Information Society Technologies

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 mechanism using non-visual signals – Quantitative evaluation of time-to-contact performance**

Date of Delivery:	14.6.2008
Organization name of lead contractor for this deliverable:	UGE
Author(s):	M. Chessa, S.P. Sabatini, F. Solari (UGE)
Participant(s):	UGE, UMU
Work package contributing to the deliverable:	WP2
Nature:	R
Version:	2.0 (revised 13/06/2008)
Total number of pages:	19
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 participants (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)	

Summary: We present a specific analysis of the problem of the time-to-contact estimation from first-order flow descriptors obtained by the Kalman-based context sensitive filters. A quantitative evaluation of our results with both synthetic and natural sequences is reported. A more extensive analysis of the performances of area based descriptors for extracting SVEs will be provided in an update Deliverable 2.2a, to also reflect changes of plan as necessary for the September meeting.

Contents

1 Premise

As described in Deliverable D2.1, a Kalman-based filter implementing spatial-context rules can be used to describe an optic flow field in terms of first-order elementary flow components. On the basis of these first-order flow descriptors, the research activities of WP2 have been directed towards the functional characterization of such descriptors for their direct use in motion interpretation and scene understanding. In particular, we are interested in extracting the following structured visual events (SVEs): (1) heading direction, (2) time-to-contact, and (3) information about the orientation of surfaces in the scene. Here, as requested by the reviewers, we present a specific analysis of the problem of the time-to-contact estimation and a quantitative evaluation of our results with both synthetic and natural sequences. From a general perspective, a thorough analysis of the "as-a-whole" performances of these area-based descriptors for extracting SVEs for high-level scene interpretation will be the goal of a complete revision of the Deliverable D2.2, that will result in an additional deliverable (D2.2a) to be submitted by the end of August and it will be discussed at the review meeting in mid September 2008. Due to the lack of standard benchmark in the literature, a quantitative comparative analysis of our results is not possible. Though, the results can be compared qualitatively with those obtained under similar conditions. We plan to make available to the Computer Vision community the benchmark sequences used in our analysis for comparing the results.

2 Introduction

In a broad sense, the time-to-collision (TTC) corresponds to the map of the temporal distance between the observer and *any* point in the scene. This implies the solution of a complex problem that concerns the determination of the heading direction of the viewer, the reconstruction of the surface orientation of the scene, as well as the segmentation of IMOs that further complicates the solution of the problem and the consequent selection/decision process of braking/steering actions. From a more restrictive perspective, we can limit on recovering only that information that is relevant for a vehicle to avoid an obstacle obstructing its path. According to this "purposive vision" paradigm, TTC is just a single scalar value that can be directly recovered from the structural properties of the optic flow field around the heading direction (and used to maneuver a vehicle in presence of obstacles). However, still in this case, a complete knowledge of the geometry and the relative motion of objects in the scene is necessary to disambiguate TTC measures in real-world situations, unless introducing further simplification assumptions.

In this report, we assume that the optical axis of the camera is always approximately aligned with the heading direction of the car, and we limit the visual analysis around the center of the image. In this way, meaningful time-to-collision estimates are extracted from planar image flow approximations [1]. The role of additional assumptions on the external environment are discussed.

The idea of employing TTC from first-order derivatives of the optical flows goes back to the early 90s [2, 3, 4]. In general, such methods share the drawback of being sensitive to errors in the estimates of optical flow, since the latter are always corrupted by noise that

is amplified by the process of differentiation. As an alternative, families of simple fixed flow divergence templates have been proposed in an attempt to overcome the problems associated with the computation of image velocity derivatives. Along this line, Meyer proposes a technique for applying the theoretical analysis of [5] in realistic situations. He assumes that the image motion field can be segmented into regions whose motion can be accurately described by affine models. The coefficients of these models along with their temporal derivatives are estimated by a multiresolution scheme and temporal continuity of motion using a Kalman Filter. The time-to-contact is then recovered using the estimated coefficients. Although the computation of dense optic flow fields and their derivatives is avoided this method is sensitive to luminance variations across scales of the image patches.

In this work, we propose an alternative to the method of Meyer, which works on the image velocity field, while avoiding an explicit differentiation of the optic flow: the affine coefficients are obtained from the Kalman-based adaptive templates defined in Deliverable D2.1, working on the optic flow obtained by the front-end modules defined in WP1.

3 Time-to-contact from first-order optic flow descriptors

A relative motion between an observer and an object induces a corresponding image motion field, which is divergent in nature when the object moves towards the camera. From this divergent image flow it is possible to derive an estimate of the time-to-contact (also known as the time-to-collision or time-to-impact) with the object in the field of view of the moving observer, which is defined as the amount of time that remains before the object in question collides with the observer, provided that they continue to maintain the same relative translational velocity [6]. For a system with a camera pointing at the same direction of heading, the time-to-contact can be computed from the ratio between the distance of the image point of the object from the focus of expansion and the magnitude of radial flow [7]. More generally, computing the time-to-contact is a complex problem that implies, in principles, solving in advance the structure from motion problem and especially separating the translational and rotational components of relative rigid motion. Though, as we will show in the following, it has been observed that under proper simplification assumptions, at least bounding values of the TTC can be directly related to the first-order differential invariants of the image flow by simple algebraic relationships.

