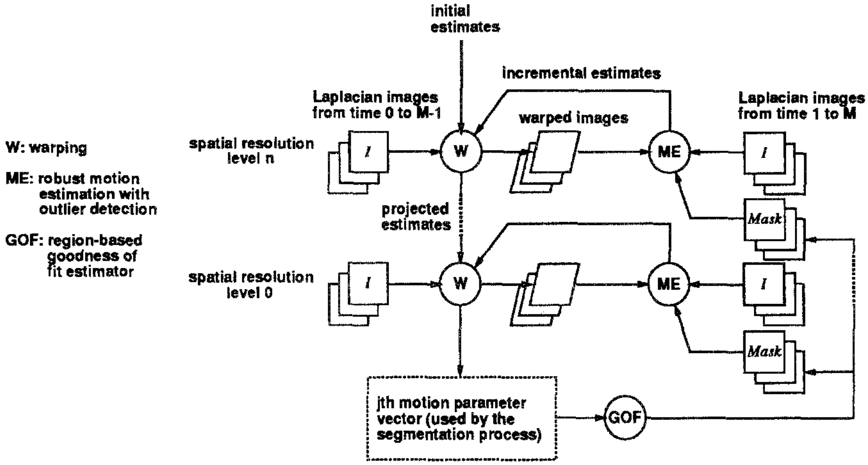

Fig. 2. General framework of the robust motion estimator

In our method, the estimation is performed in a hierarchical and iterative way. In our implementation, we use the Laplacian image pyramid as input to the algorithm. The estimation process begins at the coarsest resolution level, using zero initial estimates. An estimation step consists of two operations: the image sequence warping (using bicubic interpolation) and the estimation of the motion parameter increments. As it will be explained in Sect. 2.1 and 2.2, the incremental motion parameters are estimated by using using a robust linear regression procedure and an outlier detection technique.

Using the final time-varying motion parameter estimates (obtained at resolution level 0), a region-based goodness of fit is computed for each frame. The goal of this operation is to classify regions into two classes: the regions where the estimated parameters are valid and those where they are not. This point will be explained in more details in Sect. 2.3.

The result of the multiple motion estimation algorithm is a set of time-varying parameter vectors $\mathbf{p}_j(t)$. As explained in Sect. 3, these vectors are used together with gray level information in order to obtain an accurate segmentation of the moving objects.

2.1 From parametric to time-varying parametric estimation

If we consider only two frames in the sequence and assume intensity constancy (i.e. the brightness of a small surface patch is not changed by motion), the problem of motion model fitting can be posed as the minimization over the region of analysis of a function of

$$I(\mathbf{x}, t) - I(\mathbf{x} - \mathbf{u}(\mathbf{x}, \mathbf{p}), t - 1), \tag{1}$$

where $\mathbf{p}$ denotes the model parameters, $\mathbf{u}(\mathbf{x}, \mathbf{p})$ the flow field in that region (e.g. translational, affine or planar), and $I(\mathbf{x} - \mathbf{u}(\mathbf{x}, \mathbf{p}), t - 1)$ the image at time $t - 1$ warped towards t [2, 3, 4].

With this formulation, the use of several frames for motion estimation would lead to several distinct estimations between consecutive pairs of images. In this case, the parameter evolutions can be described as a time series. However, for robustness and efficiency purposes, we want here to integrate more measurements into a single estimation process. Since the motion to be modeled is time-varying, the model parameters to be estimated should also be time-varying. In our method, each coefficient in the model is allowed to change in time by defining it as a linear combination of some known time functions. This approach has been inspired by methods of time-varying parametric modeling in speech processing [8].

By limiting our attention to such a time-varying model, we are clearly constraining the possible types of time variations that can be allowed. However, constraints on the nature of the time variations are essential in order to limit the degrees of freedom of the time-varying parameters, so that incoherent and noisy estimations can be avoided. A judicious choice of the basis functions can provide a good approximation of a wide variety of motion parameter time variations with only a few coefficients. In our experiments, we used the trigonometric (Fourier) functions, as well as the Legendre and Hermite polynomials. A very important point is also that this method can make use of the spatial multiresolution for determining the degree of freedom of the time-varying motion parameter estimation.

