Motion Interpretation Using Adjustable Linear Models

Manuela Chessa, Fabio Solari, Silvio P. Sabatini, Giacomo M. Bisio
Department of Biophysical and Electronic Engineering
University of Genoa
manuela.chessa@unige.it

Abstract

A method to analyze first-order spatial properties of optical flow is proposed. The approach is based on the use of a set of linear models that dynamically adjust their properties on the basis of context information. These models are generated by a recursive network that takes into account spatial interaction between neighbors. By checking the presence of these models in the optic flow using a multiple model Kalman Filter it is possible to recover information about the coefficients of the affine description and the image motion invariants: divergence, curl and deformation. Reliable estimates of these quantities could help in the analysis of real world complex motion sequences. Experimental results on egomotion estimation and 3D surface reconstruction validate the approach.

1 Introduction

The analysis and interpretation of visual motion is a challenging problem in computer vision. Such interpretation aims to relate motion events in the 3D space to global spatiotemporal variations of the image (i.e., the image flow) for gaining useful information for different application domains, such as autonomous navigation, robot manipulation tasks, and 3D dynamic scene understanding. By adopting a hierarchical approach, we can resort, at least at a conceptual level, to an intermediate representation of the distribution of the local velocities (i.e. the motion field). This approach models the functional organization of the cortical motion stream of mammals [13].

At a first approximation, and under proper conditions [10], important information about egomotion, time-to-collision and the 3D layout of the scene can be obtained by looking at the spatial first-order differential properties of the motion field. Many different approaches have been proposed in the literature to recover reliable estimates of these differential properties. Cipolla and Blake [7] use B-spline snakes to track the change in the apparent area of scene object to approximate the differential invariants (divergence and deformation) of the motion field. Other authors estimate the affine motion parameters by robust maximum-likelihood estimate technique [15], or directly from the spatio-temporal derivatives of the image intensity [8]. Other approaches work on optic flow. Nelson and Aloimonos [12] use divergence for obstacle avoidance by deriving it analytically. This approach needs the integration of many results over time to produce stable results. Ancona and Poggio [1] use sparse estimates of optic flow to get information about time to collision. Fu and Kovesi [9] propose the use of a bank of filters for recovering the differential invariants from a dense optic flow with a correlation technique. This approach requires a large number of filters to obtain reliable quantitative results in real-world sequences. In summary, one can use techniques that either analyze the deformation of the image or work on the optic flow. The latter poses stability problems when one directly computes the spatial derivatives or require an high computational cost when one adopts matched filters approaches. Recently Chessa et al. [6] proposed a method for designing adjustable linear models for the analysis of first order properties of complex dense optic flow fields. These models make use of contextual information by capturing coherent linear properties and regularities over small image patches. The linear 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, we combine the advantages of the differential linear models with those of template matching since quantitative measures of first-order differentials can be obtained by a small number of templates.

In this paper, we aim to systematically validate the approach proposed in [6] by analyzing the reliability of the first-order optic flow measures obtained by the templates, and their perceptual significance by using them directly for dynamic scene interpretation.

2 Adjustable linear templates

Within any small image region, and under smooth change in viewpoint [10], an affine model of image motion

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c_5 \\ c_6 \end{bmatrix} \tag{1}$$

is often sufficient to locally provide a good approximation of 3D rigid moving objects and information about the 3D layout of the scene. The parameters c_i have qualitative interpretations in terms of the spatial variations of the associated velocity field $\boldsymbol{v}(x,y) = [v_x(x,y), v_y(x,y)]$. Formally, the parameters c_5 and c_6 represent the horizontal ($\bar{v}_x$) and vertical ($\bar{v}_y$) translational velocities in the image patch, respectively; whereas the parameters c_1, c_2, c_3, and c_4 represent the values of the coefficients of the velocity tensor:

$$\bar{\mathbf{T}} = \mathbf{T}|_{\mathbf{x_0}} = \begin{bmatrix} \frac{\partial v_x}{\partial x} & \frac{\partial v_x}{\partial y} \\ \frac{\partial v_y}{\partial x} & \frac{\partial v_y}{\partial y} \end{bmatrix}_{\mathbf{x}=\mathbf{x_0}} \tag{2}$$

of a first-order Taylor expansion calculated around the image point $\mathbf{x_0} = (x_0, y_0)$:

