

A RECURSIVE APPROACH TO THE DESIGN OF ADJUSTABLE LINEAR MODELS FOR COMPLEX MOTION ANALYSIS

Manuela Chessa, Silvio P. Sabatini, Fabio Solari, Giacomo M. Bisio
Department of Biophysical and Electronic Engineering
University of Genoa
Via all'Opera Pia 11A
16145 Genova - ITALY
email: manuela@dibe.unige.it

ABSTRACT

Parametric models are widely used in motion analysis. Traditionally, affine or learned models are adopted. Here, we propose the use of a set of linear models that dynamically adjust their properties to approximate first-order structures in noisy optic flow fields. Each model is generated by the evolution of a recursive network that can be used as a process equation of a multiple model Kalman Filter. The presence of a model is checked by computing the consistency between the observations (data) and the predictions (model). In each image region, for each model, a probability value can be computed, on which to base motion analysis. Experimental results on multiple motion detection problems and facial expressions analysis validate the approach. The algebraic transformations relating our linear descriptors with the traditional affine models are discussed.

KEY WORDS

Recursive Filtering, Motion Detection, Kalman Filter, Optic Flow.

1 Introduction

Reliable complex (e.g. multiple or non rigid) motion analysis is a challenging problem in computer vision, with several impacts in different application domains. Indeed, distinguishing on visual basis, different motion causes may help the recognition of actions and events [1] [2] [3], such as gestures [4] [5] and facial expressions [6], the location of objects whose trajectory could intersect observer's path, or to coordinate movements to interact with other moving objects (separating ego-motion from independent object motion), as well the reconstruction of the 3-D structure of the observed scene.

The multiple motion detection problem can be addressed as a segmentation problem relying on local descriptors of the optic flow. A popular class of local flow descriptors is based on parameterized models of optic flow [7]. Such models, learned from examples [8] [9], or specified a priori as constant and affine (linear) models, are characterized by a small number of parameters, which provide a concise description of the optic flow structure that can be used to recognize motion patterns from image sequences. In general, linear models can be used both for estimating

optic flow directly from the spatio-temporal image derivatives and for filtering a dense optic flow field. In the recent years, the former approach greatly affermated [10] since the recovering of the model coefficients directly from the spatiotemporal variations of image intensity improves the accuracy and stability of the motion estimates. These methods work very well when the model is a good approximation to the image motion, but they fall short when large image regions are not well modeled by a single parametric motion. This could happen because of the complexity of motion or because of the presence of multiple motions.

In this paper, we propose a method to design adjustable linear models for the analysis of complex dense optic flow fields. The models are specified as discrete space-time dynamical systems, in the velocity space, that are characterized by an unforced or "free" response, given by the structure of network interconnections, and a forced response related to the contingent local optic flow information in input. In this way, given a motion information represented by an optic flow field extracted by a "classical" algorithm, we recognize if a group of velocity vectors relates to a specific motion pattern, on the basis of their spatial relationships in a local neighborhood. More precisely, the analysis/detection occurs through a spatial recurrent filter that checks the consistency between the spatial structural properties of the input flow field pattern and a set of linear models representing (first-order) elementary components of the optic flow [11]. In order to design a filter that checks this consistency, in an adaptive way, the linear models can be considered the process equations of a multiple model Kalman Filter (KF). Motion segments emerge from the noisy flows as the output of the KF that compares its prediction to the actual observations of the local properties of the optic flow.

Many works in the literature make use of the Kalman Filter for motion estimation. It has been used to estimate kinematic parameters (rotational and translational velocities and acceleration) of three-dimensional features [12] or to track 2D features through a sequence [13]. In [14] affine motion models are used to perform a region-based tracking in long image sequences and a standard Kalman Filter generates recursive estimation of each motion parameter. The novelty of the approach presented in this paper is in the def-

inition of models, which describe the optic flow and not the motion in the 3D space.

2 Linear models

Motion flow fields usually consist of large patches of flow-patterns, which result from a common cause (e.g., from ego-motion or object motion). These flow-patterns can be characterized on the basis of their first-order (linear) differential properties. From this perspective, local spatial features around a given location of a flow field can be of two types [11]: (1) the average flow velocity at that location, and (2) the structure of local variation in the neighborhood. The former relates to the *smoothness constraint* or *structural uniformity*, the latter refers to the *linearity constraint* or *structural gradients*. Velocity gradients provide important information about the 3-D layout of the visual scene.

