existing models of the quality of range data from stereo vision and AM-CW LADAR, then used these models to derive a new model for the quality of a simple obstacle detection algorithm. This model predicts the probability of detecting obstacles and the probability of false alarms, as a function of the size and distance of the obstacle, the resolution of the sensor, and the level of noise in the range data.

We evaluated these models experimentally using range data from a total of 850 stereo image pairs of a gravel road with known obstacles at several distances. For stereo pairs at both 60×64 and 120×128-pixel resolution (reduced down from an original 480×512 by image pyramid transformation), we found that the modal value of the sample standard deviation of disparity was less than 0.05 pixels. For the field of view (40 degrees) and baseline (35 cm) used by the stereo system, this corresponds to a standard deviation of range measurements of between 25 and 50 cm at a distance of 20 m. We also found spatially correlated errors and notable systematic errors in the range data. For the quality of obstacle detection, our experimental results for detection and false alarm probabilities matched fairly well with predictions of the model. As a concrete example, we applied the obstacle detectability model to estimate detection and false alarm probabilities given the sensor and actuator characteristics of an existing military robotic vehicle, for a vehicle velocity of 10 km/h and an obstacle height of 30 cm. At this speed, the vehicle would have to look 8.4 m ahead in order to see the obstacle in time to stop. At that distance, for 60×64 resolution the detection probabilty was 0.89, but the expected number of false alarms over a reasonably sized image window was 9. This is undesirably high. For 120×128 resolution, the detection probability rose to 0.99 and the expected number of false alarms dropped to 0.03; these values are reasonable. Overall, these predictions are roughly in accord with our qualitative impressions from operating the vision system on the vehicle in question.

We draw two main conclusions from the results. First, the statistical modeling effort shows promise for being useful as a practical design tool, because it can be used to predict the reliability that will be achieved for specific resolutions or thresholds. Second, our experimental evaluation methodology gives valuable, quantitative measures of actual obstacle detection performance, independent of the predictive modeling issue. The experimental results also revealed artifacts in the range data that suggest areas for future algorithm research.

In the future, we plan to refine the statistical performance modeling and evaluation work begun here and to use it to 1) measure progress on stereo vision algorithms, 2) compare the performance of stereo vision with daylight cameras to other range image sensors, and 3) evaluate the performance of more elaborate obstacle detection algorithms. Also, we believe that quantitative, statistical performance models like those presented here are essential ingredients in algorithms for intelligent control of perception, or "active perception." We have begun to apply these performance models to perception control in [2].

References

[1] L. H. Matthies, "Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation," *Int. J. Comput. Vision*, vol. 8, no. 1, pp. 71–91, 1992.

[2] P. Grandjean and L. Matthies, "Perception control for obstacle detection by a cross-country rover," in *Proc. IEEE Int. Conf. Robotics Automat.*, May 1993.

[3] A. Kelly, A. Stentz, and M. Hebert, "Terrain map building for fast navigation on rugged outdoor terrain," in *Proc. SPIE Conf. Mobile Robots*, SPIE, November 1992.

[4] E. D. Dickmanns and B. Mysliwetz, Recursive 3-D road and relative ego-state recognition, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 2, pp. 199–213, 1992.

[5] D. Nitzan, A. Brain, and R. Duda, "The measurement and use of registered reflectance and range data in scene analysis," *Proc. IEEE*, vol. 65, no. 2, pp. 206–220, 1977.

[6] C. C. Slama, Ed., *Manual of Photogrammetry*, 4th ed. Falls Church, VA: American Society of Photogrammetry, 1980.

[7] D. B. Gennery, "Modelling the environment of an exploring vehicle by means of stereo vision," Ph.D. thesis, Stanford University, June 1980.

[8] N. Ayache and O. D. Faugeras, "Maintaining representations of the environment of a mobile robot," in *Int. Symp. Robotics Res. 4*. Cambridge, MA: MIT Press, 1987.

[9] L. H. Matthies and S. A. Shafer, "Error modeling in stereo navigation," *IEEE Trans. Robotics Automat.*, vol. 3, no. 3, pp. 239–248, 1987.