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} \bar{T}_{11} & \bar{T}_{12} \\ \bar{T}_{21} & \bar{T}_{22} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \bar{v}_x \\ \bar{v}_y \end{bmatrix}. \tag{3}$$

Equivalently, the differential invariants of image motion can be related to algebraic combinations of the affine coefficients: $div\boldsymbol{v} = c_1 + c_4$, $curl\boldsymbol{v} = c_2 - c_3$, $(def\boldsymbol{v})\cos 2\theta = c_1 - c_4$ and $(def\boldsymbol{v})\sin 2\theta = c_2 + c_3$. By comparing Eq. (1) and Eq. (3) and by breaking down the tensor in its dyadic components, the motion field can be locally described through 2D maps representing elementary flow components (EFCs):

$$\boldsymbol{v} = c_1 \boldsymbol{d}_x^x + c_2 \boldsymbol{d}_y^x + c_3 \boldsymbol{d}_x^y + c_4 \boldsymbol{d}_y^y + c_5 \boldsymbol{\alpha}^x + c_6 \boldsymbol{\alpha}^y \tag{4}$$

where $\boldsymbol{\alpha}^x : (x, y) \mapsto (1, 0)$, $\boldsymbol{\alpha}^y : (x, y) \mapsto (0, 1)$ are pure translations and $\boldsymbol{d}_x^x : (x, y) \mapsto (x, 0)$, $\boldsymbol{d}_y^x : (x, y) \mapsto (y, 0)$, $\boldsymbol{d}_x^y : (x, y) \mapsto (0, x)$, $\boldsymbol{d}_y^y : (x, y) \mapsto (0, y)$ represent cardinal deformations, basis of a linear deformation space.

It is worth noting that by distributing the pure translations and incorporating the coefficients in the deformation components, the velocity field can be described by four models of *generalized* deformations that act as adjustable linear templates parametrized by the coefficients a_i and c_i: $\boldsymbol{v}_x^x : (x, y) \mapsto (c_1 x + a_1, 0)$, $\boldsymbol{v}_y^x : (x, y) \mapsto (c_2 y + a_2, 0)$, $\boldsymbol{d}_x^y : (x, y) \mapsto (0, c_3 x + a_3)$, $\boldsymbol{d}_y^y : (x, y) \mapsto (0, c_4 y + a_4)$. In this way, we have four classes of deformation gradients: one stretching ($\boldsymbol{v}_i^i$) and one shearing ($\boldsymbol{v}_j^i$) for each of the two cardinal directions, which generate uniform samples of the linear deformation space. Due to their ability to detect the presence and the orientation of velocity gradients and kinetic boundaries, as well as large field motion invariants, these resulting templates resemble the receptive fields of the cells in the extrastriate cortical areas [13].

With reference to the Taylor expansion, it is worth noting that a template based approach cannot be used to extract single components, but to perform pattern matching operations, only. Hence, in general, to proper sampling the linear deformation space one has to use a large number of templates with very different structural properties. The introduction of the adjustability in our model allows us to reduce to only four the number of templates. In this way, we will be able to “measure” the linear properties of the motion field without performing direct differential operations, but by reading out the values of the adjusted coefficients of the templates.

2.1 Generative models

In [6] the authors demonstrated that each template that locally approximates a generalized deformation components can be generated recursively by using a lattice network:

$$\boldsymbol{v}[k] = \boldsymbol{\Phi}[k, k-1]\boldsymbol{v}[k-1] + \boldsymbol{n}_2[k-1] + \boldsymbol{s}[k-1] \ , \tag{5}$$

which describes the temporal evolution, from the previous time step $k-1$ to the current time step k, of the relationships among motion features over a fixed small spatial region $[-L, L] \times [-L, L]$ according to specific rules embedded in the transition matrix$\boldsymbol{\Phi}$. The driving input $\boldsymbol{s}[k]$, evaluated at each time step, by computing the average of the optic flow velocity components at the patch’s borders, can be interpreted as the boundary conditions of the lattice network (see Fig. 1), whereas $\boldsymbol{n}_2[k]$ represents the process noise.

