Project no.:	**IST-FP6-FET-16276-2**
Project full title.:	**Learning to emulate perception action cycles in a driving school scenario**
Project Acronym:	**DRIVSCO**
Deliverable no:	**D2.2**
Title of the deliverable:	**Adaptation mechanisms using non-visual signals**

Date of Delivery:	**29th February 2008**
Organization name of lead contractor:	**UGE**
Author(s):	**S. P. Sabatini, M. Chessa, F. Solari**
Partecipant(s):	**UGE**
Work package contributing to the deliverable:	**WP2**
Nature:	**R**
Version:	**1.0**
Total number of page:	**23**
Start date of project:	**1 Feb. 2006**
Duration:	42 **months**

Project Co-funded by the European Commission		
Dissemination level		
PU	Public	**X**
PP	Restricted to other program partecipants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Abstract

In this deliverable we analyze the integration of non-visual information to cross-validate the visual cues and to obtain a reliable description of the visual scene.

Contents

1 Introduction

The main goal of this workpackage is the integration of additional information (e.g. coming from the context) or other exteroceptive sources to obtain reliable cues. The starting point on which we have worked, by adding external information, are the feature maps obtained form the low-level vision system. In the previous deliverable D2.1 we have addressed the problem of recurrent processing for low-level feature regularization. A spatial regularization over a neighborhood resulted to be no crucial because low-level features were reliable enough.

Deliverable 2.2 focuses on adaptation mechanisms that rely upon non-visual signals. In the following section we will describe different approaches to combine visual information and non-visual signals with adaptive and recurrent techniques:

1. The visual cues can be cross-validated using non-visual signals. The motion parameters of the car ($\mathbf{T}$ and $\mathbf{\Omega}$), provided by the University of Münster and HELLA, can be used to compute structure from motion from a monocular image sequence. It is possible to obtain depth maps to validate and integrate the depth estimation from binocular disparity.

2. It is also possible to obtain a consistent description of the visual scene on a more global scale. Important information, directly usable for the interpretation of the scene, can be obtained from the linear description of optic flow fields and disparity maps. The linear (affine) description can be obtained by using the recurrent linear templates described in D2.1 and can be integrated with the motion parameters of the car for the derivation of Structured Visual Events (SVEs, e.g., time-to-contact, heading, orientation of surfaces).

The report is organized as follows:

- In Section 2 the patch-wise linear descriptors analyzed in D2.1 are summarized and some improvements to the approach are described.

- In Section 3 an approach for cross-validating depth maps by the integration of non visual signals is presented.

- In Section 4 the problem of motion interpretation by combining optic flow affine description and non visual signals is analyzed.

2 Patch-wise linear descriptors of visual feature maps

2.1 Basic principles

Kalman Filter is a powerful technique for real-time estimation of dynamic systems. The filter is based on two different models: the *process model*, that describes the evolution over

time of the state vector $\mathbf{x}(t)$, through the transition matrix $\mathbf{\Phi}(t, t-1)$ and the process noise $\mathbf{n_2}(t)$ whose autocorrelation matrix is $\mathbf{Q_2}(t)$:

$$\mathbf{x}(t) = \mathbf{\Phi}(t, t-1)\mathbf{x}(t-1) + \mathbf{n_2}(t-1) \tag{1}$$

and the *measurement model* that relates the current measures $\mathbf{y}(t)$ to the current state trough a measurement matrix $\mathbf{C}(t)$ and a measure noise $\mathbf{n_1}(t)$ with autocorrelation matrix $\mathbf{Q_1}(t)$:

$$\mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{n_1}(t) \tag{2}$$

The algorithm operates in two phases, the first is the *prediction*:

$$\hat{\mathbf{x}}^-(t) = \mathbf{\Phi}(t, t-1)\hat{\mathbf{x}}^+(t-1) \tag{3}$$

$$\mathbf{K}(t+1, t) = \mathbf{\Phi}(t+1, t)\mathbf{K}(t)\mathbf{\Phi}^T(t+1, t) + \mathbf{Q}_1(n) \tag{4}$$

where $\hat{\mathbf{x}}^-(t)$ is the a priori estiamte and $\mathbf{K}(t+1, t)$ is its autocorrelation of the error and $\hat{\mathbf{x}}^+(t)$ is the a posteriori estimate and $\mathbf{K}(t)$ is its autocorrelation of the error.

The second phase is the *update*:

$$\alpha(t) = \mathbf{y}(t) - \mathbf{C}(t)\hat{\mathbf{x}}^-(t) \tag{5}$$

$$\mathbf{\Sigma}(t) = \mathbf{C}(t)\mathbf{K}(t, t-1)\mathbf{C}^T(t) + \mathbf{Q}_2(t) \tag{6}$$

$$\mathbf{G}(t) = \mathbf{\Phi}(t+1, t)\mathbf{K}(t, t-1)\mathbf{C}^T(t)\mathbf{\Sigma}^{-1}(t) \tag{7}$$

$$\hat{\mathbf{x}}(t) = \mathbf{\Phi}(t, t+1)\hat{\mathbf{x}}^+(t) \tag{8}$$

$$\mathbf{K}(t) = \mathbf{K}(t, t-1) - \mathbf{\Phi}(t, t+1)\mathbf{G}(t)\mathbf{C}(t)\mathbf{K}(t, t-1) \tag{9}$$

where $\alpha(t)$ is the innovation process, $\mathbf{\Sigma}(t)$ is the autocorrelation of the innovation process and $\mathbf{G}(t)$ is the Kalman gain.

2.2 Generalizations and improvements

The patch-wise linear descriptors approach based on Kalman Filter, discussed in Deliverable 2.1, has been improved: a multiscale approach and a reliability measures have been added in order to obtain more stable results.

2.3 Multiscale

Spatial variations of the coefficients involve features of the scene at different resolution. To properly analyze these features a multi-resolution analysis is necessary. It is possible to follow two different approaches: (1) to analyze optic flow with patches of different dimensions; (2) to analyze optic flow at different scales. A pyramidal approach has been preferred to keep reasonable the computational load of the Kalman Filter. Figure 1 shows a frame from the sequence *Town03*, the associated optic flow and the divergence map computed at different spatial scale (the patch size is constant). At a rough scale the background has a constant divergence due to ego-motion, while on the pedestrian a spatial variation of divergence emerges. The same analysis at a finest scale is noisier because the considered neighborhood is too small.