3.1 Basic principles

The motion of an observer in a static environment can be described at each instant t as a rigid-body motion, by means of two vectors (i.e., kinetic characteristics): the translational velocity $\mathbf{T} = (T_X, T_Y, T_Z)^t$, and the angular velocity $\mathbf{\Omega} = (\Omega_X, \Omega_Y, \Omega_Z)^t$. If one considers a pinhole camera model with the optical center in the origin of a viewer-centered coordinate frame (see Fig. 1) and the optical axis oriented along the Z axis, the

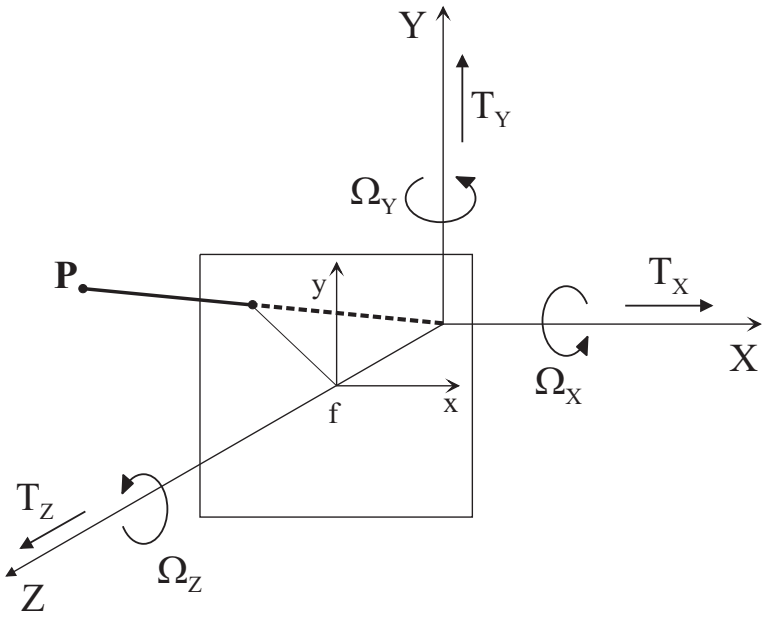

Figure 1: Viewer-centered coordinate frame.

perspective projection $\mathbf{p} = (x, y)$ of a point $\mathbf{P} = (X, Y, Z)^t$ of a visible surface in the 3-D space is defined as

$$\begin{aligned} x &= f\frac{X}{Z} \\ y &= f\frac{Y}{Z} \end{aligned}$$

where, without loss of generality, the focal length of the optical system f can be set to 1. The image motion field $\dot{\mathbf{p}} = (\dot{x}, \dot{y})^t = (u, v)^t$ is expressible as a function of image position $\mathbf{p}$ and surface depth $Z = Z(x, y)$ (i.e., the depth of the object projecting in (x, y) at current time):

$$[u(x, y), v(x, y)]^t = \frac{1}{Z}\mathbf{H}(x, y)\mathbf{T} + \mathbf{G}(x, y)\mathbf{\Omega} \tag{1}$$

with

$$\begin{aligned} \mathbf{H}(x, y) &= \begin{bmatrix} -f & 0 & x \\ 0 & -f & y \end{bmatrix} \\ \mathbf{G}(x, y) &= \begin{bmatrix} \frac{xy}{f} & -\left(f + \frac{x^2}{f}\right) & y \\ f + \frac{y^2}{f} & -\frac{xy}{f} & -x \end{bmatrix}. \end{aligned}$$

For a sufficiently small field of view (i.e., within any small image region) [8]), an affine model of image motion

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c_5 \\ c_6 \end{bmatrix} \tag{2}$$

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 $\boldsymbol{v}(x,y) = [u(x,y), v(x,y)]$. Formally, the parameters c_5 and c_6 represent the horizontal ($\bar{u}_0$) and vertical ($\bar{v}_0$) 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:

$$\bar{\mathbf{L}} = \mathbf{L}|_{\mathbf{x_0}} = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}_{\mathbf{x}=\mathbf{x_0}} = \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix}_{\mathbf{x}=\mathbf{x_0}} \tag{3}$$

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

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \bar{L}_{11} & \bar{L}_{12} \\ \bar{L}_{21} & \bar{L}_{22} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}. \tag{4}$$

It is worth noting that this corresponds to a planar approximation of the viewed surface patch around $\mathbf{P}$

$$Z \simeq Z_0 + Z_X X + Z_Y Y = Z_0 + Z_x x + Z_y y \tag{5}$$

where (Z_x, Z_y) is the local surface orientation in the image plane. (Under the assumption of (i) a small field of view and (ii) a small depth range of the object compared to the viewing distance (i.e., $\Delta Z/Z << 1$), the transformation in shape between times t and $t + \mathrm{d}t$ (for small $\mathrm{d}t$) can be approximated by a linear (affine) transformation [1].