It is worth noting that the spatial interactions occur separately for each component of the velocity vectors through 1D nearest neighbor interactions. More precisely, given the difference equation that describes the nearest neighbor cooperation among the spatial nodes n’s for the generic velocity component v: $A_{-1}v(n-1)+A_0 v(n)+A_1 v(n+1) = 0$, and solving it with the boundary conditions $v(-L) = \lambda$ and $v(L) = \mu$, we obtain the velocity profiles that approximate the linear templates parametrized by the coefficients a_i and c_i:

$$v(n) = \frac{\mathrm{e}^{-\alpha \mathrm{L}}}{1 - \mathrm{e}^{-4\alpha \mathrm{L}}} \left[(\lambda - \mu \mathrm{e}^{-2\alpha \mathrm{L}}) \mathrm{e}^{-\alpha \mathrm{n}} + (\mu - \lambda \mathrm{e}^{-2\alpha \mathrm{L}}) \mathrm{e}^{\alpha \mathrm{n}} \right] \tag{6}$$

where $\lambda = a_i - Lc_i$ and $\mu = a_i + Lc_i$, and with α depending on the coupling coefficient $A_1 = A_{-1}$ of the 1D lattice network. By a proper choice of the coupling coefficients and of the boundary values λ and μ the velocity profiles result approximately linear. To quantify the approximation error, we calculated, as a function of α and L, the average

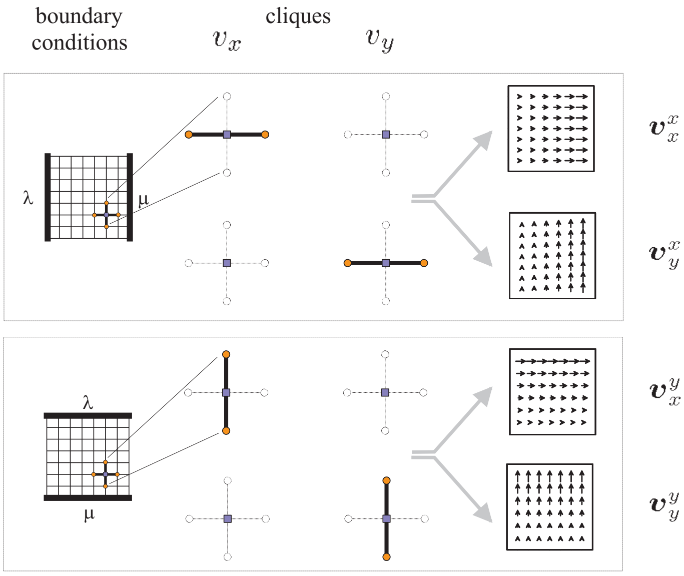


Figure 1: Basic lattice interconnection schemes for the generation of the adjustable linear templates. The lattice networks have a *structuring effect* constrained by the boundary conditions that yields to structural equilibrium configurations, characterized by the specific first-order EFCs. The resulting velocity patterns depend on the directions of the interaction scheme and on the boundary conditions. $\boldsymbol{v}_x^x$ and $\boldsymbol{v}_y^y$ represent the stretching components, whereas $\boldsymbol{v}_y^x$ and $\boldsymbol{v}_x^y$ represent the shearing components. The boundary values λ and μ control the gradient slope and the constant term.

relative integral error between the solution of the lattice network (see Eq. 6) and a straight line that joins the values at the boundaries (λ and μ). Figure 2(a) shows the curves of constant error ($\epsilon = 0.01$), for different combinations of λ and μ. Figure 2(b) shows the variability of the approximation error by varying the boundary values λ and μ for a fixed size of the template ($L = 3$) and for a fixed value of $\alpha = 0.2$. The limited increase of the error over a wide variation of the boundary values in the range of ± 30 pixel/frame demonstrates the validity of the approximation of the linear templates by the generative models.

