Estimation of Time-to-Collision Maps from First Order Motion Models and Normal Flows *

François Meyer †, Patrick Bouthemy
IRISA/INRIA
Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France
E-mail: meyer@irisa.fr

Abstract

We address in this paper the problem of estimating time-to-collision maps involving all the objects in relative motion with respect to the camera. The approach only takes into account normal flows. Moreover we prove that first-order visual motion models are sufficient to obtain time-to-collision. Experiments have been carried out on real images to validate the performance of the method.

1 Introduction

One of the main goal of a vision system is to deliver the information necessary to interact with the environment. Robotic applications often require the complete recovery of the 3-D structure and motion. Yet it appears that many biological vision systems possess some "goal-oriented" functions which provide a fast and reliable information about a specific purpose. The determination of time-to-collision is one of these mechanisms. It offers an interesting alternative approach to the classic "motion and structure from motion" issue. Time-to-collision is the time that will elapse before the moving object and the observer will collide. Lee, [1], argues that time-to-collision is the type of information used by birds before landing. Furthermore the derivative with respect to time of this variable should afford enough information for a driver to adjust his speed with respect to the obstacles. According to Marr, [2], there exists a specific mechanism in the human visual system, designed to cause one to blink or to avoid a looming object approaching too quickly. In [3], Arkin proposes an "avoid-static-obstacle" scheme for the navigation of a robot, where no 3-D world model need to be produced, but only the information for timing action with respect to the environment. The usefulness of one of the differential invariants of the motion field – the divergence, to accomplish obstacle avoidance in a qualitative way, has been demonstrated by Nelson and Aloimonos, [4]. Time-to-collision has obvious connection with divergence, but it can be computed from this invariant alone, only if the object is not slanted (i.e. its surface is parallel to the image plane). Subbarao [5] has derived theoretical expressions of bounds on time-to-collision of a single object moving in the scene, using the first-order derivatives of the optical flow. Tistarelli and Sandini [6] have recently described an estimation of time-to-collision using a particular sensor based on polar or log-polar representation. The computation of time-to-collision for every point in the image is based on the optical flow, and thus is sensitive to noise. We address in this paper the problem of estimating time-to-collision maps involving all the objects in relative motion with respect to any classical CCD sensor. This approach does not require the estimation of optic flow fields. It only takes into account normal flows. Moreover we prove that first-order visual motion models are sufficient to obtain time-to-collision information.

2 From Kinematic Parameters to Motion Descriptors

We consider a point P on the viewed surface of a rigid object in relative motion with respect to the camera, with a translational velocity vector $T = (U, V, W)^T$, and a rotational velocity vector $\Omega = (A, B, C)^T$,(refer to Fig. 5). Let $p = (x, y)^T$ be the perspective projection in the image plane of $P = (X, Y, Z)^T$, and let $(u, v)^T$ be the 2-D velocity

*This work is supported by MRT (French Ministry of Research and Technology) in the context of the EUREKA European project PROMETHEUS, under PSA-contract VY/85241753/14/Z10

†Supported by a grant of C.N.R.S. and Brittany County Council

0-8186-2910-X/92 $3.00 © 1992 IEEE

vector of p. Let f be the focal length. Well known expressions relating the 2-D motion field $(u,v)^T$ to the depth Z and the kinematics parameters can be derived. Let us consider the tangent plane to the patch of surface around P :

$$Z = Z_0 + \gamma_1 X + \gamma_2 Y \tag{1}$$

On the other hand, we can consider a first-order development of the 2-D motion field as follows :

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} + \begin{bmatrix} \frac{div+h_1}{2} & \frac{h_2-rot}{2} \\ \frac{rot+h_2}{2} & \frac{div-h_1}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{2}$$

This representation has the advantage that the variables div, rot, h_1 and h_2 correspond to the divergence, rotational, and the two hyperbolics vector subfields that can be easily interpreted, [7]. We have shown in [7], that keeping first-order terms only in the expression relating optical flow to 3-D kinematics parameters, yields the following relations for the six variables :

$$\begin{aligned} a &= f\left(\tfrac{U}{Z_0} - B\right) & b &= f\left(\tfrac{V}{Z_0} - A\right) \\ div &= -2\tfrac{W}{Z_0} - \gamma_1 \tfrac{U}{Z_0} - \gamma_2 \tfrac{V}{Z_0} & h_1 &= -\gamma_1 \tfrac{U}{Z_0} + \gamma_2 \tfrac{V}{Z_0} \\ rot &= 2C - \gamma_1 \tfrac{V}{Z_0} + \gamma_2 \tfrac{U}{Z_0} & h_2 &= -\gamma_1 \tfrac{V}{Z_0} - \gamma_2 \tfrac{U}{Z_0} \end{aligned} \tag{3}$$