Accordingly, we can derive the relationships between the linear affine coefficients and the geometric and kinematic information about the scene:

$$\begin{array}{ll} c_1 = T_Z/Z_0 + T_X Z_x/Z_0^2 & c_2 = \Omega_Z + T_X Z_y/Z_0^2 \\ c_3 = -\Omega_Z + T_Y Z_x/Z_0^2 & c_4 = T_Z/Z_0 + T_Y Z_y/Z_0^2 \\ c_5 = -T_X/Z_0 - \Omega_Y & c_6 = -T_Y/Z_0 + \Omega_X \end{array}. \tag{6}$$

Equivalently, from the Cauchy-Stokes decomposition theorem:

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

that 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.

Defining $\mathbf{A} = (T_X/Z_0, T_Y/Z_0)$ as the viewer translation, $\mathbf{F} = (Z_x, Z_y)/Z_0 = \nabla Z/Z_0$ as the surface orientation (represented by the depth gradient, scaled by depth), and $\mathbf{Q}$

to be a unit vector along the view direction, from Eqs.(6), we can obtain direct relationships between the affine coefficients and behaviorally-relevant quantities for motion interpretation and scene understanding:

$$\begin{aligned} \mathrm{div}\boldsymbol{v} &= c_1 + c_4 = \frac{2\mathbf{T}\cdot\mathbf{Q}}{Z} + \mathbf{F}\cdot\mathbf{A} \\ \mathrm{curl}\boldsymbol{v} &= c_2 - c_3 = -2\boldsymbol{\Omega}\cdot\mathbf{Q} + |\mathbf{F}\cdot\mathbf{A}| \\ \mathrm{def}\boldsymbol{v} &= [(c_1 - c_4)^2 + (c_2 + c_3)^2]^{1/2} = |\mathbf{F}||\mathbf{A}|. \end{aligned} \tag{7}$$

where θ bisects $\mathbf{A}$ and $\mathbf{F}$:

$$\theta = \frac{\angle\mathbf{A} + \angle\mathbf{F}}{2}. \tag{8}$$

It is worth noting that 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.

$$|\mathbf{F}| = \tan\sigma \tag{9}$$

$$\angle\mathbf{F} = \tau \tag{10}$$

From these equations 7 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:

$$t_c = \frac{\mathbf{Z}}{\mathbf{T}\cdot\mathbf{Q}} \tag{11}$$

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:

$$\frac{2}{\mathrm{div}\mathbf{v} + \mathrm{def}\mathbf{v}} \leq t_c \leq \frac{2}{\mathrm{div}\mathbf{v} - \mathrm{def}\mathbf{v}} \tag{12}$$

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, which is a valuable property because it allows the cues to be utilized even when the camera is not completely stabilized.

3.2 Working assumptions

The narrow field of view hypothesis ($\mathbf{p} \sim 0$) adopted in Section 3.1 it is not sufficient to derive from differential invariants quantitative measurements of TTC in real-world situations (cf. Eq. 12). To remove the residual ambiguities, additional (exteroceptive) assumptions must be introduced. Such assumptions are based on a partial a priori knowledge of either camera-scene relative geometry or motion. Although, in general, these assumptions limit the application of these methods to carefully controlled scenarios, some

of them are quite reasonable and do not represent a severe limitation of the operating conditions for the system used in the DRIVSCO project. In this section we summarize a classification of the assumptions discussing their different implications on the TTC estimation.

Although there is a connection between the divergence and TTC, as described by Eqs. (7), which can be exploited to achieve obstacle avoidance in a qualitative way, the TTC cannot be recovered from the divergence alone, since the values of divergence can be altered also by other geometric and kinematic conditions. By example, translation parallel to an inclined surface or certain motion discontinuities can produce or being interpreted as a divergent flow. In these situations, deformation components occur, which are generated by rotational velocity and/or by motion towards non-frontoparallel surfaces.

In general, non-ambiguous TTC estimates can be obtained from first-order structure of the planar motion field, without solving explicitly the rigid motion problem, provided that some constraints are set on relative motion or viewing angles:

1. *Narrow field of view* ($\boldsymbol{p} \sim 0$)[4] [2]
2. *Dominant translation* ($||\boldsymbol{\Omega}|| \sim 0$) [9] [10].
3. *Frontoparallel surface* ($||\nabla Z|| \sim 0$) [11] [3].