2.2 Recursive/adaptive filtering

The adjustable templates defined in the previous Section can be used as models for a multiple model Kalman filter (KF) to measure the structural properties of the input optic flow. The output of the KF will be the estimate of the motion field on the basis of the spatial contextual information described by the generative models of the EFCs. Since the models are continuously adapted to the measures by changing the boundary conditions for every patch, and the KF iteratively integrates the new measures with the knowledge about the motion pattern obtained by the previous measurements, we obtain adaptive estimates of the EFCs. In this way, we perform an adaptive template matching capable of tracking

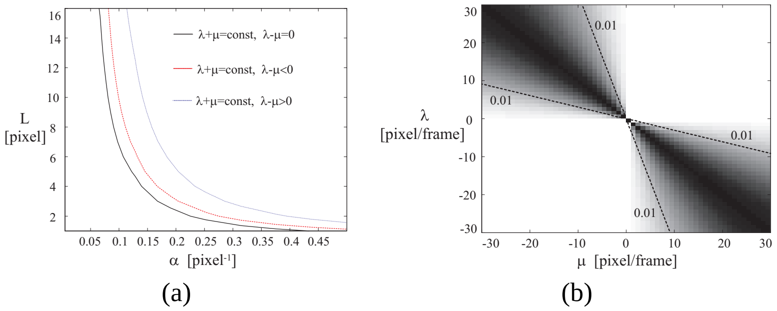


Figure 2: Variations of the approximation error for different values of the network parameters. (a) Relationships between the size of the patch L and the diffusive coefficient of the lattice network α for a constant value of the approximation error ($\epsilon = 0.01$). (b) Variation of the error for the pair $L = 3$ and $\alpha = 0.2$ over a variation of the boundary values.

the coefficients of a linear description/approximation of the optic flow.

Formally, the measurement equation is $\mathbf{v}[k] = \mathbf{C}[k]\boldsymbol{v}[k]+\boldsymbol{n}_1[k]$, where $\mathbf{v}[k]$ is a noisy measure, at current time k, of the actual motion field $\boldsymbol{v}[k]$, $\boldsymbol{n}_1[k]$ is the uncertainty of the measure, and $\mathbf{C}$ is a modified unitary operator for discarding the image points where the optic flow is not available or not reliable. The output of the filter is:

$$\hat{\boldsymbol{v}}[k|\mathbf{V}_k] = \hat{\boldsymbol{v}}[k|\mathbf{V}_{k-1}] + \boldsymbol{G}[k]\boldsymbol{\nu}[k] \tag{7}$$

where: $\hat{\boldsymbol{v}}[k|\mathbf{V}_{k-1}]$ is the *a priori* state estimate, $\hat{\boldsymbol{v}}[k|\mathbf{V}_{\mathbf{k}}]$ is the *a posteriori* state estimate, $\mathbf{V}_k$ represent all the measurements until time steps k, $\boldsymbol{\nu}[k] = \mathbf{v}[k] - \hat{\mathbf{v}}[k|\mathbf{V}_{k-1}]$ is the innovation and $\boldsymbol{G}[k]$ is the Kalman gain. In order to have a statistical measure the discrepancy between predictions and observations, as an indication of the filter's consistency, it is frequently used the Normalized Innovation Squared (NIS) [2]:

$$\mathrm{NIS}_k = \boldsymbol{\nu}^T[k]\boldsymbol{S}^{-1}[k]\boldsymbol{\nu}[k] \tag{8}$$