[10] M. Hebert and E. Krotkov, "3D measurements from imaging laser radars: How good are they?" *Image and Vision Computing*, vol. 10, no. 3, pp. 170–178, 1992.

[11] L. H. Matthies, "Toward stochastic modeling of obstacle detectability in passive stereo range imagery," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition*, June 1992.

[12] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*. New York: John Wiley and Sons, 1968.

[13] M. F. Reiley, D. C. Carmer, and W. F. Pont, "3-D laser radar simulation for autonomous spacecraft landing," in *Proc. SPIE Int. Symp. High Power Lasers*. SPIE, January 1991.

[14] M. Okutomi and T. Kanade, "A locally adaptive window for signal matching," Tech. Rep. CMU-CS-90-178, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, October 1990.

[15] L. H. Matthies, R. Szeliski, and T. Kanade, "Kalman filter-based algorithms for estimating depth from image sequences," *Int. J. Comput. Vision*, vol. 3, no. 3, pp. 209–236, 1989.

Time-to-Collision from First-Order Models of the Motion Field

François G. Meyer

***Abstract*— Time-to-collision provides vital information for obstacle avoidance and for the visual navigation of a robot. The original contribution of this paper is to demonstrate with sequences of real images that time-to-collision can be robustly and accurately recovered with a single calibrated camera, using first order models of the motion field.**

I. Introduction

One of the main goals of a vision system is to deliver the information necessary to interact with the environment in real time. Robotic applications often require the complete recovery of 3-D structure and motion. Yet it appears that many biological vision systems possess some "goal-oriented" mechanisms that provide fast and reliable knowledge for a specific task. The determination of

Manuscript received May 28, 1993; revised April 5, 1994. This work was supported in part by a C.N.R.S. and Région Bretagne fellowship, by MRT in the context of the EUREKA European project PROMETHEUS, under PSA Contract VY/85241753/14/Z10, and in the context of the French project Orasis.

The author was with IRISA/INRIA, Rennes, France. He is now with the Departments of Mathematics and Diagnostic Radiology, Yale University, New Haven, CT 06520 USA.

IEEE Log Number 9405739.

time-to-collision is one of these mechanisms. Time-to-collision is the time needed for the observer to reach the object, if the instantaneous relative velocity along the optical axis is kept unchanged. Measuring distances in temporal units may be an advantage when performing visual tasks such as landing an aircraft or braking [9]. The estimation of time-to-collision should allow an unmanned vehicle, navigating in an outdoor environment, to avoid hazardous obstacles that could cause a crash. Subbarao [16] 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 [18] have recently described the estimation of time-to-collision using a special sensor based on polar or log-polar representation. The computation of time-to-collision for every point in the image is based on the derivatives of the optical flow, and thus is sensitive to noise. Even more recently, Cipolla and Blake [6] have proposed estimating the parameters of an affine model of the motion field of a tracked object. With some additional knowledge the method can yield estimates of time-to-collision. The approach requires the temporal derivatives of the moments of area of the moving region, and thus is likely to be quite sensitive to even a small partial occlusion of the tracked object.

In this paper we investigate the problem of estimating the time-to-collision of a moving object from a monocular sequence of images. We assume that the contour of the object of interest can be detected. We propose to exploit the 3-D information captured by the first order coefficients of the motion field, and, by their temporal derivatives. To achieve this we apply the theoretical results of Subbarao and Waxman [15], [17] who showed that the unknown 3-D parameters could be expressed from one of the solutions of a fifth degree equation. The original contribution of this paper is to demonstrate with sequences of real images that time-to-collision, which provides vital 3-D information for the outdoor navigation of unmanned vehicles and mobile robots, can be robustly and accurately recovered. A novel method to estimate the first order coefficients of the flow and their temporal derivatives has been developed. This method makes it possible to apply the theoretical analysis of Subbarao and Waxman in realistic situations. The estimation of the first order coefficients of the flow exploits a multiresolution scheme and temporal continuity of motion using a Kalman filter. We focus our attention on the case when one component of the 3-D translational velocity of the moving object is equal to zero, since we can obtain in that case closed form solutions of the 3-D parameters. Such a situation includes planar motion, which is realistic for a number of indoor and outdoor situations (such as a mobile robot, moving on the ground, equipped with a camera that can rotate with three degrees of freedom). The 3-D translational velocity of the object can only be recovered up to an unknown scale factor from a monocular sequence of images. This well known problem of indeterminate depth scale is a consequence of projective geometry. However, measured in temporal units, the depth can be uniquely recovered with a single calibrated camera.