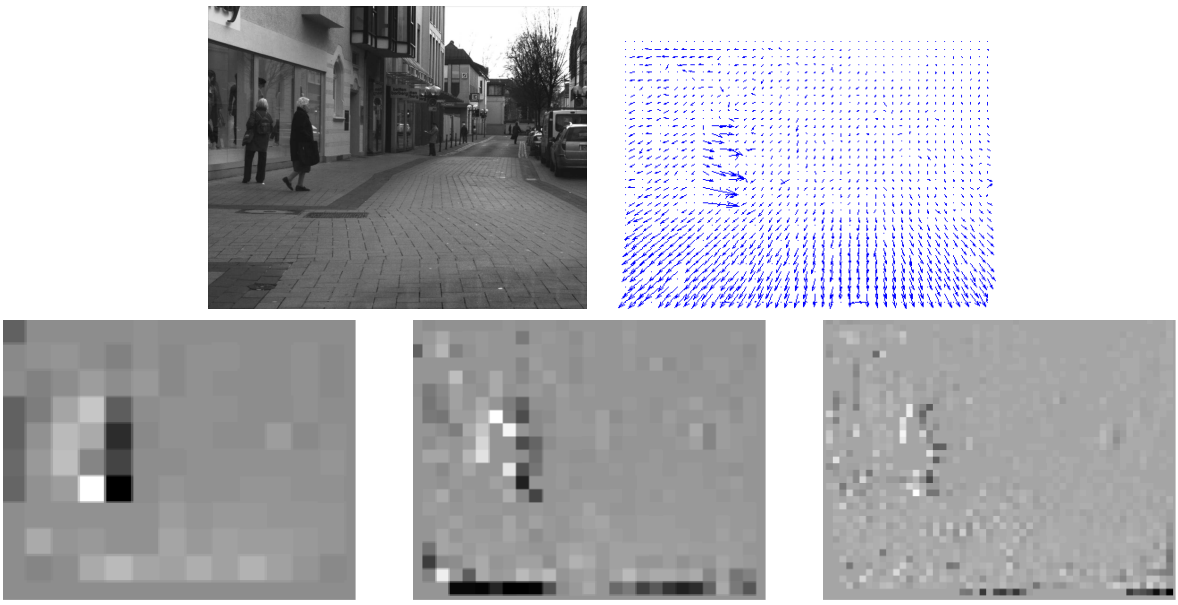

Figure 1: First row: Frame from the Town03 sequence and the associated optic flow. Second row: Divergence computed at different spatial scales, from left to right (coars eto fine): the roughest scale (on the left) allows us to detect the spatial variation of divergence on the pedestrian with respect with to the background, finer scales (on the right) are noisier.

In the Appendix A comparison among the results obtained at different spatial scales is shown.

2.4 Reliability measure

To obtain a reliability measure of the estimates, a confidence map, based on Normalized Innovation Squared, has been added to the estimation process. The Normalized Innovation Squared (NIS) measures the discrepancy between the input measures and the Kalman estimates:

$$\text{NIS} = \alpha(t)^T \mathbf{S}^{-1}(t)\alpha(t) \tag{10}$$

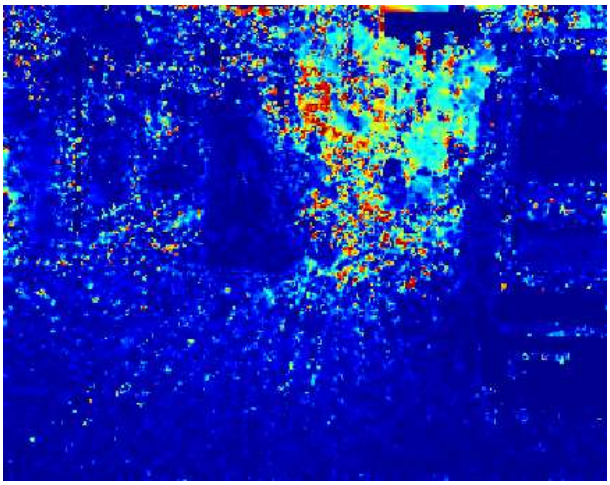

Figure 2: Estimation of the angular error. Where the angular error is large (red values) the estimation of the measure error is increased. The estimation of measure error is done independently in the different regions of the image.

NIS is chi-squared distributed with $n \times n$ degrees of freedom:

$$\text{NIS} < \chi^{-1}(p|n). \tag{11}$$

Values that are less then a threshold (chosen with 0.95 percentile) are accepted. From Eq. 10 it is evident that NIS depends on the covariance of the innovation $\mathbf{S}$, that depends on the estimate of the measure noise. The estimation of measure noise should be relative to the magnitude of the measure, so we decided to use the mean angular error between current frame n and previous frame $n-1$ to tune the measure noise. Moreover, all the parameters of the filter are adjusted independently in each patch, thus allowing a maximum flexibility and adaptability during the filtering stage.

3 Cross-validation of depth maps by integration of non visual signals

The known camera motion can be used to estimate depth from a monocular image sequences. Depth from motion can be used to cross-validate the binocular depth estimate.

A frame from a binocular image pair and the corresponding depth map, computed with the phase-based algorithm using only non-visual cues, is shown in Figure 4. The aim of this work is to combine the optic flow information (see Figure 6 and the car parameters to integrate information where binocular depth map is less reliable (e.g. on the tree on the foreground).

This approach has been described in [7], where a "classical" structure from motion technique is improved by a predictive filtering stage. For many applications having an on-line incremental estimate of depth is important. This could be achieved by using an adaptive recurrent filter, such as the Kalman Filter.

Figure 3: Reliability mask computed in accordance with the chi-squared distribution of NIS. The threshold has been chosen according to a 0.95 percentile.

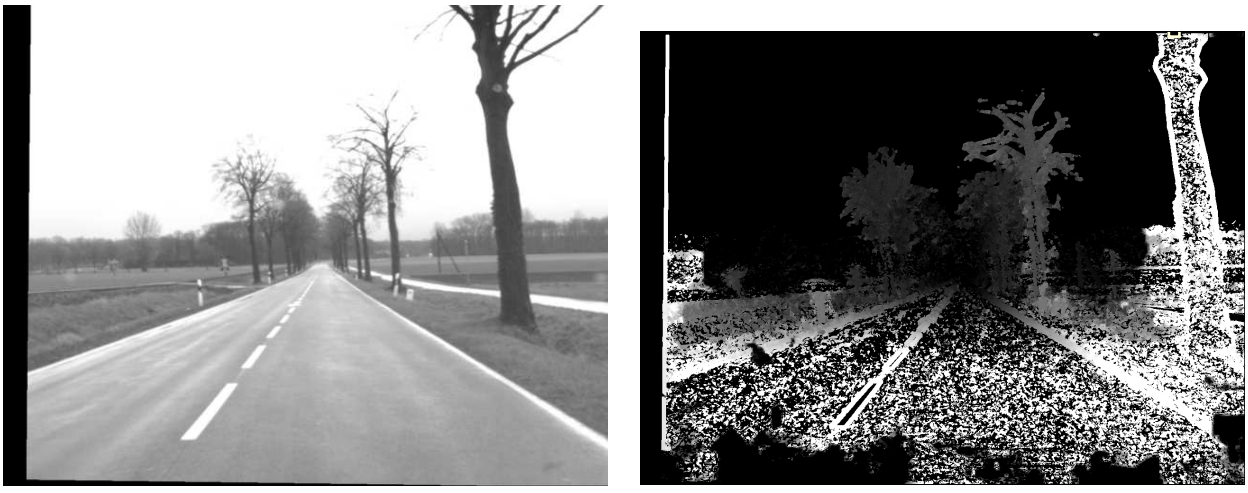

Figure 4: (a) Frame from a street sequence. (b) Depth map obtained with the phase-based algorithm for disparity (using only non-visual cues).