With a model of this form, the coefficients in the linear combination are to be estimated from the image sequence and the problem of motion estimation is posed as the minimization of a function of

$$I(\mathbf{x}, t_a) - I(\mathbf{x} - \mathbf{u}(\mathbf{x}, \mathbf{p}(t_a)), t_{a-1}), \qquad j = 1, \ldots, M \tag{2}$$

over the region of analysis and over $M + 1$ frames. In this formulation, $\mathbf{p}(t_a)$ denotes the motion parameter vector at time t_a, whose components are modeled as linear combinations of some known functions of time $f_k(t)$:

$$p_i(t) = \sum_{k=0}^{F-1} a_{i+Pk}\ f_k(t), \qquad i = 0, \ldots, P - 1 \tag{3}$$

where P is the number of motion parameters (e.g. translational model ($P = 2$), affine model ($P = 6$)) and F the number of time functions.

The problem of estimating the parameter vector $\mathbf{a}$ from equation (2) is non-linear, and this system of non-linear algebraic equations has no closed form solution. Consequently, one is compelled to solve the system of equations numerically in order to find an estimate of the parameter vector $\mathbf{a}$. Suppose that an initial estimate $\hat{\mathbf{a}}$ is available (in our case, it will be the parameter vector 0 at the coarsest resolution level and the projected estimates at the other levels). The problem may be linearized using $\hat{\mathbf{a}}$ and the first order Taylor expansion of equation (2) leads to

$$\Delta I(\mathbf{u}(\mathbf{x},\mathbf{p}))|_{\mathbf{a}=\hat{\mathbf{a}}} = [(\nabla I(\mathbf{x}-\mathbf{u}(\mathbf{x},\mathbf{p}),t_{a-1}))^T \frac{\delta \mathbf{u}}{\delta \mathbf{p}} \frac{\delta \mathbf{p}}{\delta \mathbf{a}}]|_{\mathbf{a}=\hat{\mathbf{a}}} (\mathbf{a}-\hat{\mathbf{a}}) \qquad (4)$$

where

$$\Delta I(\mathbf{u}(\mathbf{x},\mathbf{p})) = I(\mathbf{x},t_a) - I(\mathbf{x}-\mathbf{u}(\mathbf{x},\mathbf{p}(t_a)),t_{a-1}).$$

In equation (4), $\nabla \mathrm{I}(\mathbf{x}-\mathbf{u}(\mathbf{x},\mathbf{p}),t-1)$ denotes the spatial gradient vector $[\frac{\delta I}{\delta x}, \frac{\delta I}{\delta y}]^T$ of the warped image, $\frac{\delta \mathbf{u}}{\delta \mathbf{p}}$ the $2 \times P$ Jacobian matrix of the vector field $\mathbf{u}$, and $\frac{\delta \mathbf{p}}{\delta \mathbf{a}}$ the $P \times PF$ Jacobian matrix of the motion parameter vector $\mathbf{p}$.

Equation (4) shows that, after linearization of the model, the problem of time-varying model-based motion estimation can be reformulated to one of parameter estimation. Standard or robust linear regression methods may be used to estimate the parameters $\delta\mathbf{a}$ (see Sect. 2.2).

2.2 Robust linear regression estimators

The LS estimator is known to be optimal for Gaussian noise distribution. However, more recently attention has been given to the fact that LS analysis is very sensitive to minor deviations from the Gaussian noise model and to the presence of outliers in the data. In order to reduce the impact of these negative influences, robust methods that are much less affected by outliers have been developed (e.g. LMedS and LTS [12]). In computer vision the problem of regression analysis is an important statistical tool and recently the interest for robust estimators has increased [10].

In this paper, we discuss the application of two robust estimators to the problem of time-varying motion estimation, namely the *least median of squares* (LMedS) and the *least-trimmed squares* (LTS) given by

$$\min_{\hat{\Theta}} \operatorname{median}_i(r_i^2) \qquad \text{and} \qquad \min_{\hat{\Theta}} \sum_{i=1}^{h} (r^2)_{i:n} \qquad (5)$$

where $(r^2)_{1:n} \leq \ldots \leq (r^2)_{n:n}$ are the ordered squared residuals.