Formally, the velocity gradient tensor can be written as follows:

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \begin{bmatrix} \partial v_x / \partial x & \partial v_x / \partial y \\ \partial v_y / \partial x & \partial v_y / \partial y \end{bmatrix}. \quad (1)$$

If we consider a point $\mathbf{x} = (x, y)$ in the spatial image domain, the linear properties of a motion field $\mathbf{v}(x, y) = (v_x, v_y)$ around the point $\mathbf{x}_0 = (x_0, y_0)$ can be characterized by a first-order Taylor expansion:

$$\mathbf{v} = \bar{\mathbf{v}} + \bar{\mathbf{T}}\mathbf{x} = \bar{\mathbf{v}} + \begin{bmatrix} \bar{T}_{11} & \bar{T}_{12} \\ \bar{T}_{21} & \bar{T}_{22} \end{bmatrix} \mathbf{x} \quad (2)$$

where $\bar{\mathbf{v}} = \mathbf{v}(x_0, y_0) = (\bar{v}_x, \bar{v}_y)$ and $\bar{\mathbf{T}} = \mathbf{T}|_{\mathbf{x}_0}$. By breaking down the tensor in its dyadic components, the motion field can be locally described through two-dimensional maps representing elementary flow components (EFCs) and Eq. (2) can be written as:

$$\mathbf{v} = \alpha^x \bar{v}_x + \alpha^y \bar{v}_y + \mathbf{d}_x^x \bar{T}_{11} + \mathbf{d}_y^x \bar{T}_{12} + \mathbf{d}_x^y \bar{T}_{21} + \mathbf{d}_y^y \bar{T}_{22} \quad (3)$$

where α^i are pure translations:

$$\alpha^x : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \alpha^y : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and \mathbf{d}_j^i are cardinal deformations:

$$\mathbf{d}_x^x : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x \\ 0 \end{bmatrix}, \quad \mathbf{d}_y^x : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} y \\ 0 \end{bmatrix}$$

$$\mathbf{d}_x^y : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ x \end{bmatrix}, \quad \mathbf{d}_y^y : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} 0 \\ y \end{bmatrix}.$$

The components of pure translations α^i can be incorporated in the corresponding deformations components, thus obtaining *generalized deformation components*:

$$\begin{aligned} \mathbf{v}_x^x &= a_1 \alpha^x + a_2 \mathbf{d}_x^x \triangleq \mathbf{m}_1 \\ \mathbf{v}_y^x &= a_3 \alpha^x + a_4 \mathbf{d}_y^x \triangleq \mathbf{m}_2 \\ \mathbf{v}_x^y &= a_5 \alpha^y + a_6 \mathbf{d}_x^y \triangleq \mathbf{m}_3 \\ \mathbf{v}_y^y &= a_7 \alpha^y + a_8 \mathbf{d}_y^y \triangleq \mathbf{m}_4 \end{aligned} \quad (4)$$

In this way, we have four classes of deformation gradients: one stretching (\mathbf{v}_i^i) and one shearing (\mathbf{v}_j^i) for each cardinal direction. As it will be clear in the following, this choice gives to the model maximum flexibility: every gradient deformation within a single class will be built through the same recurrent network, just by changing its driving inputs on the basis of direct local measures on the input optic flow. Figure 1 shows the four classes of deformation gradients.