3.1 Structure from motion

If the camera motion is known, the motion of a 3D point in the scene is described by the following equation:

$$\frac{d\mathbf{P}}{dt} = -\mathbf{T} - \mathbf{\Omega} \times \mathbf{P} \tag{12}$$

If we consider an unitary focal length the projection of the point $\mathbf{P} = (X, Y, Z)$ onto the image plane is given by:

$$x = \frac{X}{Z} \qquad y = \frac{Y}{Z}. \tag{13}$$

By taking the derivatives of (x, y) with respect to the time we obtain the equation of optic flow:

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} -1 & 0 & x \\ 0 & -1 & y \end{bmatrix} \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} + \begin{bmatrix} xy & -(1+x^2) & y \\ (1+y^2) & -xy & -x \end{bmatrix} + \begin{bmatrix} \Omega_x \\ \Omega_y \\ \Omega_z \end{bmatrix}. \tag{14}$$

These equation relate the depth Z of the point to the camera motion $\mathbf{T}$, $\mathbf{\Omega}$ and the optic flow. Thus, by combining the information of optic flow coming from the vision front-end and the car motion information read from the can bus it is possible to obtain an estimate of depth from a monocular sequence. By defining:

$$\mathbf{H} = \begin{bmatrix} t_x \\ t_y \end{bmatrix} = \begin{bmatrix} -1 & 0 & x \\ 0 & -1 & y \end{bmatrix} \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}, \tag{15}$$

a disparity measure can be obtained from the noisy optic flow by solving numerically the following equation:

$$d = (\mathbf{H}^T \mathbf{P}_m^{-1} \mathbf{H})^{-1} H^T \mathbf{P}_m^{-1} \mathbf{v}, \tag{16}$$

where $\mathbf{P}_m$ is the covariance of the noise in the flow measurements.

3.2 Kalman Filter-based Algorithm.

It is possible to define a Kalman filter to integrate at each step new disparity measurements (obtained with Eq. 16) with the predicted disparity map. The information about the translation and rotation of the car are obtained from the data provided by Hella. It is necessary to have a measure of optic flow (from the front-end) and of the associated noise covariance. The translational components of velocity in the image plane (t_x, t_y) are directly obtained from the motion of the car (T_x, T_y, T_z). At first we have assumed that each value in the measured or predicted disparity map is not correlated with its neighbors, so the state vector is composed by a single value of disparity for each pixel and the Kalman Filter can proceed independently for each pixel in the image. Then it is necessary to define the covariance matrix of the process noise $\mathbf{Q}_2$ and the covariance matrix of the measure noise $\mathbf{Q}_1$. At each step an a-priori prediction of the disparity $\hat{\delta}^-(t)$ is combined with the innovation process $\alpha(t) = \delta(t) - \hat{\delta}^-(t)$ through the Kalman Gain $\mathbf{G}(t)$ to obtain the a-posteriori estimate of disparity $\hat{\delta}^+(t) = \hat{\delta}^-(t) + \mathbf{G}(t)\alpha(t)$. Figure 5 shows the steps of the Kalman Filter.

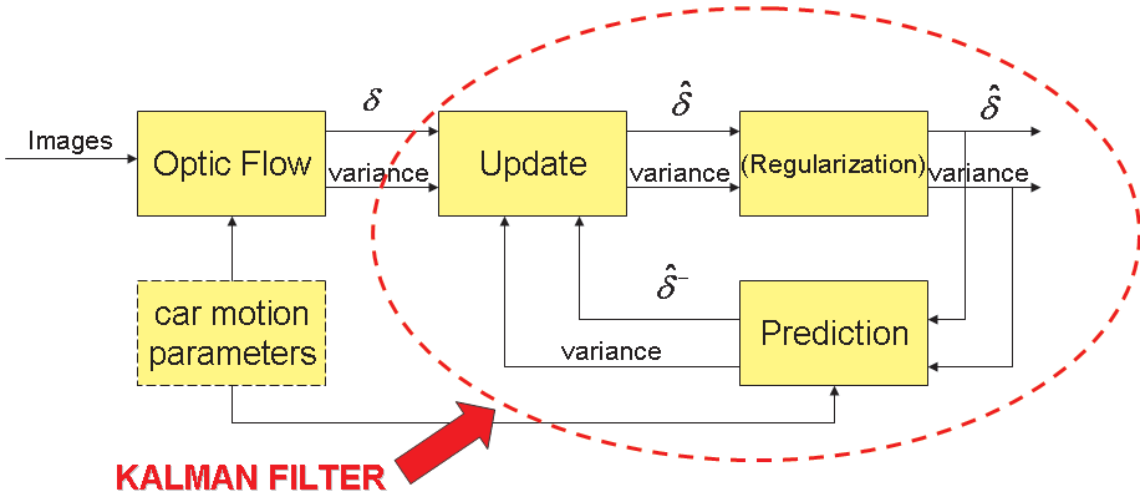

Figure 5: Integration of the car motion parameters to compute disparity with Kalman Filter. The raw disparity (δ) is combined with the a priori estimation ($\hat{\delta}^-$) to obtain the a posteriori estimation ($\hat{\delta}$) after the Kalman filter update.

3.3 Results

In this section the results obtained with the described approach are analyzed. All the results have been obtained starting from the sequence and the related optic flow map of Fig. 6 (from the low-level vision front-end).

Figures 7 and 8 show a depth map obtained with the described approach. In Figure 7 there is no update according with the Kalman Filter equation, while Figure 8 shows the results obtained by using the a posteriori estimate of the Kalman Filter. A constant value of translation T_z has been considered, with a covariance $P_m = 0.05$, a measure noise with autocorrelation 0.001 and a process noise with autocorrelation 0.008.

It is worth noting that it is necessary to warp the estimate of disparity according to the optic flow information to continue the estimation process in the correct image location in the following steps. Figure 9 shows the depth map if the estimated values are not correctly warped: a "ghost" effect is visible especially on the tree on the foreground.

Figures 10 shows an other example of depth map obtained for a different road situation.

4 Motion interpretation by combining optic flow affine description with non visual signals

In the literature there are many different approaches that aim to estimate rigid scene structure and the relative 3D motion of a camera from an image sequence. All these approaches start from the description of the models used for the camera imaging geometry, the motion of the scene relative to the camera and the structure of rigid surfaces in the scene.