In the case of motion estimation, the sensitivity of LS estimators to the presence of outliers may be exhibited by the following example. In Fig. 3, we show a 3-D plot of the translational model objective function for the TAXI sequence (see Fig. 5). In these plots, dx (not shown) and dy axes represent the two translational motion parameters (dx, dy) and the z axis represents the objective function to be minimized by the motion estimation algorithm. In this case, the minimum objective function value should be at $(dx, dy) = 0$ (zero background

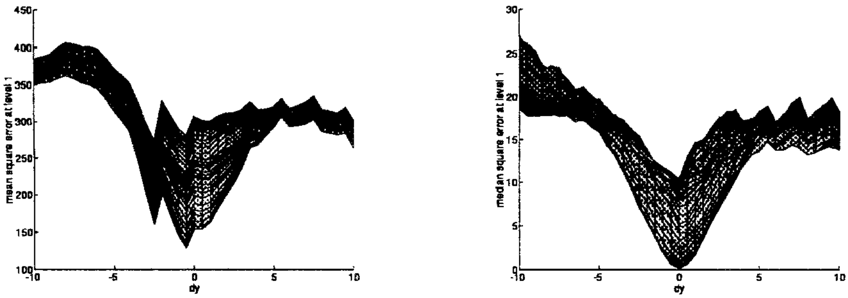

Fig. 3. Objective function plots for the TAXI sequence (translational model): (left) mean square error (right) median square error

motion). These plots exhibit the resistance of the LS and LMedS estimators to outliers, which are caused by the multiple motions present in the image sequence. The biases introduced in the LS estimates make their use difficult for outlier detection and for segmentation purposes.

2.3 Region-based goodness of fit measures

In order to detect the different moving objects present in the scene, we need to find the regions where the estimated parameters are not valid, onto which further motion estimations should be applied. For this purpose, we compute a goodness of fit measure in each region obtained from the static segmentation [13]. Note that the first computed motion is obtained generally by applying the motion estimation algorithm on the whole image.

As the robust linear regression estimators are combined with a least-squares procedure, the value minimized by the motion estimation algorithm may be expressed as

$$\gamma = (\Delta I(\mathbf{u}(\mathbf{x},\mathbf{p})))^T \; \Delta I(\mathbf{u}(\mathbf{x},\mathbf{p})) \tag{6}$$

which is the sum of the residuals. With the Gaussian noise assumption, this sum is the sum of the squares of MN independent scalar random variables with zero mean and unity variance (assuming that the errors are normalized), that is γ has a chi-square distribution with $n - PF$ degrees of freedom. Here, MN is the number of points in the image sequence and PF is the dimension of the parameter vector $\mathbf{a}$.

Based on this, γ is a measure of the goodness of fit. This measure is computed over each statically segmented region of each original frame of the sequence and the parameters are considered as not valid if

$$\gamma > c \tag{7}$$

where the threshold c is obtained such that the probability of a $n - PF$ degree of freedom chi-square random variable exceeding it is α (here 5 %).

Once the regions have been classified as valid or not valid, connected regions are merged. In Fig. 5 (row 1, left), we show the results of this computation obtained for the TAXI sequence after the first robust motion estimation, which returns estimated parameters very close to zero. In this figure, the black regions stands for not valid regions. In order to remove tiny regions, a median filtering is applied (Fig. 5 (row 1, right)).

At this stage we need to find masks for further motion estimation. For this purpose, correspondences have to been found between the different valid regions. Due to the assumption of unchanged object configuration and to the fact that the motion estimation behaves very consistently over time, this is not a very complicated correspondence problem. Methods such that presented in [14] can be used for this purpose. The final result of the region-based goodness of fit estimator is a set of mask sequences that can be used for further motion estimations.

3 Segmentation of multiple moving objects

This section describes a method that combines motion and luminance for an accurate segmentation of multiple moving objects. An attempt in this direction was proposed in [15]. A different approach (based only on motion) consists in computing the prediction error pixelwise or in a small neighborhood and then segmenting the resulting error image [3, 9]. In our experiments, this approach worked well with sequences containing textured moving objects but failed for objects where the gradient information is low. Here, we propose to overcome this problem by constraining the pixels within a spatially segmented subregion to follow the same motion.