defined on the basis of the innovation and on its covariance $\boldsymbol{S}$. Since the covariance of the innovation depends on the estimate of the measure noise $\boldsymbol{n}_1$, it is important to have a reliable estimate of the noise in the measure. Thus, the noise covariance matrices are tuned on the basis of the differences (in terms of the mean angular error [3]) of the velocity values measured inside a patch between two consecutive frames. Where the optic flow smoothly changes in time, the measure noise $\boldsymbol{n}_1$ remains low, whereas, where optic flow changes more abruptly, the noise becomes higher and the estimates have a lower confidence. In the multiple model KF the NIS value is used to compute, for each model, the likelihood of the measurements, on which to base the selection among the different models. This choice varies continuously while the filter is operating. In such a case, we cannot make a fixed *a priori* choice of the filter's parameters, but we have to use a continuously varying model-conditioned combination of the candidate state and error covariance estimates. It is worth noting that, in our dynamic multiple model approach, we do not want the probabilities to converge to fixed values, but we want them to be free to change at each new measurement. In the multiple model approach [2] it is assumed that the system obeys one of a finite number of models $\boldsymbol{m}_i$ with $i = 1, 2, \ldots, r$ (with $r = 4$, in our case, corresponding to the four classes of deformation gradients). The likelihood of the measurement $\mathbf{v}$ given a model $\boldsymbol{m}_i$ at time step k is given by:

$$f(\mathbf{v}|\boldsymbol{m}_\mathbf{i}) = |\mathbf{2}\pi \boldsymbol{S}_{\boldsymbol{m}_\mathbf{i}}|^{-\frac{1}{2}} \mathbf{e}^{-\frac{1}{2}\boldsymbol{\nu}^{\mathrm{T}}_{\boldsymbol{m}_\mathrm{i}} \boldsymbol{S}^{-1}_{\boldsymbol{m}_\mathrm{i}} \nu_{\boldsymbol{m}_\mathrm{i}}} \tag{9}$$

where $\boldsymbol{m}_i$ is the considered model. The probability that the candidate model $\boldsymbol{m}_i$ is the correct one is given by the following equation:

$$p_{\boldsymbol{m}_i}[k] = \frac{f(\mathbf{v}|\boldsymbol{m}_\mathbf{i})}{\sum_{j=1}^{r} f(\mathbf{v}|\boldsymbol{m}_\mathbf{j})}. \tag{10}$$

With this approach the probability value approaches 1 when the optic flow has the same structure of the model. None of the models gives a high probability value if none of the EFCs is present in the optic flow. In this way, noisy and unstructured motions are automatically discarded. Figure 3 shows the evolution in time of the four models related to an optic flow patch in the same position for different frames. The four models are continuously adjusted on the basis of the input optic flow and a probability value is associated to each model. We can observe through frames the behavior of each model for different motion situations: at frame 2, the patch contains the motion of the background, only; from frame 8 to frame 17, motion discontinuities appear in the models (e.g., kinetic edges) in correspondence of the passage of the motorbike; at frame 21, the patch contains the motion of the motorbike, only.

To quantitatively assess the reliability of the first-order differential measures and their robustness to noise, we calculated the error for synthetic optic flow patterns and we compared the results with the error obtained by a direct numerical differentiation of the noisy flow. We observe that the KF allows us to obtain correct estimates for high values of the noise, with an almost constant error below 0.07 (see Fig. 4).

3 Motion interpretation

The affine description of the optic flow are related to the motion of the observer $\boldsymbol{T} = (T_x, T_y, T_z)$ and $\boldsymbol{\Omega} = (\Omega_x, \Omega_y, \Omega_z)$ and to the depth gradient of the surfaces $\boldsymbol{F} = (p, q)$ in the following way [7]:

$$c_1 = \frac{T_z}{Z_0} + \frac{pT_x}{Z_0} \qquad c_2 = \omega_z + \frac{qT_x}{Z_0} \qquad c_4 = -\omega_z + \frac{pT_y}{Z_0} \qquad c_5 = \frac{T_z}{Z_0} + \frac{qT_y}{Z_0}. \tag{11}$$

There are many ways of proceeding to solve these equations for the unknown 3D parameters. One approach is to directly derive the optic flow measurements within a small region and solve for the parameters by minimizing an error function. The main problem of this approach is the instability of the numerical derivative, as we have shown in the previous section. An alternative is to fit a quadratic parametrization to the optic flow measurements to obtain the affine coefficients [8], than solving for the 3D parameters by a minimization. Still, the main problem could be the instability in the affine coefficients estimates.