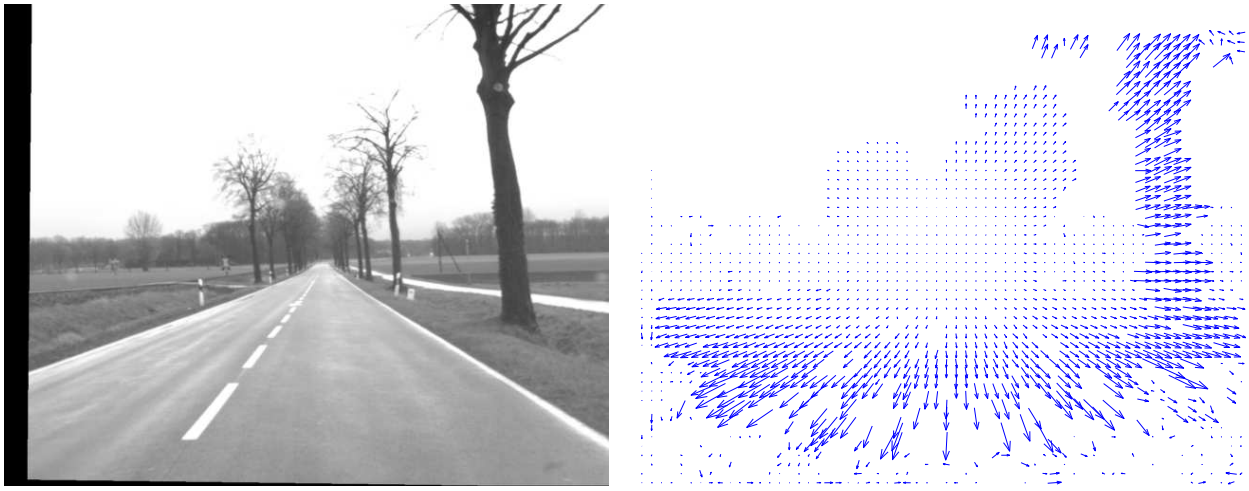

Figure 6: (a) Frame from a street sequence. (b) Optic flow obtained with the phase-based approach.

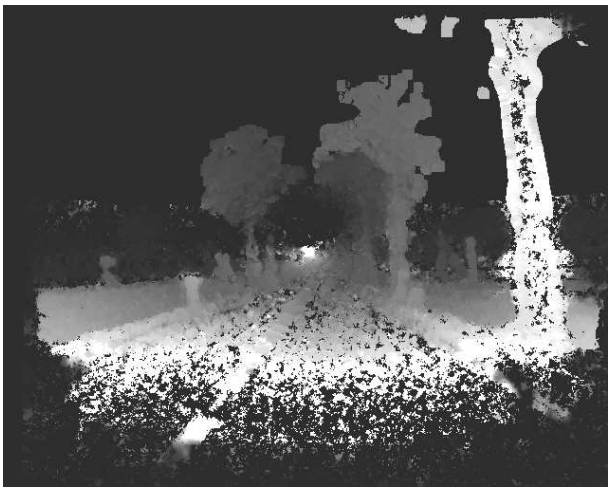

Figure 7: Depth map computed using the non visual signals together with visual information (optic flow), without the Kalman Filter update.

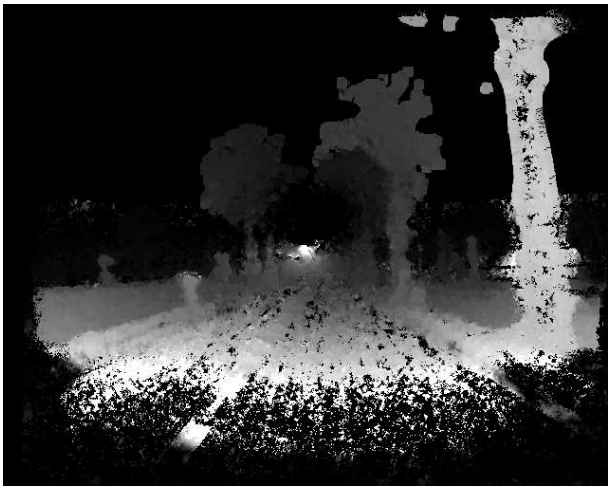

Figure 8: Depth map obtained from the a posteriori estimate of the depth map by using the Kalman Filter update.

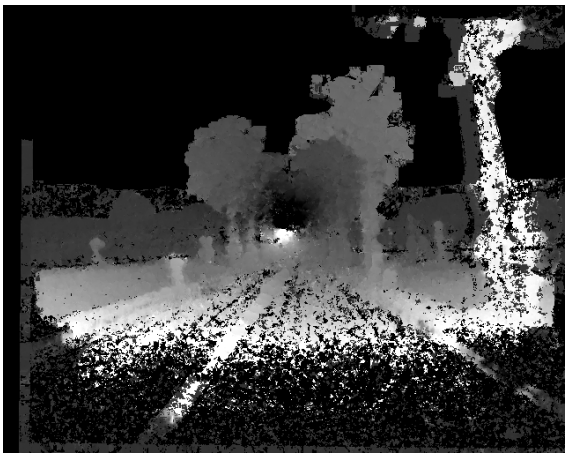

Figure 9: Depth map obtained without warping the estimates. It is possible to see a "ghost" effect, especially evident on the tree on the right of the image.

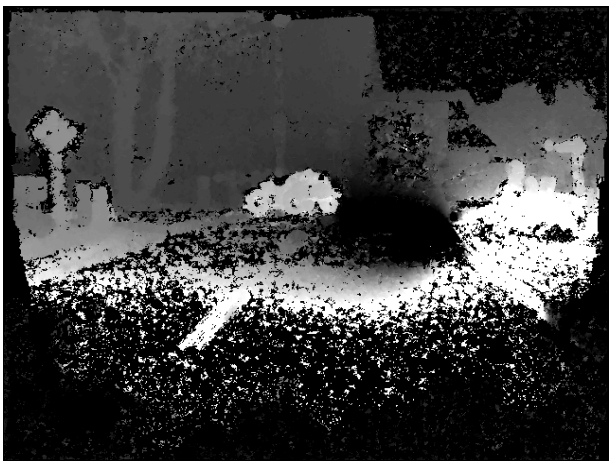

Figure 10: Depth map computed using the non visual signals together with visual information (optic flow) with a recurrent estimation in an other road situation (sequence Tour03).

4.1 Basic principles

At a first approximation, and under proper conditions [6] [8], important information about heading, 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, and many different approaches have been proposed in the literature to recover reliable estimates of these differential properties. It is worth noting that a a complete solution for the 3D motion estimation using only a first-order approximation is not possible, without considering additional information. Several approaches can be used to overcome the problem: (1) to give a qualitative interpretation of the first-order approximation under proper assumptions; (2) to solve for the interesting parameter by minimizing an error function in different area of the patch [2]; (3) to use additional sources of information if they are available.