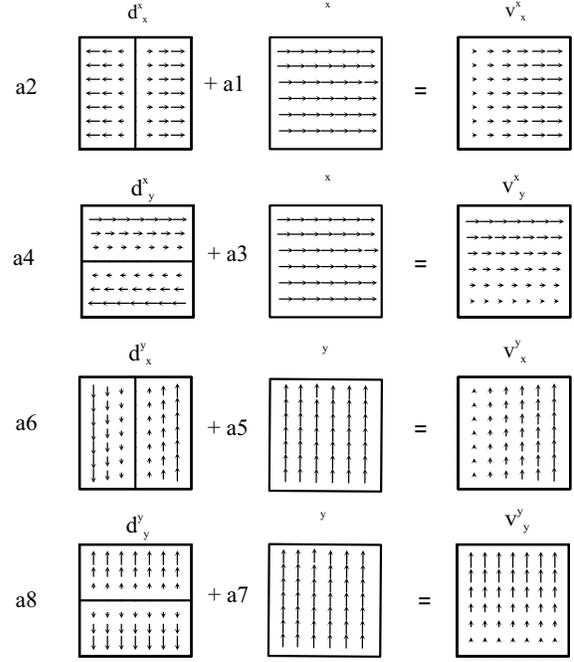


Figure 1. The generalized deformation components (\mathbf{v}_x^x , \mathbf{v}_y^x , \mathbf{v}_x^y , \mathbf{v}_y^y) are obtained by incorporating the pure translations in the corresponding cardinal deformations.

It is worthy to note that Eqs. (3) and (4) describe, in fact, an affine model:

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_4 \end{bmatrix} + \begin{bmatrix} c_2 & c_3 \\ c_5 & c_6 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (5)$$

where c_i are constants and v_x and v_y are the horizontal and the vertical components of the flow. The parameter vector $[c_1, c_2, \dots, c_6]$ describes a specific configuration of optic flow that locally provides a good approximation of 3D rigid moving objects. The six parameter affine model is reasonable to describe the motions of smooth surface in small image regions. The affine model is not sufficient to describe non rigid motion, like the motion of a human face [6]. However the motion within small patches can be still approximated by a first-order model. The relationships between these patches will describe the global motion of the face. The parameters c_i have qualitative interpretations in terms of image motion, for example c_1 and c_4 represent horizontal and vertical translation and we can express divergence (isotropic expansion), curl (rotation about the viewing direction), and the two components of shear defor-

mation (squashing, def_1 , or stretching, def_2) as combination of the c_i 's:

$$\begin{aligned} div &= c_2 + c_6 \\ curl &= c_3 - c_5 \\ def_1 &= c_3 + c_5 \\ def_2 &= c_2 - c_6 \end{aligned} \quad (6)$$

3 Kalman filtering

The problem of evidencing the presence of a certain complex pattern in the optic flow is posed as an adaptive filtering problem. The Kalman Filter is an optimal recursive adaptive filter [15], in the sense that it can iteratively process new measures as they arrive, on the basis of the knowledge about the system obtained by previous measurements. Kalman filtering is an optimal estimator if noise is independent, zero-mean and normally distributed. The output of the filter will be the *a posteriori* estimate of motion field improved by the additional (contextual) information provided by Kalman innovation.

Kalman filtering needs a *measurement equation* and a *process equation*. Formally we can write the following *measurement equation*:

$$\mathbf{v}[k] = \mathbf{C}[k]\mathbf{v}[k] + \mathbf{n}_1[k] \quad (7)$$

where $\mathbf{v}[k]$ is the optic flow at current time k , an intensity-based measure of the actual velocity field $v[k]$ and $\mathbf{n}_1[k]$ models the uncertainty of the algorithm. The linear operator \mathbf{C} represents a general "early-vision filter" providing a noisy measure of an observable property of the visual stimulus.

The *process equation* models the temporal evolution, from the previous step $k - 1$ to the current time k , of the relationships among visual features over a fixed spatial region, according to specific rules embedded in the transition matrix Φ :

$$\mathbf{v}[k] = \Phi[k, k - 1]\mathbf{v}[k - 1] + \mathbf{n}_2[k - 1] + \mathbf{s}[k - 1] \quad (8)$$

where $\mathbf{s}[k]$ is a driving input that can be interpreted as the boundary conditions of a lattice network (see Figure 2) and $\mathbf{n}_2[k]$ represents the process noise. Matrix Φ together with driving inputs $\mathbf{s}[k]$ implements a specific linear deformation component (see Eq. (4)). More precisely, this matrix models space-invariant nearest neighbor interactions within a finite region Ω in the image plane.