First, we apply a static segmentation algorithm [13] on the last frame of an image sequence in order to obtain the subregions with similar gray-level. This static segmentation is embedded in a multiresolution framework using quadtrees [17]. Multiple resolutions are very useful because at lower resolutions the noise is reduced, allowing the class-centers to be better defined, whereas higher resolutions are needed to obtain accurate borders. Here the boundary refinement step will be applied only once the motion labeling is done.

The estimates of the different motions are used to classify each subregion, thus merging the subregions with the same motion. This can be done by computing the prediction errors corresponding to the different detected motions for all the subregions, and to assign them a motion label corresponding to the minimum error. We also propose to use more frames to increase the certainty that a subregion is correctly classified. If a subregion is detected as having the same motion in successive frames then its classification certainty increases.

Figure 4 shows the block diagram of the motion segmentation algorithm. Suppose that we have a set of $M+1$ successive frames at times $t_a, a = 0, \ldots, M$. Time-varying motion parameter vectors, $\mathbf{p}_j(t_a)$ are first estimated by the multiple motion estimation algorithm (Sect. 2). The image sequence is then warped towards the last frame with the composition of all motions $C(\mathbf{p}_j(t_a), a =$

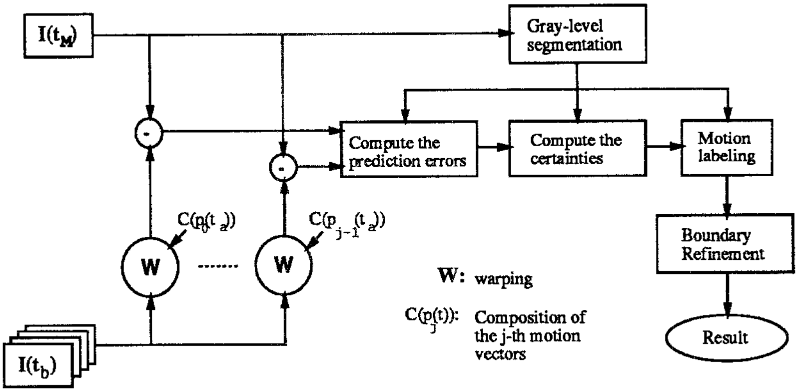

Fig. 4. Block diagram of the motion segmentation algorithm

$M-1,\ldots,b)$, for $b=M-1,\ldots,0$. For example, assume that we have only translational parameters and 3 frames ($M=2$):

$$
\begin{aligned}
b=1&: C(\mathbf{p}_j(t_a), a=1,\ldots,1) = \mathbf{p}_j(t_1)\\
b=0&: C(\mathbf{p}_j(t_a), a=1,\ldots,0) = \mathbf{p}_j(t_1)+\mathbf{p}_j(t_0)
\end{aligned}
\tag{8}
$$

This results in $M \times N_m$ warped images where N_m is the number of detected motions.

Starting from $b=M-1$ (until $b=0$), the prediction errors are computed, at the coarsest level of the pyramid, by using the following robust objective function

$$
e_{ij}(t_b) = \sum_{k=1}^{N_i/2} (r_{ij}^2)_{k:N_i}, \qquad j=0,\ldots,N_m-1, \qquad i=0,\ldots,N_s-1 \tag{9}
$$

where

$$
r_{ij} = I(\mathbf{x}_{ik}, t_M) - I(\mathbf{x}_{ik} - \mathbf{u}(\mathbf{x}_{ik}, C(\mathbf{p}_j(t_a), a=M-1,\ldots,b)), t_b) \tag{10}
$$

and $(r_{ij}^2)_{1:N_i} \leq \cdots \leq (r_{ij}^2)_{N_i:N_i}$. Here, I represents the original frames, N_i is the size of region S_i, $\mathbf{x}_{ik} \in S_i$, N_s the number of subregions and $\mathbf{u}(\mathbf{x}_{ik}, C(\mathbf{p}_j(t_a), a= M-1,\ldots,b))$ is the flow field obtained with the composition of the parameter vectors. A robust objective function (like in equation (5)) is used because a small fraction of pixels of a subregion with high residuals can have a strong influence on the motion labeling and can even cause a false classification of a subregion.