In the context of the DRIVSCO project we have a source of information about non visual signals that can be useful to integrate the visual information and to help to solve the ill-posed problem.

4.2 First-order description of optic flow

Within any small image region, and under smooth change in viewpoint [5], an affine model of image motion

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c_5 \\ c_6 \end{bmatrix} \tag{17}$$

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:

$$\bar{\mathbf{L}} = \mathbf{L}|_{\mathbf{x_0}} = \begin{bmatrix} \frac{\partial v_x}{\partial x} & \frac{\partial v_x}{\partial y} \\ \frac{\partial v_y}{\partial x} & \frac{\partial v_y}{\partial y} \end{bmatrix}_{\mathbf{x}=\mathbf{x_0}} \tag{18}$$

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

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} \bar{L}_{11} & \bar{L}_{12} \\ \bar{L}_{21} & \bar{L}_{22} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \bar{v}_x \\ \bar{v}_y \end{bmatrix}. \tag{19}$$

Equivalentely, the differential invariants of image motion can be related to algebraic combinations of the affine coefficients:

$$\begin{aligned} \mathrm{div}\boldsymbol{v} &= c_1 + c_4 \\ \mathrm{curl}\boldsymbol{v} &= c_2 - c_3 \\ (\mathrm{def}\boldsymbol{v})\cos 2\theta &= c_1 - c_4 \\ (\mathrm{def}\boldsymbol{v})\sin 2\theta &= c_2 + c_3 \end{aligned}$$

and they represent: an isotropic expansion specifying a change in scale, a 2D rigid rotation specifying a change in orientation, and the components of a pure shear along the axis of expansion described by the orientation θ, respectively.

If the viewer motion is known, divergence, curl and deformation are sufficient to unambiguosly recover the surface orientation and the distance to the object (time to collision) [3] [4]. The car motion is defined by its translational $\mathbf{T} = (T_x, T_y, T_z)$ and rotational components $\mathbf{\Omega} = (\Omega_x, \Omega_y, \Omega_z)$. Both can be recovered from the can bus of the car. The viewing direction is defined by the vector $\mathbf{Q}$. The component of translational velocity parallel to the image plane scaled by depth Z is defined by:

$$\mathbf{A} = \frac{\mathbf{T} - (\mathbf{T} \cdot \mathbf{Q})\mathbf{Q}}{\mathbf{Z}}. \tag{20}$$

The surface orientation is represented by the depth gradient, scaled by depth:

$$\mathbf{F} = f\nabla(\log \mathbf{Z}) = \frac{f\nabla \mathbf{Z}}{\mathbf{Z}} \tag{21}$$

The magnitude of the depth gradient is the tangent of the slant (σ) of the surface and its direction corresponds to the tilt (τ) of the surface tangent plane.

$$|\mathbf{F}| = \tan\sigma \tag{22}$$

$$\angle\mathbf{F} = \tau \tag{23}$$

The differential invariants can be rewritten using these quantities:

$$\begin{aligned} \text{div}\boldsymbol{v} &= \frac{2\mathbf{T}\cdot\mathbf{Q}}{Z} + \mathbf{F}\cdot\mathbf{A} \\ \text{curl}\boldsymbol{v} &= -2\boldsymbol{\Omega}\cdot\mathbf{Q} + |\mathbf{F}\cdot\mathbf{A}| \\ \text{def}\boldsymbol{v} &= |\mathbf{F}||\mathbf{A}|. \end{aligned}$$

From these equations and under proper conditions it is possible to recover important information about the 3D shape of the scene without the knowledge of the motion parameters. For example in presence of a pure translation along the ray towards the surface patch ($|\mathbf{A}| = 0$) the divergence can give important information about the time to contact t_c:

$$t_c = \frac{\mathbf{Z}}{\mathbf{T}\cdot\mathbf{Q}} \tag{24}$$

Even without this assumption it is possible to recover useful information from the first-order differential invariants. The information about time to collision can be expressed as bounds:

$$\frac{2}{\text{div}\mathbf{v} + \text{def}\mathbf{v}} \leq t_c \leq \frac{2}{\text{div}\mathbf{v} - \text{def}\mathbf{v}} \tag{25}$$

If we consider a pure translational motion perpendicular to the visual direction it will result an image deformation with a magnitude which is determined by the slant of the surface σ and with an axis depending on the tilt of the surface τ. It is worth noting that divergence and deformation are unaffected by viewer rotations such as panning and tilting of the cameras.

In the context of the DRIVSCO project we can use both non visual information and estimates coming from the early-vision front-end in order to obtain more reliable information.

4.3 Kalman Filter Implementation

To this goal the affine coefficients estimated with the recursive algorithm (see Deliverable 2.1 and Appendix A) can be used together with the known translational and rotational values of the car to obtain information about time to collision and the slant of the surfaces (see Figure 11)

If we are not in presence of pure translational motion and the rotational components are known, the relationship between the divergence of optic flow and the time to collision becomes [1]:

$$\text{div}\mathbf{v} = -\frac{2}{t_c} + \frac{3x\omega_y}{f} + \frac{3y\omega_x}{f}, \tag{26}$$

and the two components of the depth gradient vector (see Eq. 21) $\mathbf{F} = (p, q)$ are

related to the affine coefficients of the optic flow by:

$$\begin{aligned} c_1 &= \frac{T_z}{Z_0} + \frac{pT_x}{Z_0} \\ c_2 &= \omega_z + \frac{qT_x}{Z_0} \\ c_4 &= -\omega_z + \frac{pT_y}{Z_0} \\ c_5 &= \frac{T_z}{Z_0} + \frac{qT_y}{Z_0}. \end{aligned} \tag{27}$$

Figure 11 shows a schematic representation of the approach we have followed: optic flow is analyzed with the Kalman-based patch-wise filter, from which it is possible to recover the 6 coefficients of the affine description. Those coefficients can be combined with the non-visual inputs leading to a motion interpretation (slant of the surface and time-to-collision).

4.4 Results

We applied the recursive KF to the optic flows computed from real-world driving sequences recorded in different situations. The aim of the experimental analysis is to obtain a description of motion in different multiple motion situations. A multiscale approach has been followed to describe the linear properties at different image resolutions.

The green colormap in Figures 12 and 13 shows the information about TTC: lighter green corresponds to higher values related to far objects or to forward moving objects.

The maps in Figures 14 and 15 show the information about the slant of the surfaces. Green patches correspond to "horizontal" structures (e.g. the street) while "red" patches represent standing objects (e.g. a wall).