The partial derivatives with respect to time of the four differential invariants : div, rot, h_1, h_2 will provide us with another set of equations. Assuming a constant translational velocity and a constant rotational velocity within the considered time interval, we have :

$$\begin{aligned} div' &= -\gamma_1' \tfrac{U}{Z_0} - \gamma_2' \tfrac{V}{Z_0} - \tfrac{Z_0'}{Z_0} div \\ rot' &= -\gamma_1' \tfrac{V}{Z_0} + \gamma_2' \tfrac{U}{Z_0} - \tfrac{Z_0'}{Z_0}(-\gamma_1' \tfrac{V}{Z_0} + \gamma_2' \tfrac{U}{Z_0}) \\ h_1' &= -\gamma_1' \tfrac{U}{Z_0} + \gamma_2' \tfrac{V}{Z_0} - \tfrac{Z_0'}{Z_0} h_1 \\ h_2' &= -\gamma_1' \tfrac{V}{Z_0} - \gamma_2' \tfrac{U}{Z_0} - \tfrac{Z_0'}{Z_0} h_2 \end{aligned} \tag{4}$$

with :

$$\begin{aligned} \tfrac{Z_0'}{Z_0} &= \tfrac{W}{Z_0} - \gamma_1 \tfrac{U}{Z_0} - \gamma_2 \tfrac{V}{Z_0} - (\gamma_1 B - \gamma_2 A) \\ \gamma'_1 &= -B - \gamma_2 C - \gamma_1(\gamma_1 B - \gamma_2 A) \\ \gamma'_2 &= A + \gamma_1 C - \gamma_1(\gamma_1 B - \gamma_2 A) \end{aligned} \tag{5}$$

We advocate the use of a first-order model of the motion field for two reasons. Firstly the higher-order terms are generally of quite low magnitude. Secondly the estimation of these terms is usually less robust than the estimation of first-order ones, [7, 8]. Therefore the determination of time-to-collision is certainly more stable using only first-order terms, and the quantitative difference is quite negligible when considering higher-order terms.

3 Derivation of Time-to-Collision

3.1 Time-to-Collision Expression

Time-to-collision is defined as the time that will elapse before the object and the sensor are in contact due to the instantaneous relative translation along the optical axis. We can compute it for every point of the image plane $(x,y)^T$, projection of a point $(X,Y,Z)^T$ of the viewed surface of the object, (See Fig 5).Using (1), its expression is :

$$\tau_c(x,y) = -Z/W = -Z_0/\left[W(1-\gamma_1 x - \gamma_2 y)\right] \tag{6}$$

${\tau_c}^{-1}$ is the sum of two terms. The first one $-\frac{W}{Z_0}$ represents the inverse of time-to-collision of the virtual point at the intersection of the tangent plane of the surface patch, and the optical axis. The second term provides the necessary correction to estimate time-to-collision on the surface patch around P, that we approximate with the tangent plane to P. We need to retrieve $\frac{W}{Z_0}, \gamma_1$ and γ_2.From relation (3) we notice that $-\frac{Z_0}{W}$ depends on divergence (div), but is usually not equal to $\frac{2}{div}$. Systems (3) and (4) provide us with ten equations in eight unknowns : $A, B, C, \frac{U}{Z_0}, \frac{V}{Z_0}, \frac{W}{Z_0}, \gamma_1, \gamma_2$. We are interested only in the three last variables.

3.2 Case of an Object Surface Parallel to the Image Plane

In the simple case where the local tangent plane remains parallel to the image plane (i.e. $\gamma_1 = \gamma_2 = 0, Z = Z_0$), time-to-collision depends only on the divergence. It follows from (6) and (3) :

$$\tau_c(x,y) = -Z_0/W = 2/div \tag{7}$$

for all (x,y) within the region corresponding to the projection, in the image plane, of the moving object.

3.3 Case of Planar Motion

We consider here the case of objects and sensor undergoing a relative planar motion. This assumption is realistic for a number of outdoor and indoor scenes. It includes the situation where a vehicle equipped with a camera is moving on a planar ground. This assumption allows us to derive an explicit formula for time-to-collision. Planar motion only implies : $V = 0$. From (3), (4), (5), and after some mathematical developments we obtain :

$$\gamma_1 = \frac{1}{a}\left(\frac{W}{Z_0} + \frac{div' - h_1'}{div - h_1}\right) \qquad \gamma_2 = \frac{h_2}{h_1}\gamma_1 \tag{8}$$

$$\frac{W}{Z_0} = \frac{1}{2}(h_1 - div) \quad (9)$$

It follows from (6), (8) and (9) :

$$\tau_c(x,y) = \left[\frac{1}{2}(div - h_1) + \frac{1}{2a}\left\{\frac{1}{2}(h_1 - div)^2 + (h_1' - div')\right\}\left(x + \frac{h_2}{h_1}y\right)\right]^{-1} \quad (10)$$

There are two situations where two solutions exist.