The motion labeling is computed by means of a measure of classification certainty that evolves with the number of frames. For each subregion i and for each motion j, the certainty $c_{ij}(t_b)$ is obtained by computing the normalized difference between products of prediction errors:

$$
\begin{array}{ll}
c_{ij}(t_b) = \dfrac{Pd_2 - Pd_1}{Pd_2 + Pd_1} & 0 \le j' \neq j < N_m \\
Pd_1 = \prod_{a=b}^{M-1} e_{ij}(t_a) & Pd_2 = \prod_{a=b}^{M-1} e_{ij'}(t_a)
\end{array}
\tag{11}
$$

where $e_{ij'}(t_a)$ is the closest prediction error at time t_a (corresponding to motion j') smaller than $e_{ij}(t_a)$ if it exists, or the closest prediction error greater than $e_{ij}(t_a)$ otherwise. Note that $c_{ij}(t_b)$ may take negative values when $Pd_2 \le Pd_1$. If the motion label j corresponding to the minimum error of prediction remains the same (i.e. $e_{ij'}(t_a) > e_{ij}(t_a)$ for all t_a) then the certainty $c_{ij}(t_b)$ will tend to increase. However, if the motion labels change with time or if more than one motion is present within a subregion (i.e. $e_{ij'}(t_a) \approx e_{ij}(t_a)$), $c_{ij}(t_b)$ will be close to or smaller than zero.

After all frames have been considered (i.e. $b = 0$), the subregions can be classified according to the maximum certainty, resulting in a motion label image $L(\mathbf{x}_{ik})$ at the coarsest level of the quadtree:

$$
L(\mathbf{x}_{ik}) = \arg\max_j c_{ij}(t_b = 0) \qquad 0 \le i < N_s, \quad 0 \le j < N_m, \quad 1 \le k \le N_i \tag{12}
$$

The tiny subregions of $L(\mathbf{x}_{ik})$ can then be reassigned to their neighborhood. Here, we assume that a small subregion enclosed in a larger subregion must have the same motion than the larger one. Finally, the boundary refinement [13] is applied only on the boundaries corresponding to motion boundaries of $L(\mathbf{x}_{ik})$.

4 Results

The method described in the previous sections has been applied to different sequences. The first one, called the BBC sequence, shows a car moving to the left and tracked by the camera inducing a motion of the background to the right. The difficulties of this sequence are due to the fact that the car is close to the camera and that the car motion is not parallel to the image plane. For simplicity, we used here a translational model but the affine model yields, in this case, similar results. Figure 5 (row 3, right) shows the last frame of the sequence ($M = 4$) and its static segmentation (row 3, left) computed with $c = 4$ classes at the second level of the quadtree, yielding 251 subregions. In order to prevent subregions having different motions from being merged, different values of c ($c = 4, 5$ or 6 typically) can be used in the clustering algorithm, causing only slight changes in the final segmentation result.

Figure 5 (row 2) shows the evolution of the motion labeling. Black labels correspond to subregions that are classified with certainty under a threshold of 0.9. We can observe that the number of such subregions is reduced as the number of frames increases. Only tiny subregions remain unclassified after the last iteration partly because of unrecovered areas caused by the warping operation. Figure 5 (row 3) shows the final motion labeling after the reassignment of the tiny unclassified (or misclassified) subregions and boundary refinement. Accurate boundaries can be observed and it can be seen that small details like most parts of the antenna are preserved. As the shade underneath the car is moving with the same motion as the car, the segmentation at this place is correct. However, the back window of the car is misclassified because it is merged to the background by the static segmentation algorithm.

The same procedure (with the same parameters) was applied to the TAXI sequence that contains three moving cars on a static background. The final segmentation results and the superimposed boundaries can be seen in Fig. 5 (row 4). First, it can be noticed that the three cars are correctly classified and that the white taxi boundaries are accurate. On the left car, the gray level of part of the roof was merged to the background by the static segmentation. On the right, the car is passing behind a tree disturbing the motion segmentation. Results for a sequence (SAL) where only the camera is moving and where different motions are induced due to depth discontinuities is also shown in Fig. 5 (row 4).