Specifically, it is possible to show that the deformation vector **defv** can be expressed as the sum of two terms, taking into account translations and rotations, respectively:

$$\mathbf{defv} = \mathbf{def_T}(\mathbf{p}; \nabla Z; Z; \mathbf{T}) + \mathbf{def_\Omega}(\mathbf{p}; \boldsymbol{\Omega}). \tag{13}$$

The term $\mathbf{def_\Omega}$ vanishes in the case of pure translation ($\boldsymbol{\Omega} = 0$) or, whatever $\boldsymbol{\Omega}$, at the image origin ($\mathbf{p} = 0$). According to the narrow field hypothesis ($\mathbf{p} \simeq 0$) the $\mathbf{def_\Omega}$ can be neglected without introducing the "dominant translation" hypothesis. On the other hand, $\mathbf{def_T}$ vanishes either in the case of pure rotation ($\mathbf{T} = 0$) or, whatever $\mathbf{T}$, if the tangent plane at $\mathbf{P}$) is parallel to the image plane (frontoparallel condition, $\nabla Z = 0$). Considering that we assumed a motion in the direction of the optical axis, the residual term $\mathbf{def_T}$ can be neglected, without assuming the "frontoparallel surface" hypothesis.

3.3 Algorithm

Our method for TTC detection is based on the direct use of the flow field differential invariants, as originally proposed by [4] [2] further developed by [12] [13]. Though, differently from those approaches we estimate the differential invariants through an adaptive technique that recovers first-order flow components from patch-wise context-sensitive filtering of a dense optic flow field. Here, we assume that the image optic flow is given by the front-end developed in WP1. In this way, we exploit the full vector field information (but see [12]) while avoiding the problem of numerical differentiation of the noisy flow field. The affine coefficients obtained by the adaptive templates are indeed more reliable and robust to noise than those derived by smoothed differentials [14] or by bilinear interpolation (see Deliverable D2.1), and come together with their confidence measures.

Considering the small vergence angle of the cameras fixed on the car, we assume that the viewing direction is almost aligned with the heading direction. This simplifies the relations of **A** and allows us o estimate the TTC with an object on the path of the observer by limiting the analysis to the central (foveal) part of the image. It is worth noting that, though this is a realistic situation for the DRIVSCO experimental set-up, in general, it is possible to first derive the heading direction [1], still from the linear flow properties, and then focus the attention around the heading direction to estimate the TTC with possible obstacles along the path. This will be part of a future work.

The complete algorithm is given as follows:

1. Subsample the flow field at a low resolution. We subsample an optic flow of size 512×640 to a size of 32×40.

2. Calculate the differential invariants for a 3×3 neighborhood around the image center, by using the Kalman-based filtering procedure, described in Deliverable D2.1. The size of each patch is 6×6. By considering the dimension of the patch the total size of the considered neighborhood is 12×12 at the lower resolution.

3. Verify that the **def** components are negligible, under the simplification assumptions of narrow field-of-view and of motion towards a frontoparallel surface (this hypothesis will be removed in a future work).

4. Derive an estimate of the local divergence by spatial averaging of the different estimates over the neighborhood, rescaling the divergence value respect with the original size of the image.

5. Finally, the time-to-collision is computed according to Eqs. (7).

4 Evaluation of performances

4.1 Approach and methodology

To quantitatively assess the performance of our technique, the estimated TTC must be compared with the "ground truth". The different methods proposed in the literature were tested either with synthetic image sequences or with real-world sequences captured from a moving vehicle equipped with a laser range finder and a camera. Unfortunately, to the best of the authors' knowledge, no systematic benchmark sequences are available from public databases, and this prevents a comparative analysis of our results with the state-of-the-art.

In the following, we will test our method with a VRML synthetic sequence and with real videos provided by Hella, from camera mounted on a car, for which the LIDAR information from the CAN bus is used as the ground truth. Artificial video sequences with known ground truth can also be constructed using a stop motion technique, where the

[1] e.g., by solving an overconstrained system of equations obtained by Eq. 7 for a sufficient number of points in the image plane.

camera (or the object) is moved by a controlled amount between exposures. However, this approach does require particular care, since accurate increments in position are required. In addition, as reported in [15], stop motion sequences produced with ordinary digital cameras suffer from the effects of automatic focus and automatic exposure adjustments, as well as artifact introduced by image compression. The effects of the estimation error on the reaction time of a potential automatic emergency system are discussed in Section 4.4.

4.2 Evaluation of TTC with synthetic sequences

First we create a synthetic benchmark sequence to evaluate the performances of the proposed approach for time-to-contact estimation. All the parameters of the synthetic scene are known and can be varied and controlled.

The sequence has been created in OpenGL environment and it is composed of different textured-surfaces representing a street and two side walls and an object (a cube) in the middle of the scene, object respect to which we want to measure the TTC. The virtual camera moves toward the object, along the Z axis. Figure 2 shows some frames captured by the virtual camera. We tested the approach by varying the speed of the camera.

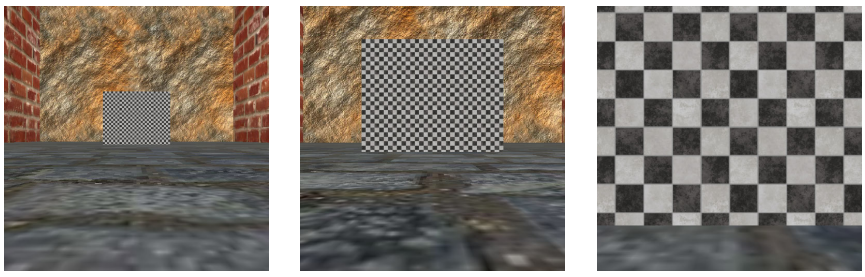

Figure 2: Virtual scene used to benchmark time-to-contact. The camera has been moved toward the black and white cube at different speeds.