The driving input $\mathbf{s}[k]$ is evaluated at each step, by computing the average of optic flow velocity components at the boundary. So, the four models are adapted to the measures continuously. The spatial interactions occur separately for each component of the velocity vectors through anisotropic nearest neighbor interconnection

schemes. Specifically, for the x component we have:

$$\begin{aligned} v_x(i, j)[k] &= w_N^x v_x(i, j - 1)[k - 1] + \\ &+ w_S^x v_x(i, j + 1)[k - 1] + \\ &+ w_W^x v_x(i - 1, j)[k - 1] + \\ &+ w_E^x v_x(i + 1, j)[k - 1] + \\ &+ w_T^x v_x(i, j)[k - 1] + \\ &+ n_2^x(i, j)[k - 1] + \\ &+ s_x(i, j) \end{aligned} \quad (9)$$

and the same equation applies for v_y . The resulting pattern depends on the anisotropy of the interaction scheme and on the boundary conditions. By example, considering, for the sake of simplicity, a rectangular domain $\Omega = [-L, L] \times [-L, L]$, the EFC \mathbf{m}_1 can be obtained through:

$$\begin{aligned} w_T^x &= 0.1 & w_N^y &= w_S^y = 0 \\ w_N^x &= w_S^x = 0 & w_W^y &= w_E^y = 0 \\ w_W^x &= w_E^x = 0.45 & & \end{aligned} \quad s_y(i, j) = 0$$

$$s_x(i, j) = \begin{cases} \lambda & \text{if } i = -L \\ \mu & \text{if } i = L \\ 0 & \text{otherwise} \end{cases}$$

where the boundary values λ and μ are related to the coefficients c_1 and c_2 , and control the gradient slope and the constant term. In a similar way we can obtain the other components (see Figure 2). In this way, all the structural constraints necessary to model the continuum of linear deformations are embedded in the lattice interconnection scheme of the process equation. The resulting lattice network has a *structuring effect* constrained by the boundary conditions that yields to structural equilibrium configurations, characterized by the specific first-order EFCs that properly describe the input flow.

To describe the Kalman filtering processing, we define $\hat{\mathbf{v}}[k|\mathbf{V}_{k-1}]$ as the *a priori* state estimate at step k , given the knowledge of the process at step $k - 1$, and $\hat{\mathbf{v}}[k|\mathbf{V}_k]$ as the *a posteriori* state estimate at step k given the measurement at step k . \mathbf{V}_{k-1} and \mathbf{V}_k represent all the measurements until step $k - 1$ and k respectively, the aim of the filter is to compute an *a posteriori* estimate starting from the *a priori* estimate and from the weighted difference between the current and the predicted measurement:

$$\hat{\mathbf{v}}[k|\mathbf{V}_k] = \hat{\mathbf{v}}[k|\mathbf{V}_{k-1}] + \mathbf{G}[k](\mathbf{v}[k] - \hat{\mathbf{v}}[k|\mathbf{V}_{k-1}]). \quad (10)$$

The difference term $\mathbf{v}[k] - \hat{\mathbf{v}}[k|\mathbf{V}_{k-1}]$ is the *innovation* $\mathbf{v}[k]$, while the matrix $\mathbf{G}[k]$ is the Kalman gain that minimizes the *a posteriori* error covariance:

$$\mathbf{K}[k] = E\{(\mathbf{v}[k] - \hat{\mathbf{v}}[k|\mathbf{V}_k])(\mathbf{v}[k] - \hat{\mathbf{v}}[k|\mathbf{V}_k])^T\}. \quad (11)$$

The covariance matrix $\mathbf{K}[k]$ provides us only information about the properties of convergence of the Kalman Filter and not whether it converges to the correct values. Hence, we have to measure the discrepancy between predictions and observations in statistical terms, as an indication of the filter's consistency. A frequently used quantitative measure