- $\frac{U}{Z_0} = B$. In this case we can only obtain $\left[\frac{U}{Z_0}\right]^2$.

$$\tau_c(x,y) = \left[\frac{1}{2}(div - h_1)\left\{1 \pm \frac{h_1 x + h_2 y}{\sqrt{\frac{h_1}{2}(div - 3h_1) - h_1'}}\right\}\right]^{-1}$$

- $\gamma_1 = 0$. We get two different solutions for $\frac{U}{Z_0}$.

$$\tau_c(x,y) = \left[\frac{1}{2}(div - h_1)\left\{1 + \frac{h_1 x + h_2 y}{\frac{a}{2} \pm \sqrt{\frac{a^2}{4} - h_1'}}\right\}\right]^{-1}$$

In the two above cases, we can use the continuity of time-to-collision with respect to time to remove the instantaneous ambiguity between the two possible terms.

There are two situations where the only information available from first-order motion model is : $\frac{Z_0}{W}$. If $U = 0$ and $B \neq 0$, γ_1 only can be retrieved : $\gamma_1 = \frac{1}{a}\left(\frac{div'}{div} - \frac{div}{2}\right)$. If $U = 0$ and $B = 0$, neither γ_1 nor γ_2 can be obtained. One the one hand we have in these two cases : $-\frac{Z_0}{W} = \frac{2}{div}$. On the other hand (6) implies that τ_c can be approximated with $2/div$ if the object is close to the optical axis, (e.g. if the camera is directed toward the object), or if its surface is not too slanted. Consequently, $\frac{2}{div}$ provides here a good approximation of τ_c.

3.4 Case of General Motion

If the relative motion of the objects in the scene does not fall in the two previous cases, time-to-collision is obtained by the system formed by the set of equations (3), (4) and (5). The eight unknowns are $\frac{U}{Z_0}, \frac{V}{Z_0}, \frac{W}{Z_0}, A, B, C, \gamma_1, \gamma_2$. The system is sufficiently constrained, and it is possible to determine the solution from numerical methods for non linear algebraic equations.

4 Motion Descriptors Estimation

We consider scenes containing multiple moving objects. Before attempting to obtain the time-to-collision of the corresponding projections of these objects, we must delineate each region. This involves a motion-based segmentation stage. The approach relies on the spatio-temporal variations of the intensity function while making use of 2-D first-order motion models, [7]. It does not require neither explicit 3-D measurements, nor the estimation of optic flow fields. Moreover the segmentation also manages to link two successive partitions in time.

4.1 Estimation of the First Order Motion Parameters

Within the segmentation stage a maximum-likelihood estimation of the vector : $\theta = [a, b, div, rot, hyp1, hyp2]$ is performed for each extracted region. The following "data-model adequacy" random variable is considered :

$$\varepsilon_\theta(x,y) = \omega_\theta(x,y).\nabla I(x,y) + I_t(x,y) \quad (11)$$

where ω_θ is the 2-D affine model of the velocity given by (2). It relies somehow on the well-known image flow constraint equation : $\omega.\nabla I + I_t = 0$ Therefore the estimation of the motion descriptors involves only the computation of the normal flows : that is the flow parallel to the image gradient.

However, when the image sequence involves large displacements it appears that this method yields underestimated values. For this reason, we have developed a multiresolution method to provide more reliable motion descriptors. The idea is to estimate the six motion parameters with a classic "coarse-to-fine" strategy where a rough estimate of the parameters is obtained at the lowest resolution. The estimate is subsequently refined using the higher resolution images. We build two low pass Gaussian pyramids for each image at time t and $t + \delta t$ and a pyramid for the segmented image. Delineations of regions are provided at each level by the pyramid of the segmented image.