The parameters that describe the scene are the following:

- Initial distance between the observer (the camera) and the target: 48 m
- Dimension of the target: 3 m $\times$2 m

We tested the approach in three different situations:

- Seq1: speed of the camera 1 m/frame (about 90 km/h);
- Seq2: speed of the camera 0.5 m/frame (about 45 km/h);
- Seq3: speed of the camera 0.5 m/frame (about 45 km/h) with a small difference (10 degrees) between the direction of the camera and the direction of sight;

In all the cases the speed has been kept constant, and in the first case the observer arrives very close to the target object. For each frame of the sequences the affine coefficients (then the divergence) have been estimated, then from the divergence we obtain

the TTC, following the approach described in Section 3.3. The plots in Figure 3, in Figure 4 and in Figure 5 show, as a function of the time, the estimation of the time to collision (continuous line), the linear fitting of the experimental results (dotted line) and the ground truth (thick line). The values of ground truth have been obtained by considering the known speed of the observer and the initial distance between the camera and the target. Since we kept constant the velocity, the time-to-contact decreases as a straight line. The equations of the linear fitting are $TTC = -t + 48$, $TTC = -0.51t + 48$ and $TTC = -0.48t + 43$ respectively. The slope of the linear fitting is coherent with the velocity of the camera and the starting point is the initial distance between the camera and the target. In the third situation the initial distance is 43 m, as evidenced by the plot in Figure 5 and by the linear fitting. The correspondence between the TTC expressed in seconds and in frames has been done by considering a frame rate of 25 frame/sec. By analyzing the results it is worth noting that the first two/three frames do not give a good estimation of TTC because the Kalman estimate of the affine coefficients needs some frames to reach a steady-state, then the error in the TTC estimation is constant in time. We do not notice an improvement or a deterioration of the performances when approaching the target, except for the situation in Figure 3, where after 1.4 seconds the camera is too close to the cube-target and it is not possible to recover a reliable optic flow.

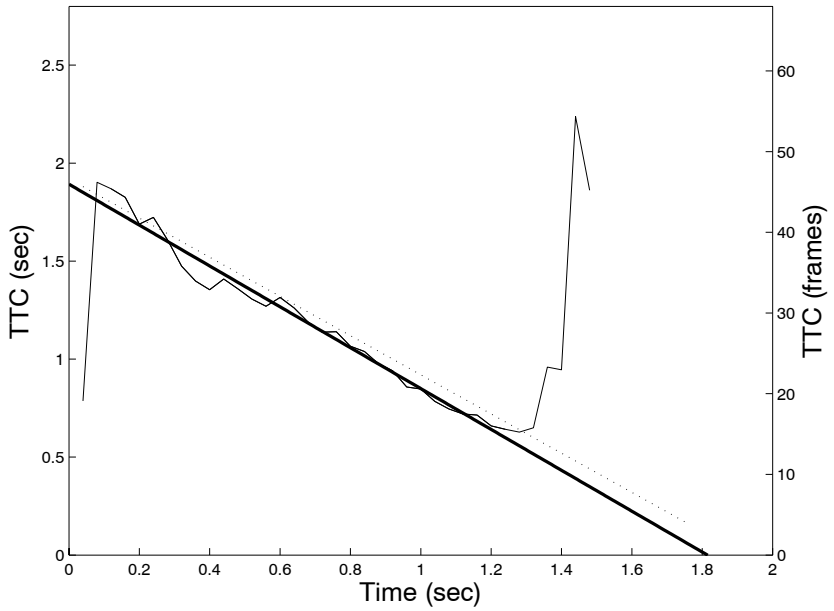

Figure 3: Estimation of TTC for Seq1. The speed of the camera is 1 m/frame. The continuous line is the estimation of TTC with the described approach, the dotted line represents the linear fitting of the data and the thick line is the ground truth. From frame 33 the camera is too close to the target, so both the optical flow estimation and the TTC computation are wrong. The linear fitting does not consider data beyond frame 33.