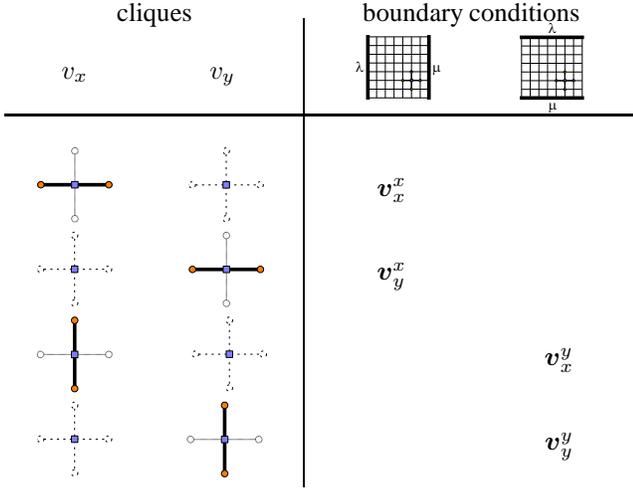


Figure 2. Basic lattice interconnection schemes for the elementary flow components considered. By a proper choice of the interconnection weights and of the boundary values λ and μ the velocity profiles result approximately linear.

of consistency is the Normalized Innovation Squared (NIS) [16]:

$$NIS_k = \nu^T[k] \mathbf{S}^{-1}[k] \nu[k] \quad (12)$$

where \mathbf{S} is the covariance of the innovation. In our model the NIS value is used to compute the likelihood of the measurement.

3.1 Multiple model approach

The structure of the optic flow approximation leads to a multiple model adaptive estimator: we can use a bank of parallel Kalman Filters, each with a different model (a different process equation embedding a generalized deformation component). We need a dynamic multiple model approach because the choice between the four possible models varies continuously while the filter is operating. In such a case, we cannot make a fixed *a priori* choice of the filter parameters, but we need to use a continuously varying model-conditioned combination of the candidate state and error covariance estimates. It is worthy to note that, in the dynamic multiple model approach, we do not want the probabilities to converge to fixed values, but we want them free to change at each new measurement. In the multiple model approach [16] [17] it is assumed that the system obeys one of a finite number of models. Thus, we must assume that the correct model m is one among all the possible models m_i with $i = 1, 2, \dots, r$.

The likelihood of the measurement \mathbf{v} given a particular model m_i at time step k is given by:

$$f(\mathbf{v}|m_i) = |\mathbf{2}\pi\mathbf{S}_{m_i}|^{-\frac{1}{2}} e^{-\frac{1}{2}\nu_{m_i}^T \mathbf{S}_{m_i}^{-1} \nu_{m_i}} \quad (13)$$

where m_i is the considered model. The probability that the candidate model m_i is the correct one is given by the

following equation:

$$p_{m_i}[k] = \frac{f(\mathbf{v}|m_i)}{\sum_{j=1}^r f(\mathbf{v}|m_j)} \quad (14)$$

with $p_{m_i}[0] = 1/r$, $i = 1, 2, \dots, r$ and $\sum_{i=1}^r p_{m_i}[k] = 1$ at each time step k . The final model-conditioned estimate of the state \mathbf{v} is computed as a weighted combination of the *a posteriori* states of each candidate filter:

$$\hat{\mathbf{v}}[k] = \sum_{i=1}^r p_{m_i}[k] \hat{\mathbf{v}}_{m_i}[k]. \quad (15)$$

For the 4 models considered (see Eqs. (4)):

$$\hat{\mathbf{v}} = p_{m_1} \hat{v}_x^x + p_{m_2} \hat{v}_y^x + p_{m_3} \hat{v}_x^y + p_{m_4} \hat{v}_y^y \quad (16)$$

where $p_{m_1}, p_{m_2}, p_{m_3}, p_{m_4}$ are the probabilities related to each model and $\hat{v}_x^x, \hat{v}_y^x, \hat{v}_x^y, \hat{v}_y^y$ are the state estimates for each Kalman filter.