5 Conclusion

In this paper, we presented a method that combines motion and gray level information in a set of successive frames in order to detect and segment multiple moving objects. We showed that robust regression methods allow us to estimate accurate motion parameters where the standard least squares estimation scheme fails. Unlike two-frames based motion estimation, our scheme estimates time varying motion parameters over a set of successive frames by defining them as linear combinations of a set of known time functions. This approach constrains temporally the motion parameters in order to avoid incoherent and noisy estimations. Here, we assume that the number of different moving objects does not change in the considered sequence. A strategy to detect the appearance or disappearance of moving objects still needs to be investigated. More attention should also be devoted to the automatic choice of the motion model to be used and of the degree of freedom of the time-varying parameter estimation.

The second part of the algorithm finds accurate boundaries between the moving objects by combining all the detected motions over the successive frames and the gray level information. Subregions with similar gray levels are given a motion label according to the a motion certainty that increases with the number of frames. Here, we first find the different subregions of a frame by means of a clustering algorithm which constrains the pixels of a subregion to move with the same motion. This approach allows to assign a motion label to pixels where the motion information alone is not sufficient for a correct motion labeling. However, problems might arise if a statically segmented subregion contains two or more different motions. In this case, some pixels of such a subregion will be misclassified. With robust objective functions we already reduced the risk of subregion misclassification. We are currently investigating a method that re-segments (into more classes) subregions with low motion certainty. This operation should further reduce the number of misclassified pixels.