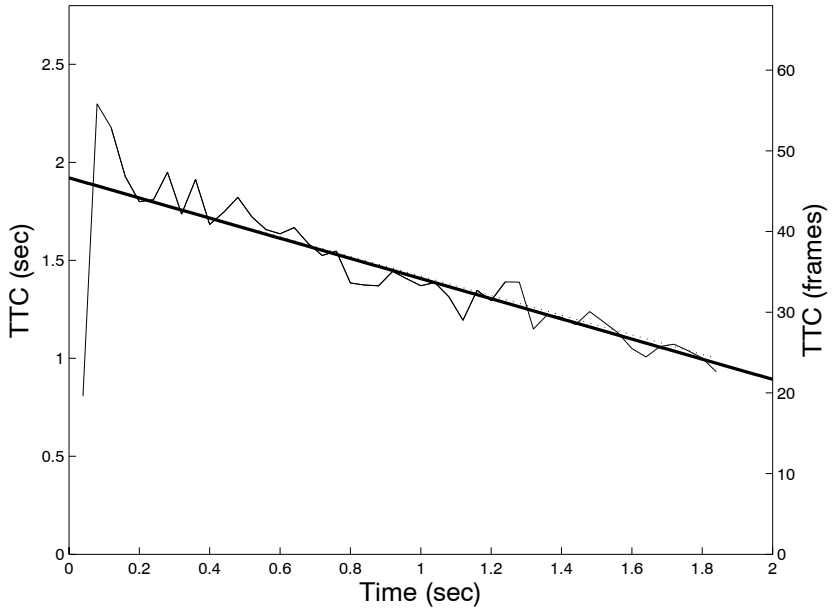

Figure 4: Estimation of TTC for Seq2. The speed of the camera is 0.5 m/frame. The continuous line is the estimation of TTC with the described approach, the dotted line represents the linear fitting of the data and the thick line is the ground truth.

Table 1 shows the mean error in the TTC estimation and its standard deviation for the 3 synthetic sequences. It is worth noting that the mean error is similar in the three different situations.

	mean error (msec)	st. deviation (msec)
seq1	68	47
seq2	62	72
seq3	65	67

Table 1: Mean error and standard deviation of the TTC estimation for the synthetic sequences.

4.3 Evaluation of TTC with real-world sequences

To evaluate the performances of TTC estimation in real-world situations, we used a set of sequences recorded by Hella with the DRIVSCO project setup (see Fig 6). In these sequences the car is driven towards a target (a car-balloon, usually used in crash test experiments). The velocity of the car is similar in the two situations and it is about 37 km/h.

Both the image sequences and the CAN-bus data are available. The CAN-bus data cannot be considered a real ground truth signal because it is affected by errors itself

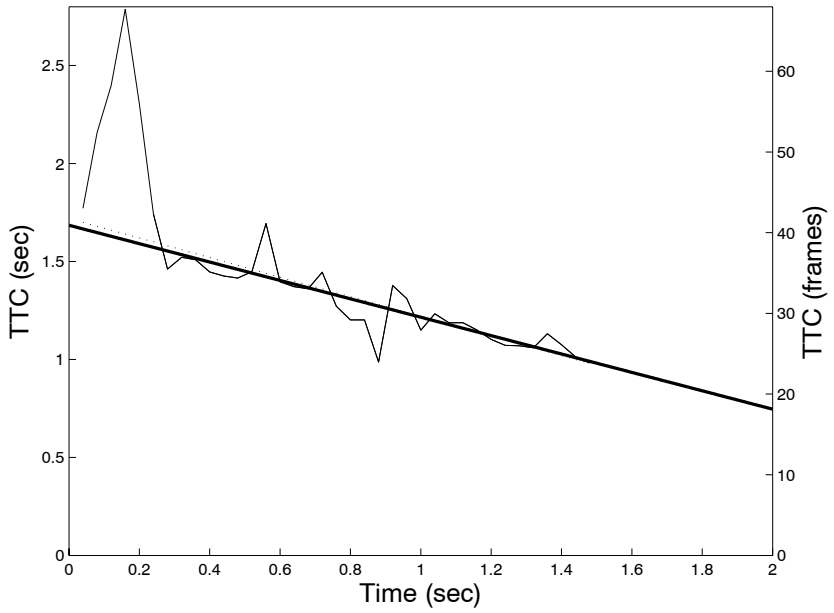

Figure 5: Estimation of TTC for Seq3. The speed of the camera is 0.5 m/frame and there is a small difference (about 10 degrees) between the direction of the camera and the direction of sight. The continuous line is the estimation of TTC with the described approach, the dotted line represents the linear fitting of the data and the thick line is the ground truth.

Figure 6: Crash-test sequence.

(i.e. acquisition errors). Moreover, the sampling frequency of the data from the LIDAR is lower than the frame rate from the cameras and this implies that we do not have a value from the CAN-bus for each estimation. Anyway we have compared our results with the LIDAR signals, when such information is available. From Figure 7 and Figure 8 it is evident that the TTC evaluation with our algorithm is coherent with the LIDAR measures. The presence of possible outliers, as the one occurring in Figure 7, can be reduced by taking an average of the last n time-to-contact estimates (e.g., with $n = 3 \div 4$) as the final measure of the time-to-contact (cf. [7]) or some other means of averaging, such as exponentially decaying filter.