In this work, we use a recursive adaptive approach to obtain estimates of the affine coefficients of the optic flow that are stable in time, then we solve the Eq. (11) by a

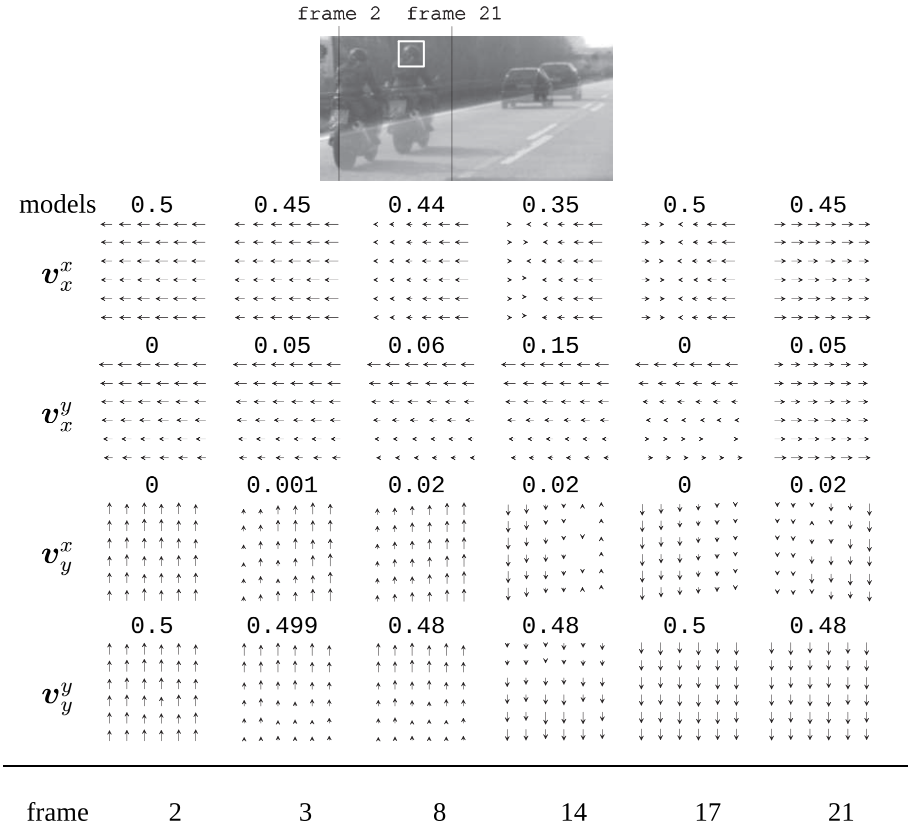
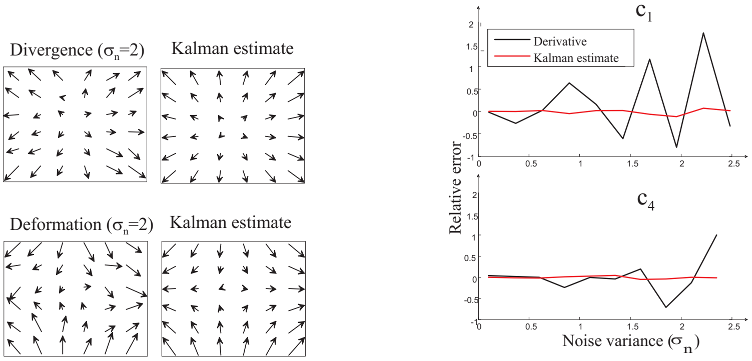


Figure 3: Evolution in time of the four optic flow models in the same image patch. The white square that localize the image patch is enlarged for the sake of representation. The sequence is acquired by a car moving on a highway: the independent motion of the motorbike superimposes to the self-motion of the car. The number on the top of each model indicates the associated probability.

minimization. Since we have 4 equation in 7 unknowns (T_x, T_y, T_z, Ω_z, p, q, Z_0,), by over-determining the system and by considering a sufficient number of points, it is possible to recover the 3D parameters, i.e. the motion parameters of the observer, from which it is possible to derive the heading direction, and the depth gradient of the surfaces, from which it is possible to derive the normal to the surface.