First we estimate at the lowest resolution level the six motion parameters of each region in the image, with (11). Then we refine the estimate. For a given level, the first order velocity field estimated at the coarser level is projected on the current level and accounts for an initial displacement. The refinement of the motion descriptors is based on another zero-mean Gaussian random variable :

$$\xi_{\Delta\theta^l}(x,y) = \nabla I((x,y) + 2\omega_{\theta^{l+1}}.\delta t, t + \delta t).\omega_{\Delta\theta^l}(x,y) + \frac{I((x,y) + 2\omega_{\theta^{l+1}}.\delta t, t + \delta t) - I(x,y,t)}{\delta t}$$

where δt is the time step between two successive frames, θ^{l+1} is the vector of the motion parameters computed at level $l+1$, and $\omega_{\Delta\theta^l}$ is the residual of the velocity model to be computed at level l. $\xi_{\Delta\theta^l}$ has a similar expression to ε_θ. $\Delta\theta^l$ is obtained using the same least-squares estimation scheme.

4.2 Recursive Estimation

The multiresolution method provides us with instantaneous measurements of the six components of θ. These measurements are generally corrupted by noise, hence we need a recursive estimator to convert observation data into accurate estimates. Furthermore stable derivatives with respect to time of these variables can then be obtained. In the absence, in the general case, of any explicit simple analytical function describing the evolution of the variables with respect to time, we use a Taylor-series expansion of each function. We obtain a linear dynamic system which models the temporal evolution of each of the motion descriptors. A standard Kalman filter generates recursive estimates of each of the descriptor. Details about the method can be found in [9].

5 Results and Conclusion

The computation of time-to-collision for real data must take into account the video rate. We have, real time-to-collision $= \frac{1}{N}\tau_c$; where N is the number of frames per second ($N = 25$).

The first experiment was performed on real images to compare the measurements given by our method with the ground truth determined by independent means. The sequence *robot*, Fig. 1, has been acquired with an experimental cell constituted by a camera mounted on the end effector of a 6 d.o.f robot. The camera is undergoing a planar motion toward a poster pined on a wall. The constant velocity of the camera, in the camera coordinate system (refer to Fig. 5), is $U = 125$ mm/s, $V = 0$, and $W = 250$ mm/s. Initially the camera is at 70 cm from the wall, and slightly slanted (17 deg), i.e. the image plane is not parallel to the wall. Expected time-to-collision of a tracked point, located far from the optical axis, is computed analytically with (6). It is linear with respect to frame number. The expected time-to-collision and the experimental time-to-collision, computed with (10), are plotted in Fig. 2. $2/div$ is also plotted on the same figure to illustrate the improvement gained with our method when computing time-to-collision far away from the optical axis. This experiment validates the explicit formula that gives τ_c with respect to the first order motion descriptors.

The second experiment illustrates the significance of time-to-collision to characterize complex dynamic evolution during a long time interval. The sequence *van* is composed of 55 real images acquired with a camera attached to a moving car. The car is driving behind a van (Fig. 3). The relative motion of the car and the van is an axial translation along the optical axis. We plot the value of τ_c at the center of gravity of the viewed surface of the van. This surface remains parallel to the image plane. In the first part of the sequence (Fig.3, images 5 to 18) the van and the car slow down because of a red traffic light. The car is getting closer to the van, but the relative velocity decreases. Thus time-to-collision increases, (Fig. 4). In the image 18 the relative velocity is equal to zero and the time-to-collision is infinite. Then the traffic light turns green, and the two vehicles start again. The van accelerates faster than the car and moves away from it (images 18 to 60), hence time-to-collision gets negative. In this case τ_c can not be interpreted as a time-to-collision, since there will not be any collision ; but it nevertheless delivers information about the rate at which the van is moving away from the car. It increases first, then it decreases and is asymptotically linear, (Fig. 4). It should be pointed out that the time-to-collision curve has the same shape as the theoretical curve in the case of a motion with constant acceleration in the opposite direction of the initial velocity. The results illustrate the efficiency of the approach to deliver a reliable and relevant information for visual navigation.

We have explored an original approach to dynamic scene analysis. We have presented a new method which delivers dense time-to-collision maps. For many applications, such as industrial robot guidance, obstacle avoidance, the use of time-to-collision is relevant. The approach recalls the suggestions expressed by Thompson, [10], who advocated to study what information is needed to accomplish specific tasks, such as obstacle avoidance. The estimation of time-to-collision is performed directly from the image sequence and relies on first-order motion descriptors. It does not require the recovery of the structure and motion of the moving objects. Experiments have been carried out on real images to validate the performance of the method. The promising results obtained indicate the strength of the approach.

References

[1] D. N. Lee. The optical flow field : the foundation of vision. *Phil. Trans. R. Soc. of London*, Vol. B 290:169–179, 1980.

[2] D. Marr. *Vision*. Freeman, 1982.

[3] R. C. Arkin. The impact of cybernetics on the design of a mobile robot system : a case study. *IEEE Trans. SMC*, Vol.20, No.6:1245–1257, Dec. 1990.