This paper is organized as follows. In the next section, we introduce the problem, and we give the closed form solutions of the 3-D orientation and motion parameters that enable us to recover time-to-collision. Section III describes the multiresolution estimation of the first order coefficients of the motion field. Results of experiments conducted on sequences of synthetic and real images are presented in Section IV.

II. Problem Statement and Approach

We assume that the projection onto the image of an object moving in the scene, has been detected. We can use a motion-based segmentation algorithm, previously developed in the laboratory [3] that divides images into distinct objects moving with different velocities. Rather than segmenting the whole image at each instant, a region tracking algorithm [1], [2] could also be used to track the contour of an object. Our method would then estimate the time-to-collision of the tracked object. In fact, experiments show that the contour of the region need not be known accurately.

A. Camera Model and Problem Statement

We use the pinhole camera model (see Fig. 1).The optical center of the camera, **O**, is the origin of the coordinate system. The optical axis is the Z axis. The camera calibration provides us with estimates of the intrinsic parameters: (x_c, y_c) the intersection of the optical axis with the image plane, and (f_x, f_y) the product of the focal length with the two scale factors along the x axis and y axis respectively. In this issue we focus our attention on a rigid object in relative motion with respect to the camera, with a translational velocity $\mathbf{T} = (U, V, W)^T$, and a rotational velocity $\mathbf{\Omega} = (A, B, C)^T$. Let $\mathbf{P} = (X, Y, Z)^T$ be a point on the viewed surface of the object. Its instantaneous 3-D velocity is $\dot{\mathbf{P}} = \mathbf{T} + \mathbf{\Omega} \times \mathbf{OP}$ (see Fig. 1). Let $(x, y)^T$ be its perspective projection on the image plane. We have

$$\frac{x - x_c}{f_x} = \frac{X}{Z} \quad \text{and} \quad \frac{y - y_c}{f_y} = \frac{Y}{Z} \tag{1}$$

The patch of surface can be locally described by a first order Taylor series expansion:

$$Z = Z_0 + Z_X X + Z_Y Y \tag{2}$$

This relation is a valid approximation if the curved surface is smooth and the depth of the object Z is large compared to the depth variations. From (1), (2), and using the expression for $\dot{\mathbf{P}}$, we obtain the well known equations [10] for the motion field:

$$\begin{aligned} u &\triangleq \dot{x} = f_x\left[u_0 + u_x\tilde{x} + u_y\tilde{y} + u_{xy}\tilde{x}\tilde{y} + u_{xx}\tilde{x}^2\right] \\ v &\triangleq \dot{y} = f_y\left[v_0 + v_x\tilde{x} + v_y\tilde{y} + v_{xy}\tilde{x}\tilde{y} + v_{yy}\tilde{y}^2\right] \end{aligned} \tag{3}$$

with

$$\tilde{x} = \frac{x - x_c}{f_x} \quad \tilde{y} = \frac{y - y_c}{f_y}$$

and

$$\begin{aligned} u_0 &= \frac{U}{Z_0} + B \\ u_x &= -Z_X\frac{U}{Z_0} - \frac{W}{Z_0} \\ u_y &= -Z_Y\frac{U}{Z_0} - C \\ v_0 &= \frac{V}{Z_0} - A \\ v_x &= -Z_X\frac{V}{Z_0} - C \\ v_y &= -Z_Y\frac{V}{Z_0} - \frac{W}{Z_0} \end{aligned} \tag{4}$$

$$\begin{aligned} u_{xy} &= Z_Y\frac{W}{Z_0} - A \\ u_{xx} &= Z_X\frac{W}{Z_0} + B \\ v_{xy} &= Z_X\frac{W}{Z_0} + B \\ v_{yy} &= Z_Y\frac{W}{Z_0} - A \end{aligned} \tag{5}$$