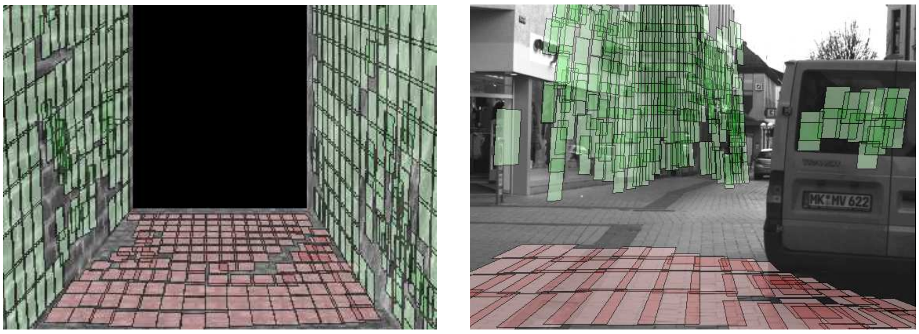
We have tested the proposed approach with both synthetic and real-world sequences recorded by a camera on a moving car. We apply the KF to the optic flow in order to obtain stable and reliable estimates of its linear properties, then we recover the 3D parameters. Figure 5 shows the estimates of the slant of the different surfaces in a virtual environment and in a real-world situation. The virtual scene is composed by 3 planes with different orientations: the two side walls have been rotated along the vertical axis clockwise and counter-clockwise respectively, the ground plane has been tilted toward the observer. The scene has been recorded by a virtual perspective camera moving both along the Z axis and along the X axis. In general with the proposed approach the major surfaces are correctly



(a) (b)

Figure 4: (a) Divergence and deformation components of an optic flow corrupted by a Gaussian noise with variance $\sigma_n = 2$ and the corresponding KF estimates. (b) Comparison between the relative error on c_1 and c_4 obtained by directly differentiating the optic flow (black lines) and by the recurrent adaptive templates (red lines).

estimated, as we quantitatively measured comparing the results with the ground truth in the virtual scenario. For the same sequences we have computed the egomotion (see Fig. 6). The white cross is the true value, derived from the known motion of the virtual camera and obtained from the can-bus data of the car.

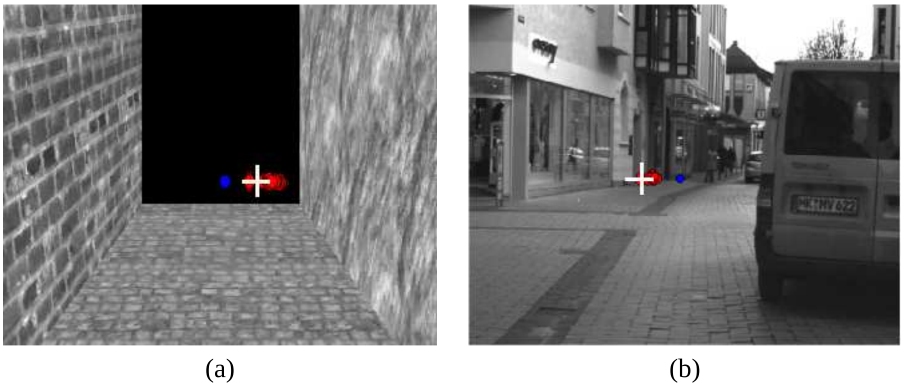


(a) (b)

Figure 5: Estimation of the slant of the surfaces. Each small patch superimposed on the original frame is build from an estimation of the normal to the surface. (a) Virtual environment. (b) Real-world scenario. It is worth noting that the patches on the building on the left correctly follow its shape. Missing data are due to unreliable estimates of the affine coefficients.

4 Conclusions

We have presented a recurrent technique to adaptively detect structural properties from the optic flow, by casting the problem as a KF based on multiple models of spatial variations



(a) (b)

Figure 6: Heading estimation. Red points are the estimates, the white cross is the true heading position and the blue dot is the center of the image. (a) Virtual environment. (b) Real-world scenario.

of velocity. We have shown that it is possible to recover information about first-order properties of optic flow in a reliable way by using a set of linear adjustable models that make use both of contextual information and direct inputs coming from the optic flow. 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. We have tested the approach with synthetic sequences to validate the estimation of the slant of the surfaces and the egomotion. Then we have applied it to several frames of real world sequences.