[4] C. Nelson, J. Aloimonos. Obstacle avoidance using flow field divergence. *IEEE Trans.PAMI*, Vol.11, No.10:1102–1106, Oct. 1989.

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

[6] M. Tistarelli and G. Sandini. Direct estimation of time-to-impact from optical flow. *Proc. of the IEEE Workshop on Visual Motion, Princeton*, pp 226–233, Oct. 1991.

[7] E. François, P. Bouthemy. Multiframe-based identification of mobile components of a scene with a moving camera. *Proc. CVPR, Hawaii*, pp 166–172, June 1991.

[8] S. Negahdaripour, S. Lee. Motion recovery from image sequences using first-order optical flow information. *Proc. of the IEEE Workshop on Visual Motion, Princeton*, pp 132–139, Oct. 1991.

[9] F. Meyer, P. Bouthemy. Region-based tracking in an image sequence. *Proc. of ECCV-92, S. Margherita Ligure, Italy*, May 1992.

[10] W.B. Thompson, J.K. Kearney. Inexact vision. *Proc. of IEEE Workshop on Motion: Representation and Analysis*, pp 791–794, May 1986.

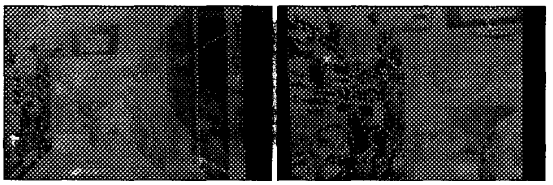

Figure 1 : Sequence *robot* time 1 and 32

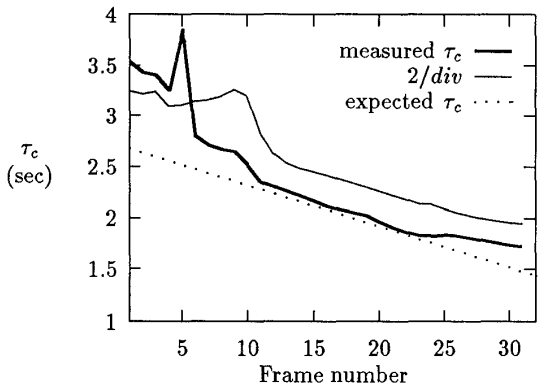

Figure 2 : Sequence *robot*, evolution of τ_c

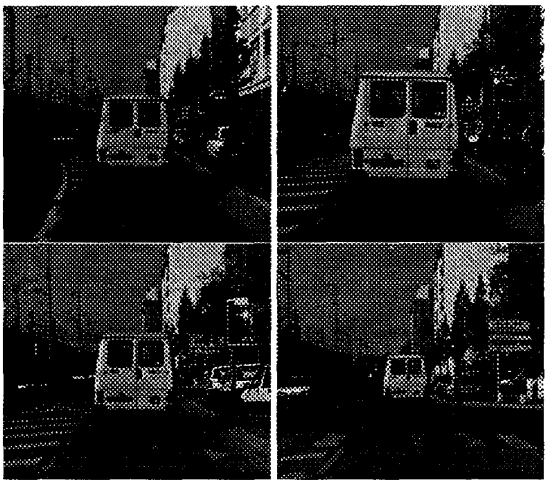

Figure 3 : Sequence *van*, from top to down and left to right, time 5, 18, 36 and 60

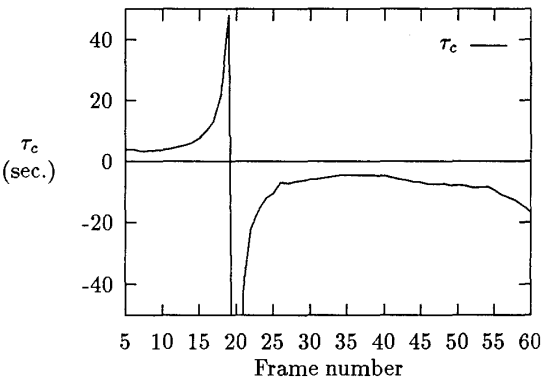

Figure 4 : Sequence *van* , evolution of τ_c

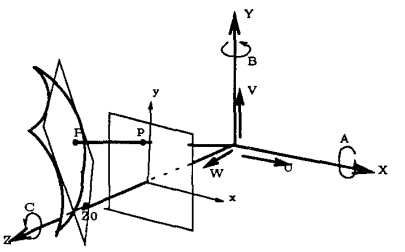

Figure 5 :Camera coordinate system