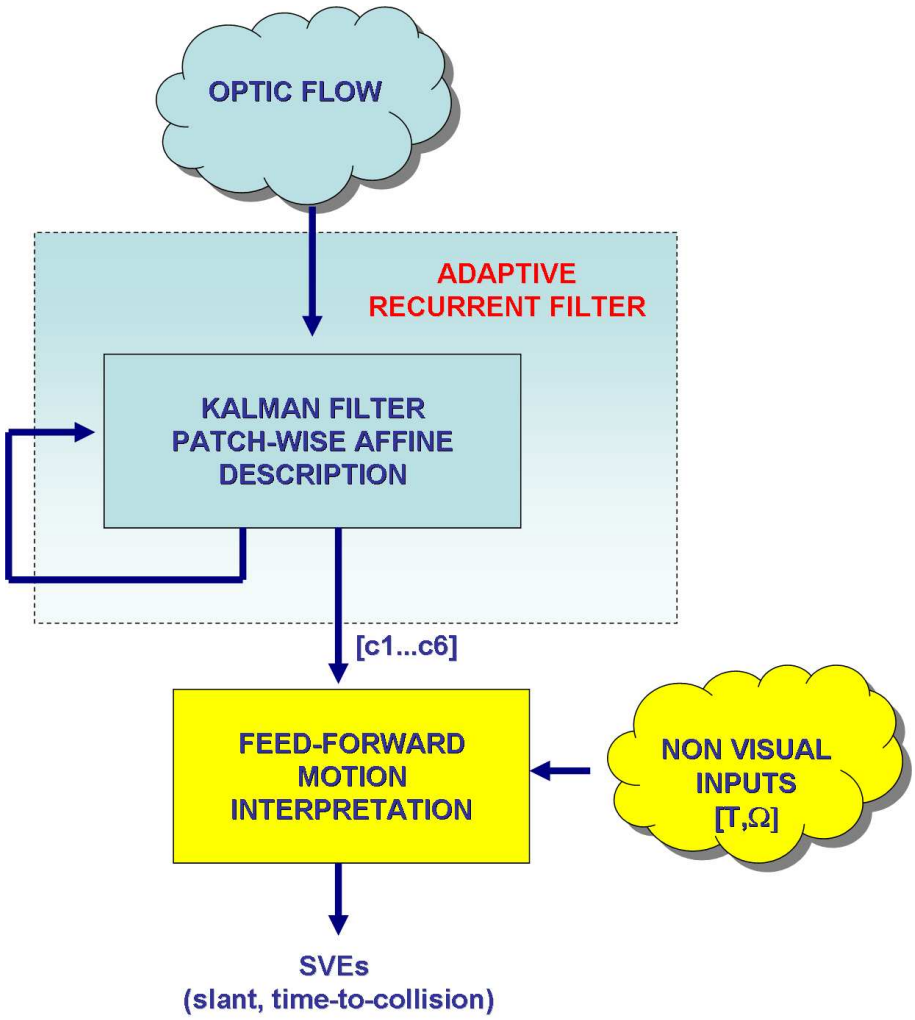

Figure 11: The optic flow coming from the low-level vision front-end is analyzed with the recurrent patch-wise approach. The Kalman Filter estimate are continuosly adapted as new measures of optic flow arrive. The resulting affine description is then combined with the non visual inputs allowing us to obtain a 3D scene interpretation and to recover Structural Visual Events (e.g. slant of the surfaces, time-to-collision).

Figure 12: Map representing the time to collision. The truck is moving along the same direction of the observer, and its TTC is higher than the one to the still objects at the same depth

Figure 13: Map representing the time to collision. The van is moving along the same direction of the observer, and its TTC is higher than the one to the still objects at the same depth

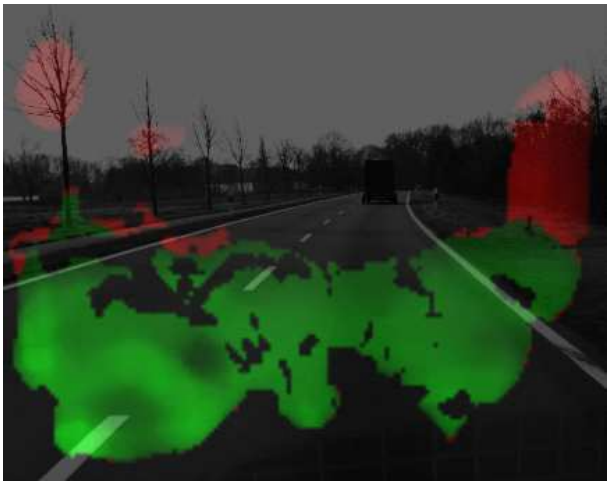

Figure 14: Higher-level feature maps obtained from ego-motion information. (a) Optic flow. (b) Slanted surfaces present in the scene: reddish regions correspond to standing objects aside of the road, greenish regions correspond to the road plane.

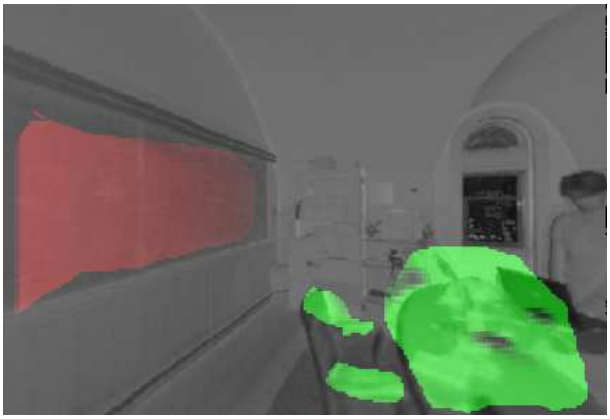

Figure 15: Higher-level feature maps obtained from ego-motion information using the Brown Range Image Database. (a) Optic flow. (b) Slanted surfaces present in the scene: reddish regions correspond to vertical planes (e.g. the blackboard), greenish regions correspond to horizontal planes (e.g. the table).

References

[1] A. Del Bimbo and S. Santini. *Motion Analysis in: Human and Machine Vision: Analogies and Divergence*. Plenum Press, 1994.

[2] A. Calway. Recursive estimation of 3d motion and surface structure from local affine flow parameters. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 27(4):562–574, 2005.

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

[4] F. Domini and M.L. Braunstein. Recovery of 3d structure from motion is neither euclidean nor affine. *Journal of Experimental Psychology: Human Perception and Performance*, 24(7):1273–1295, 1998.

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

[6] J.J. Koenderink and A.J. Van Doorn. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, 22:773–791, 1975.

[7] R. Szeliski L. Matthies and T. Kanade. Kalman filter-based algorithms for estiamting depth from image sequences. *International Journal of Computer Vision*, 3(3):209–238, 1989.

[8] A. Verri, M. Straforini, and V. Torre. Computational aspects of motion perception in natural and artificial systems. 337:429–443, 1992.