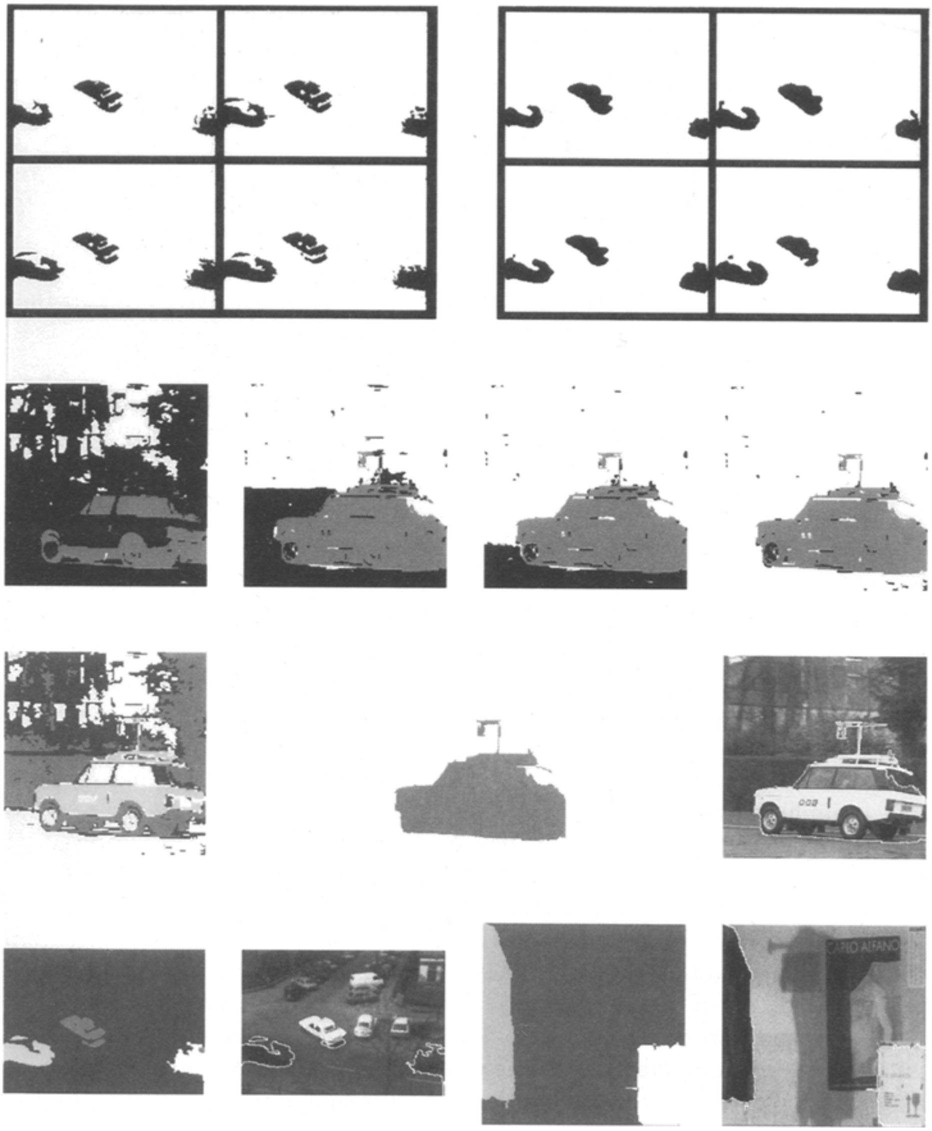

Fig. 5. (Row 1 from top) Region-based goodness of fit masks for the TAXI sequence after the first motion estimation (increasing time from left to right and from top to bottom) (Row 2) Evolution of the motion labeling by using 2, 3, 4 and 5 frames. Black labels correspond to subregions that are classified with certainty under a threshold of 0.9. (Row 3) (left) Static segmentation of the last frame (middle) Segmentation result using 5 frames after the reassignment of tiny subregions. (right) Boundaries superimposed on the fifth frame. (Row 4) (left) Segmentation result on the TAXI sequence (right) Segmentation result on the SAL sequence.

References

1. G Adiv. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:384–401, July 1985.
2. S. Ayer and P. Schroeter. Hierarchical robust motion estimation for segmentation of moving objects. In *Eigth IEEE Workshop on Image and Multidimensional Signal Processing*, pages 122–123, Cannes, France, September 1993.
3. S. Ayer, P. Schroeter, and J. Bigün. Tracking based on hierarchical multiple motion estimation and robust regression. In *Time-Varying Image Processing and Moving Object Recognition, 3*, Florence, Italy, June 1993.
4. J.R. Bergen, P. Anandan, K.J. Hanna, and J. Hingorani. Hierarchical model-based motion estimation. In *Second European Conference on Computer Vision*, pages 237–252, Santa Margherita Ligure, Italy, May 1992.
5. M.J. Black. Combining intensity and motion for incremental segmentation and tracking over long image sequences. In *Second European Conference on Computer Vision*, pages 485–493, Santa Margherita Ligure, Italy, May 1992.
6. M.J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *Fourth International Conference on Computer Vision*, pages 231–236, Berlin, Germany, May 1993.
7. T. Darrell and A. Pentland. Robust estimation of a multi-layered motion representation. In *IEEE Workshop on Visual Motion*, pages 173–178, Nassau Inn, Princeton, NJ, October 1991.
8. M.G. Hall, A.V. Oppenheim, and A.S. Willsky. Time-varying parametric modeling of speech. *Signal Processing*, 5:267–285, 1983.
9. M. Irani, B. Rousso, and S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In *Second European Conference on Computer Vision*, pages 282–287, Santa Margherita Ligure, Italy, May 1992.
10. P. Meer, D. Mintz, A. Rosenfeld, and D.Y. Kim. Robust regression methods for computer vision : A review. *International Journal of Computer Vision*, 6(1):59–70, 1991.
11. A. Rognone, M. Campani, and A. Verri. Identifying multiple motions from optical flow. In *Second European Conference on Computer Vision*, pages 258–266, Santa Margherita Ligure, Italy, May 1992.
12. P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection.* John Wiley and Sons, New York, 1987.
13. P. Schroeter and J. Bigün. Image segmentation by multidimensional clustering and boundary refinement with oriented filters. In *Gretsi Fourteenth symposium*, pages 663–666, Juan les Pins, France, Septembre 1993.
14. S.K. Sethi and R. Jain. Finding trajectories of feature points in a monocular image sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:56–73, January 1987.
15. W.B Thompson. Combining motion and contrast for segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2:543–549, 1980.
16. W.B. Thompson, P. Lechleider, and E.R. Stuck. Detecting moving objects using the rigidity constraint. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15:162–166, February 1993.
17. R. Wilson and M. Spann. *Image Segmentation and Uncertainty.* Research Studies Press Ltd., Letchworth, England, 1988.