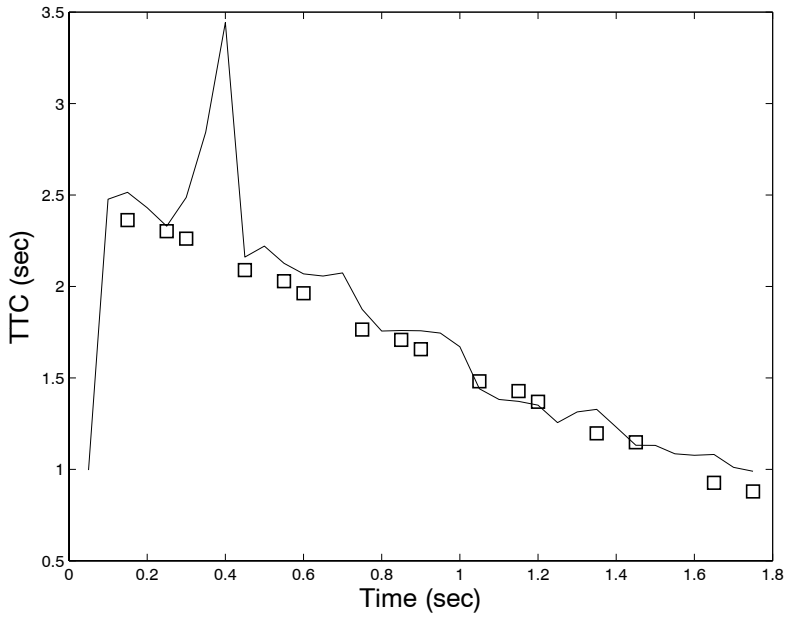

Figure 7: Estimation of TTC. The continuous line is the estimation of TTC with the described approach, the squares represent the data read from the LIDAR.

Since in real-world situations the ground-truth is not available, we have not computed the mean error and the standard deviation of the TTC estimation.

4.4 Behavioral considerations

By assuming to have available the TTC at the potential onset of braking, this information can be used to derive the available manoeuvring space at the moment the evasive action starts. On the basis of the average error in the estimate of the TTC, we can analyze how underestimations of the time-to-contact affect the driver's reaction responses to avoid the collision. To this end, we used the reaction threshold paradigms used in designing emergency braking systems. These systems usually adopts a function with two parameters: the first is the braking distance and the second is the lateral acceleration [16]. The calculation of the thresholds is done in two steps: For the braking system,

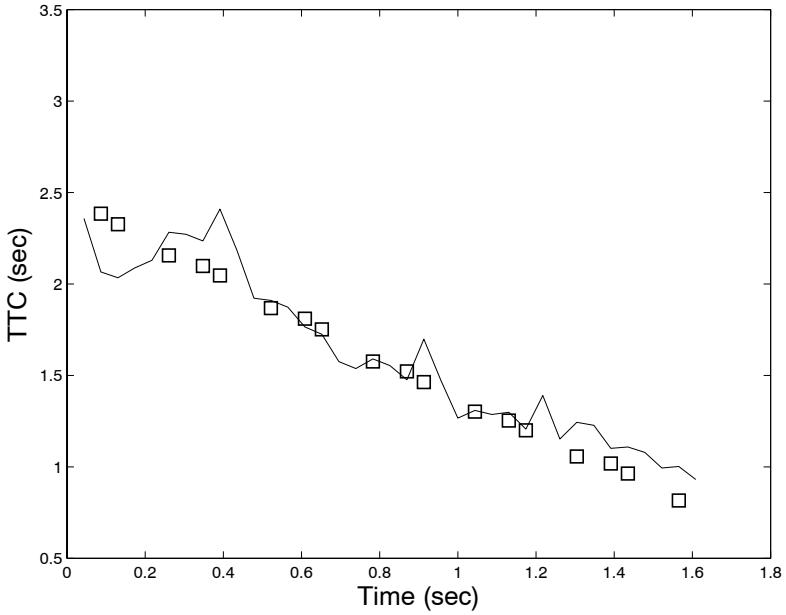

Figure 8: Estimation of TTC. The continuous line is the estimation of TTC with the described approach, the squares represent the data read from the LIDAR.

estimating that the maximum deceleration is not larger than $10m/s^2$, we can calculate if the driver has a chance to stop in front of the obstacle. For the lateral acceleration, we can calculate the maximum lateral distance the driver can handle in the remaining time. If this value is smaller than the size width of the car, then the red corner of the moving vehicle has no chance to pass the blue corner of the obstacle without a collision (see Fig. 9).

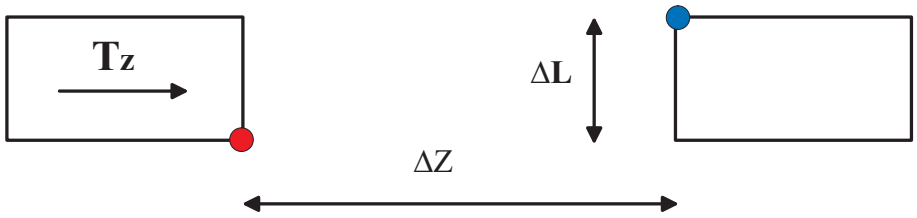

Figure 9: Typical time-to-contact scenario.