Appendix A

In this Appendix some comparisons between the results obtained with the patch-wise approach at different spatial scales are shown. The multiscale approach is necessary to analyze the features at different spatial scale. If we are interested in "global" features of the scene, such as divergence to detect time-to-collision or the slant of the surfaces, it is more convenient to work at a rough scale. In other situations (e.g. if we are interested in features like edges) to work at a finer resolution is better. Figure 16 shows the variation of the probability values in two different patches of the "motorway sequence" analyzed at a proper spatial scale. The temporal behaviour of the probability values for the 4 models is similar if we consider two patches belonging to the car.

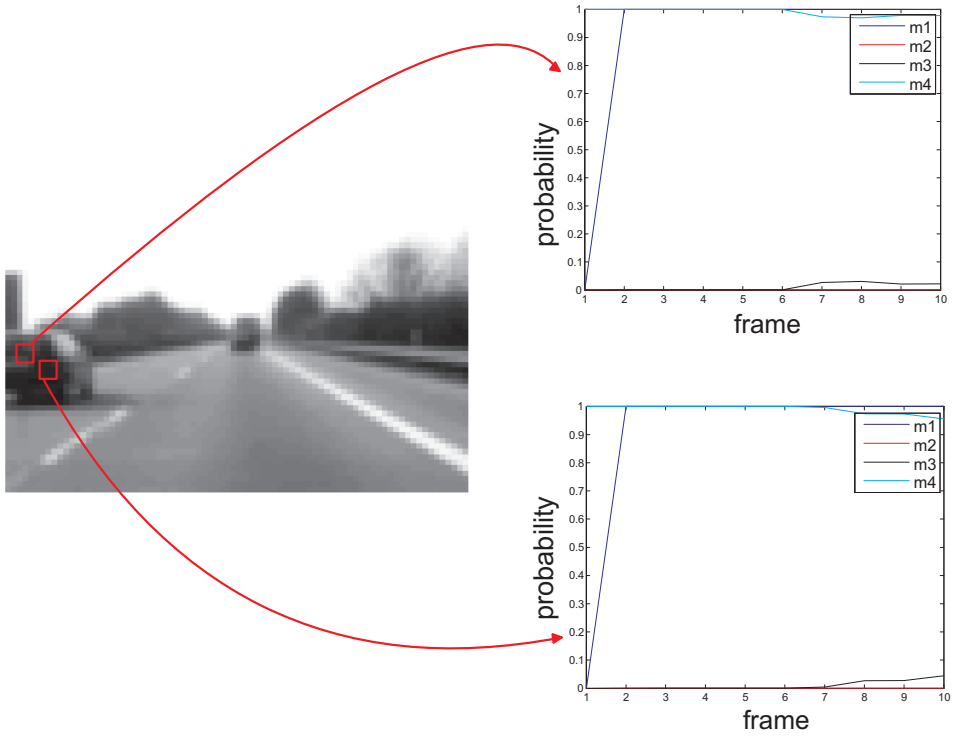

Figure 16: Patch-wise analysis computed at a coarse scale. In the image two adjacent patches are highlighted and the temporal behaviour of the probability values for the 4 models are shown. The behaviour of the coefficients in the two patches is similar. It is coherent with the fact that the two patches belong to the same object (the car).

If we use the patch-wise approach to analyze optic flow at a finer spatial scale we will obtain unstable results due to the fact that the spatial neighborhood is too small. An example of this behaviour can be seen in Figure 17

It is worth noting that if we work at a wrong scale even the temporal behaviour of the estimated models could be unstable. Figures 18 and 19 show the differences in the temporal behaviour for two patches at the same spatial location analyzed at different resolutions.

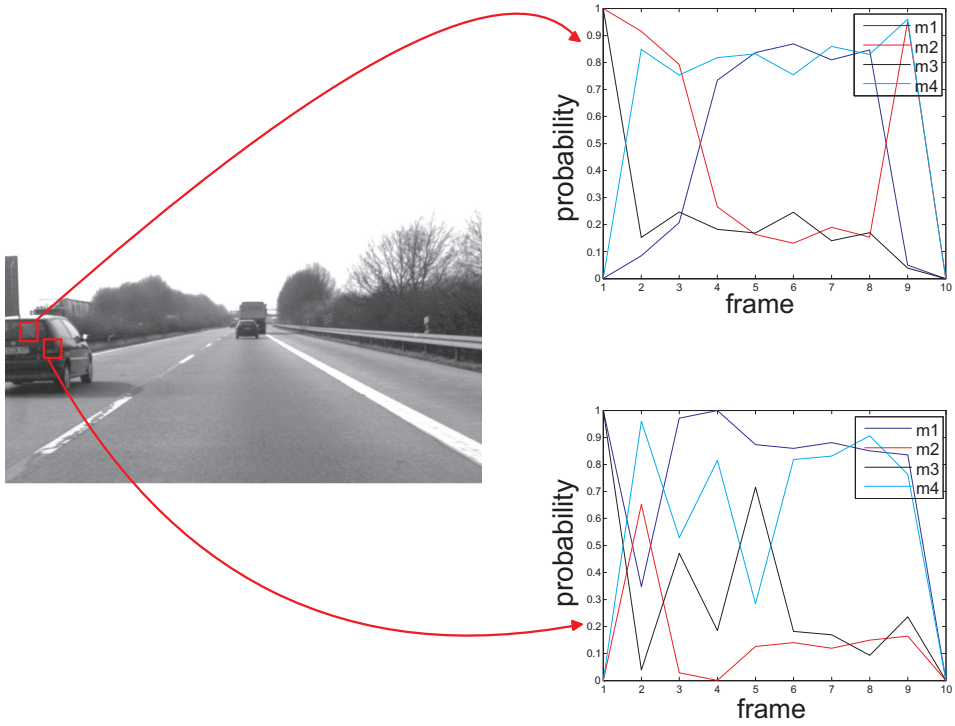

Figure 17: Patch-wise analysis computed at a fine scale. In the image two adjacent patches are highlighted and the temporal behaviour of the probability values for the 4 models are shown. The behaviour of the coefficients in the two patches is different from one patch to another even if the two patches belong to the same object (the car). This is due to the fact that the patch is too small.