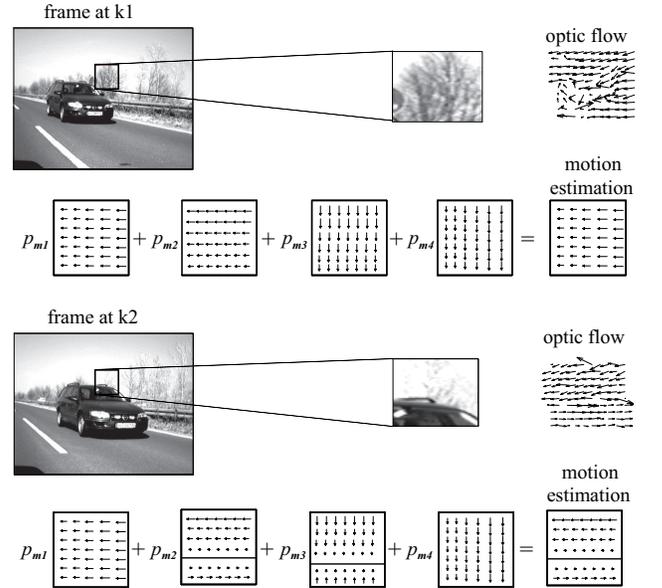


Figure 3. Multiple-model motion estimation in a road scene taken by a rear-view mirror of a moving car under an overtaking situations. The model-based decompositions are evidenced for the same image patch for two different frames at time $k1$ and $k2$. For each optic flow patch the motion is estimated from the actual generalized deformation components weighed by the corresponding probability values, see Eq.(15).

Combining Eqs. (4) and (16) we have:

$$\begin{bmatrix} \hat{v}_x \\ \hat{v}_y \end{bmatrix} = \begin{bmatrix} p_{m_1} \hat{a}_1 + p_{m_2} \hat{a}_3 \\ p_{m_3} \hat{a}_5 + p_{m_4} \hat{a}_7 \end{bmatrix} + \begin{bmatrix} p_{m_1} \hat{a}_2 & p_{m_2} \hat{a}_4 \\ p_{m_3} \hat{a}_6 & p_{m_4} \hat{a}_8 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (17)$$

from which it is possible to derive the estimated coefficients of the affine model:

$$\begin{aligned} \hat{c}_1 &= p_{m_1} \hat{a}_1 + p_{m_2} \hat{a}_3, & \hat{c}_2 &= p_{m_1} \hat{a}_2, & \hat{c}_3 &= p_{m_2} \hat{a}_4 \\ \hat{c}_4 &= p_{m_3} \hat{a}_5 + p_{m_4} \hat{a}_7, & \hat{c}_5 &= p_{m_3} \hat{a}_6, & \hat{c}_6 &= p_{m_4} \hat{a}_8 \end{aligned} \quad (18)$$

Figure 3 shows how the multiple model approach is used to estimate the presence of the different generalized deformation components in the optic flow. First, the deformation components are adapted accordingly with the optic flow values in input, then a probability value is associated to each component and the final estimate is evaluated by the weighted sum of the single components, see Eq.(15).

4 Results

To assess the performances of the approach, we applied recursive Kalman filtering to optic flows related to both real-world driving sequences and facial expressions. A “classical” algorithm [18] has been used to extract the optic flows.

If we analyze a multiple motion sequence we expect that objects in the background have the same divergence values, whereas other objects moving in the scene will have a different divergence. Therefore, by mapping the sum of c_2 and c_6 , we are able to obtain a good segmentation of the objects in the scene. Figures 4 and 5 show examples of multiple motion segmentation using divergence information for different real-world traffic scenes.

Figure 6 shows how this approach can be used to analyse different areas of optic flow in a complex motion sequence like a facial expression. If we consider the values of the affine model coefficients we are able to describe the motions of the different areas of the face. In the figure five different areas of the face have been chosen and the coefficients of the affine models have been computed and plotted as a function of time. The relationships between the temporal behaviour of these values and their spatial positions could describe quite well the face motion.

5 Conclusions

The problem of evidencing the presence of a certain complex feature in the optic flow is an important step towards motion segmentation. We have shown that it is possible to solve this problem on the basis of both direct input and contextual information, by recurrent adaptive filtering of the optic flow. Direct information comes from the input measures and the context from reference signals, represented as a set of specific linear models. Kalman-filter based techniques to switch between models have been known for some time in the control literature [16]. Here, we propose a similar approach to permit multiple linear models as multiple competing hypotheses. Accordingly, the multi-model Kalman Filter yields the optimal estimates of the weights of the adjustable linear models. A great potential advantage of the multiple-model approach is that recognition and feature extraction can be performed jointly, and so the form