In case the result of both calculations shows that the driver has no chance to avoid the collision, the automatic emergency system has to be activated. Formally, by assuming a uniform deceleration, the time to stop can be obtained by:

$$t_{stop} = \frac{T_Z}{a_{Zmax}} \tag{14}$$

where a_{Zmax} is the maximum deceleration (e.g. $a_{Zmax} = 10m/s^2$). The maximum possible

lateral displacement (along the X axis) is:

$$\Delta X = 0.5 a_{Xmax} t_c^2 \tag{15}$$

where a_{Xmax} is the maximum deceleration (e.g. $a_{Xmax} = 7m/s^2$). Given t_c as the maximum time to react, and assuming $t_c > t_{stop}$, the error (ϵ) in the t_c estimate ($t'_c = t_c + \epsilon$) must be below the difference value $t_c - t_{stop}$:

$$t_c - t_{stop} > \epsilon. \tag{16}$$

In other words, an overestimate of the TTC is equivalent to a reduction by the same amount of the effective time-to-stop:

$$t_{stop}|_{eff} = t_{stop} - \epsilon. \tag{17}$$

Although it is not possible to conclude on the maximum tolerable error, since it depends on the specific situation, it is evident that the danger of the error increases as t_c approaches t_{stop}. This can be evidenced by defining the relative error measure:

$$e_{rel} = \frac{\epsilon}{t_c - t_{stop}}. \tag{18}$$

Accordingly, we can state that if t_c is larger than $2t_{stop}$ a relative error of 10% is an acceptable situation. From the results of our simulation, we observed an error $\epsilon \sim 65msec$. Assuming a $T_Z = 90km/h$ and $t_c = 5sec$, $t_{stop} = 2.5sec$, the resulting relative error is:

$$e_{rel} = \frac{0.065}{2.5} = 0.026 \tag{19}$$

which remains below the threshold of 10% until t_c is equal to $3.15sec$.

Similarly, we can derive the effect of an erroneous t_c estimate on the lateral displacement threshold. In this case, $t'_c = t_c + \epsilon$ affects the measure of the lateral displacement as:

$$\Delta X' = \Delta X + a_{Xmax} t_c \epsilon + 0.5 a_{Xmax} \epsilon^2 \tag{20}$$

Disregarding the quadratic term, we can impose that

$$\Delta X - \Delta L > a_{Xmax} t_c \epsilon \tag{21}$$

Again, we can define a relative error:

$$e_{rel} = \frac{a_{Xmax} t_c \epsilon}{0.5 a_{Xmax} t_c^2 - \Delta L} \tag{22}$$

If $\Delta L \ll \Delta X$, the e_{rel} is twice the relative error of the t_c estimate, and the amplification of the danger becomes more severe as ΔL approaches ΔX.

5 Conclusions

At a first approximation, and under proper conditions [1] [17], 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 [18]; (3) to use additional sources of information if they are available.

In this report we focused on the analysis of the time-to-contact estimates, presenting a specific introduction to the problem and a quantitative assessment of the performance of our technique with both synthetic and natural sequences. The preliminary results demonstrate the validity of the approach and the significance of the optic flow descriptors extracted by the Kalman Filter software module in cascade to the front-end optic flow module. A more comprehensive analysis of the meaningfulness of the first order differentials for extracting SVE will be subject of future work.

References

[1] 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.

[2] M. Subbarao. Bounds on time-to-collision and rotational component from first-order derivatives of image flow. *CVGIP*, 50(3):329–341, 1990.

[3] M. Tistarelli and G. Sandini. On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(4):401–410, 1993.

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

[5] M. Subbarao and A.M. Waxman. Closed form solutions to image flow equation for planar surfaces in motion. *CVGIP*, 36(2/3):208–228, 1986.

[6] D.N. Lee. A theory of visual control of braking based on information about time-to-collision. *Perception*, 5(4):437–459, 1992.

[7] T. Camus. Calculating time-to-contact using real-time quantized optical flow. Technical Report 14, Max-Planck-Institut, 1995.

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

[9] M. Campani and A. Verri. Motion analysis from first-order properties of optical flow. *CVGIP: Image Understanding*, 56(1):90–107, 1992.

[10] A. Giachetti and V. Torre. The use of optical flow for the analysis of non-rigid motions. *International Journal of Computer Vision*, 18(3):90–107, 1996.

[11] R. Sharma. Active vision in robot navigation: Monitoring time-to-collision while tracking. In *Proc. Int. Conf, on Intelligent Robots and Systems*, pages 2203–2208. 1992.

[12] F.G. Meyer. Time-to-collision from first-order models of the motion field. *IEEE Trans. on Robotics and Automation*, 10(6):792–798, 1994.

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

[14] J.J. Little and A. Verri. Analysis of differential and matching methods for optical flow. In *Proc. Visual Motion*, pages 173–180. 1989.

[15] B.K.P. Horn, Y. Fang, and I. Masaki. Time to contact relative to a planar surface. In *Proc. IEEE Intelligent Vehicles Symposium*, pages 68–74. 2007.

[16] Hella. Personal communication.

[17] A. Verri, M. Straforini, and V. Torre. Computational aspects of motion perception in natural and artificial systems. *Phil. Trans. R. Soc. Lond. B*, 337:429–443, 1992.

[18] 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.