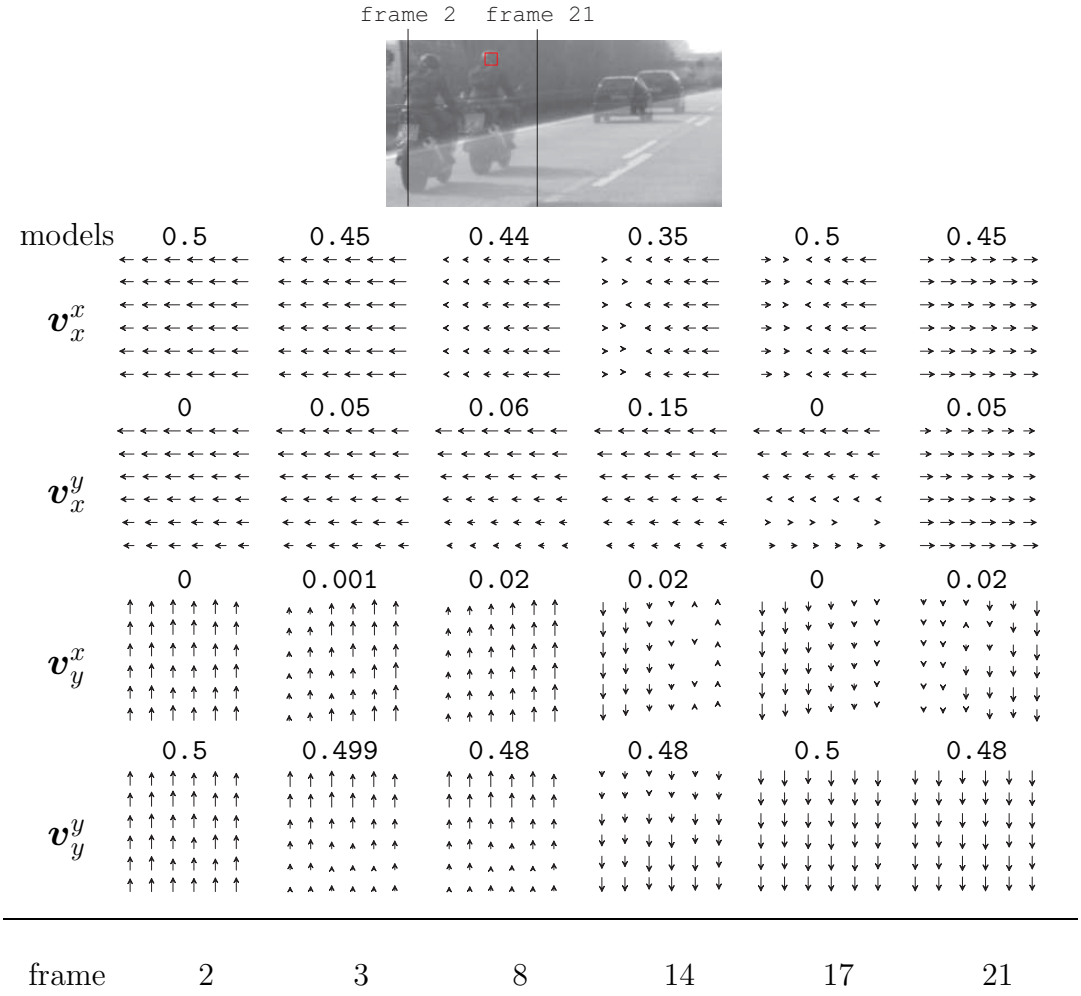

Figure 18: Evolution in time of the four optic flow models in the same image patch at a rough scale. The red square that localize the image patch is enlarged for the sake of representation. We can observe through frames the behavior of each model for different motion situations: at frame 2, the patch 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 contain the motion of the motorbike, only. The number on the top of each model indicates the associated probability. By using a proper scale it is possible to detect slow changes in the image motion and in the optic flow models.

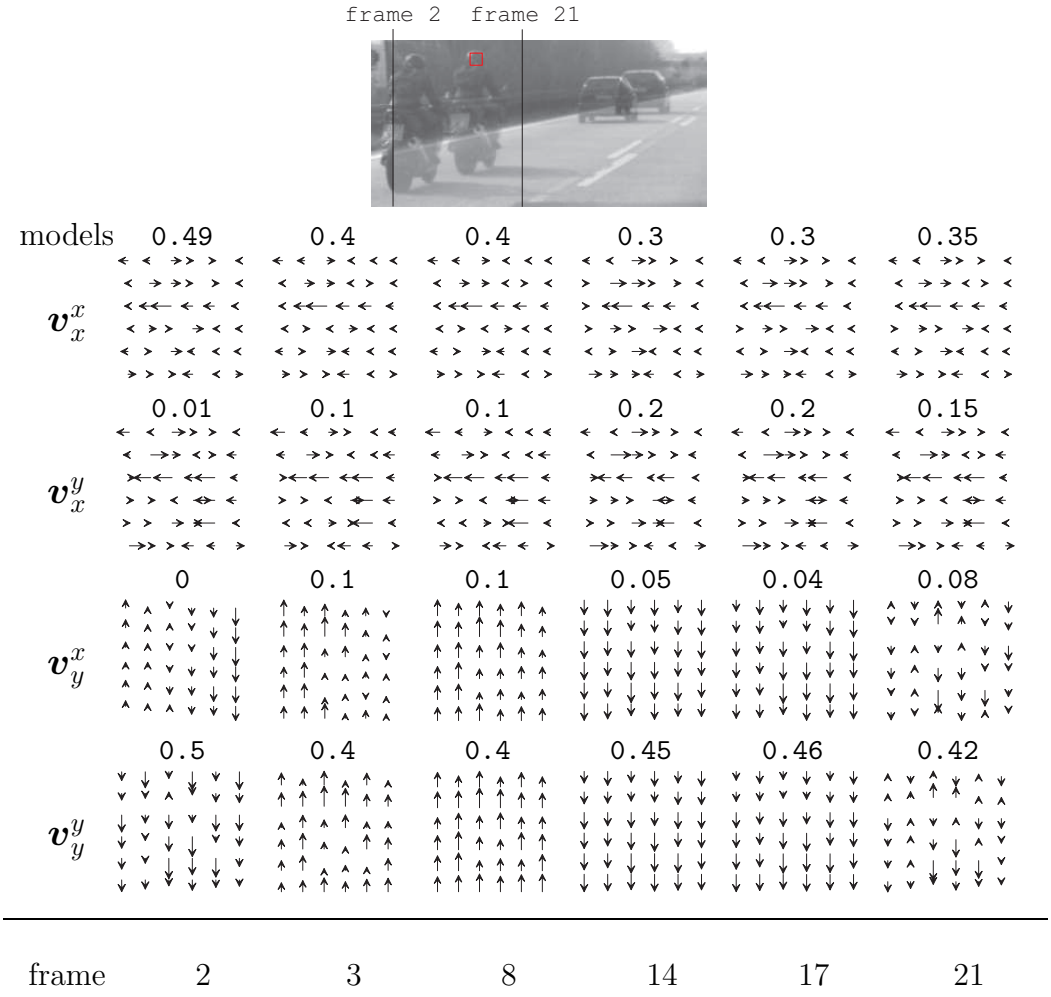

Figure 19: Evolution in time of the four optic flow models at a fine scale. The patch location is the same of Figure 18. The red square that localize the image patch is enlarged for the sake of representation. In this case the models do not represent correctly the motion in the scene because the considered neighborhood is too small. The resulting patches are noisy and do not correspond to the linear models we have defined.