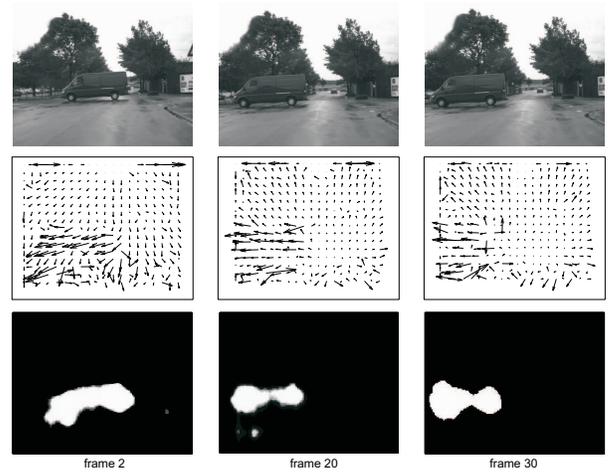


Figure 4. Example of multiple motion segmentation. The camera is moving towards the van that is crossing the street.

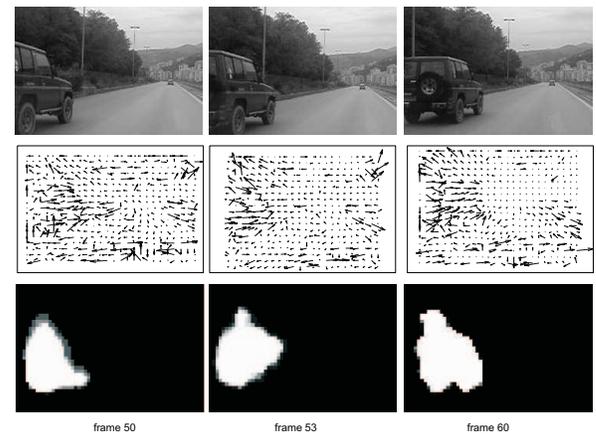


Figure 5. Example of multiple motion segmentation. The camera and the car are moving along the same direction.

of the expected linear component can be used to guide feature search, potentially making it more efficient and robust. The use of linear models to analyze image motion has been previously investigated in [7], where the authors proposed the use of parameterized motion models to represent complex motions. In that paper, they adopted both linear and learned basis flow fields to describe the motion of large portions of the face. Here, by considering small areas of facial expression, we are able to approximate image flows with linear models. A systematic comparison between the two approaches will be tackled in a future work.

Acknowledgements

This work has been partially supported by the EU Project NEST-2003-12963 “Multi-channel cooperativity in visual processing (MCCOOP)” and by the EU Project IST-2003-016276-2 “Learning to emulate perception-action cycles in