Many works in the literature make use of the KF for motion estimation. It has been used to estimate kinematic parameters (rotational and translational velocities and acceleration) of 3D features [16] or to track 2D features through a sequence [14]. In [11] affine motion models are used to perform a region-based tracking in long image sequences and a standard KF generates recursive estimation of each motion parameter. In [5] the author uses parametrization of the local flow fields to obtain the estimate of the affine coefficients and a recursive approach to solve for the 3D unknowns.

The novelty of the approach presented in this paper is in the definition of adjustable models, which describe the optic flow and not the motion in the 3D space. The presented results use real-world sequences containing rigid-body multiple motions. First-order analysis is not sufficient to describe non rigid motion and such kind of situations have been tackled in the literature by using models learned by example [8] or considering higher-order optic flow approximation [4]. Similarly, the work presented could be extended to include higher-order adjustable models to account for more complex contextual information.

Acknowledgements

This work has been partially supported by the EU Project IST-2003-016276-2 “DRIVSCO”, by the EU Project NEST-2003-12963 “MCCOOP” and by “Progetto di Ricerca di Ateneo 2007” (University of Genoa).

References

[1] N. Ancona and T. Poggio. Optical flow from 1-d correlation: Application to a simple time-to-crash detector. *Int. J. Computer Vision*, 14(2), 1995.

[2] Y. Bar-Shalom and X.R. Li. *Estimation and Tracking, Principles, Techniques, and Software*. Artech House, 1993.

[3] J.L. Barron, D.J. Fleet, and Beauchemin S.S. Performance of optical flow techniques. *Int. J. of Computer Vision*, 12:43–76, 1994.

[4] M.J. Black and Y. Yacoob. Recognizing facial expression in image sequences using local parameterized models of image motion. *Int. J. of Computer Vision*, 25(1):23–48, 1997.

[5] A. Calway. Recursive estimation of the 3d motion and surface structure from local affine flow paramters. *IEEE Trans. on PAMI*, 27(4):562–574, 2005.

[6] M. Chessa, S. P. Sabatini, F. Solari, and G. M. Bisio. A recursive approach to the design of adjustable linear models for complex motion analysis. In *SPPRA'07*, pages 33–38, 2007.

[7] R. Cipolla and A. Blake. Image divergence and deformation from closed curves. *International Journal of Robotics Research*, 16(1):77–96, 1997.

[8] D.J. Fleet, MJ. Black, Y.Yacoob, and A.D. Jepson. Design and use of linear models for image motion analysis. *Int. J. of Computer Vision*, 36(3):171–193, 2000.

[9] S. C. Fu and P. Kovesi. Extracting differential inavriants of motion directly from optical flow. *13th SCSSE*, 1:108–116, 2004.

[10] J.J. Koenderink. Optic flow. *Vision Res.*, 26(1):161–179, 1986.

[11] F. G. Meyer and P. Bouthemy. Region-based tracking using affine motion models in long image sequences. *CVGIP: Image Understanding*, 60(2):119–140, 1994.

[12] R.C. Nelson and J. Aloimonos. Obstacle avoidance using flow field divergence. *IEEE Trans. on PAMI*, 11(10):1102–1106, 1989.

[13] G.A. Orban. The analysis of motion signal and the nature of processing in the primate system. In *Artificial and Biological Vision System*, pages 24–56. ESPRIT Basic Research Series, 1992.

[14] S. M. Smith and J. M. Brady. Asset-2: Real-time motion segmentation and shape tracking. *IEEE Trans. on PAMI*, 17(9):814–820, 1995.

[15] Huang Yu, Guangyou Xu, and Yuanxin Zhu. Extraction of spatial-temporal features for vision-based gesture recognition. *J. Comput. Sci. Technol.*, 15(1):64–72, 2000.

[16] Z. Zhang and O. D. Faugeras. Three-dimensional motion computation and object segmentation in a long sequence of stereo frames. *Int. J. of Computer Vision*, 7(3):211–241, 1992.