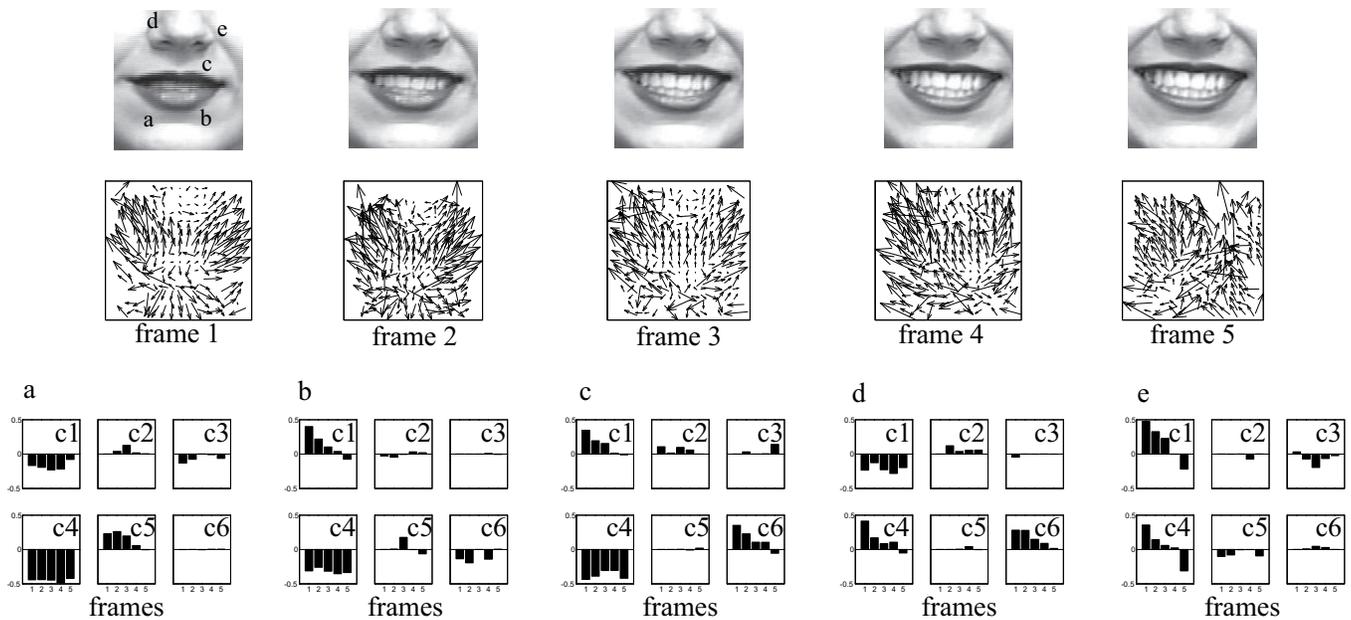


Figure 6. Top: five frames from a facial expression sequence. On the first frame the letters indicate the positions of the analysed areas. Middle: Optic flows computed from the sequence. Bottom: temporal evolutions of the six affine model coefficients for the five selected positions (a-e).

a driving-school scenario (DRIVSCO)”.

References

- [1] M.J. Black, “Explaining optical flow events with parameterized spatio-temporal models”, *CVPR*, 1999.
- [2] Y. Rui and P. Anandan, “Segmenting visual actions based on spatio-temporal motion patterns”, *CVPR*, 2000.
- [3] J. Davis and A. Bobick, “The representation and recognition of action using temporal templates.”, *CVPR*, 1997.
- [4] Y. Yacoob and M.J. Black, “Parameterized modeling and recognition of activities”, *Computer Vision and Image Understanding*, vol. 73, pp. 232–247, 1999.
- [5] T. Darrell and A. Pentland, “Space-time gestures”, *CVPR*, 1993.
- [6] M.J. Black and Y. Yacoob, “Recognizing facial expression in image sequences using local parameterized models of image motion”, *International Journal of Computer Vision*, vol. 25, pp. 23–48, 1997.
- [7] D.J. Fleet, M.J. Black, Y. Yacoob, and A.D. Jepson, “Design and use of linear models for image motion analysis”, *International Journal of Computer Vision*, vol. 36, pp. 171–193, 2000.
- [8] M.J. Black, Y. Yacoob, A.D. Jepson, and D.J. Fleet, “Learning parameterized models of image motion”, *CVPR*, 1997.
- [9] Y. Yacoob and L. Davis, “Learned temporal models of image motions.”, *International Journal of Computer Vision*, pp. 446–453, 1998.
- [10] M.J. Black and P. Anandan, “The robust estimation of multiple motion: Parametric and piecewise-smooth flow fields”, *Computer Vision and Image Understanding*, vol. 63, pp. 75–104, 1996.
- [11] J.J. Koenderink, “Optic flow”, *Vision Res.*, vol. 26, pp. 161–179, 1986.
- [12] Z. Zhang and O. D. Faugeras, “Three-dimensional motion computation and object segmentation in a long sequence of stereo frames.”, *International Journal of Computer Vision*, vol. 7, pp. 211–241, 1992.
- [13] S. M. Smith and J. M. Brady, “Asset-2: Real-time motion segmentation and shape tracking.”, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 814–820, 1995.
- [14] F. G. Meyer and P. Bouthemy, “Region-based tracking using affine motion models in long image sequences.”, *CVGIP: Image Understanding*, vol. 60, pp. 119–140, 1994.
- [15] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall International Editions, 1991.
- [16] Y. Bar-Shalom and X.R. Li, *Estimation and Tracking, Principles, Techniques, and Software*, Artech House, 1993.
- [17] G. Welch and G. Bishop, “An introduction to the kalman filter”, in *SIGGRAPH 2001*, Los Angeles, USA, 2001.
- [18] B. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision”, *Proc. DARPA Image Understanding Workshop*, pp. 121–130, 1981.