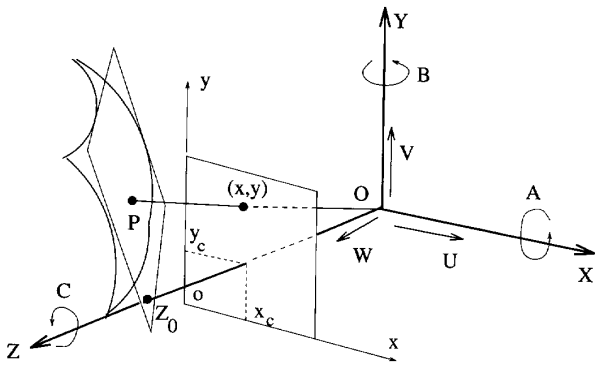


Fig. 1. Camera coordinate system.

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. Let $\tau_c(x,y)$ be the time-to-collision of the point $(X,Y,Z)^T$, estimated at location (x,y) in the image. We have $\tau_c(x,y) = -Z/W$. From (1) and (2) we get

$$\tau_c(x,y) = -\frac{Z_0}{W}\frac{1}{\left(1 - Z_X\frac{x-x_c}{f_x} - Z_Y\frac{y-y_c}{f_y}\right)} \tag{6}$$

In order to obtain time-to-collision with (6) we will recover the 3-D orientation parameters Z_X, Z_Y, and the motion parameter $\frac{W}{Z_0}$, from the first-order local behavior of the motion field and its temporal variation.

B. Orientation and Motion Recovery

The motion field (3) is quadratic. However, second order components are usually quite small [3], [5], [16]. Since their contribution to the total field is negligible, their estimation tends to be unstable and inaccurate [13], [16]. We advocate therefore the use of the first order coefficients of the motion field. We consider the case when V is equal to zero. This allows us to obtain closed form solutions of the 3-D parameters. This assumption is realistic for a number of situations, which include the case when a vehicle is moving on a planar ground. If $V = 0$, we have seven unknowns: $\frac{U}{Z_0}$, $\frac{W}{Z_0}$, Z_X, Z_Y, A, B, and C, (we note that because of the scale factor ambiguity we can only hope to recover the translational velocity up to a scale factor $\frac{1}{Z_0}$). From (4) we obtain six equations that relate the first order coefficients of the flow to the unknown 3-D motion and orientation parameters. Previous approaches [10], [19] have used the second order spatial derivatives (5) of the motion field in order to further constrain the system (4). The temporal variation (given by the partial derivative with respect to time) of a zero order coefficient of the flow (u_0) offers a more robust alternative to obtain one additional independent equation as proposed in [15], [17].

For the sake of brevity we only present the theoretical expressions of the solutions derived by Subbarao and Waxman in [15], [17]. From (4) we obtain [15], [17] after some algebra:

$$\frac{U}{Z_0} = \frac{\dot{u_0} + u_0(u_x - v_y) + v_0(v_x + u_y)}{v_y}$$

$$\frac{W}{Z_0} = -v_y$$

$$Z_X = \frac{(v_y - u_x)v_y}{\dot{u_0} + u_0(u_x - v_y) + v_0(v_x + u_y)}$$

$$Z_Y = -\frac{(v_x + u_y)v_y}{\dot{u_0} + u_0(u_x - v_y) + v_0(v_x + u_y)}$$

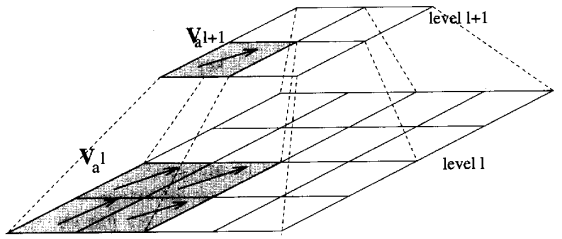


Fig. 2. Multiresolution estimation of the affine model of the 2-D field. The affine model, $v_{a^{l+1}}$, of the motion field at level $l+1$ is projected onto level l of the pyramid, and accounts for an initial displacement. We calculate an incremental estimate, and obtain v_{a^l}.

with

$$\dot{u_0} = -\frac{U}{Z_0}\left(\frac{W}{Z_0} - Z_X(B + \frac{U}{Z_0}) + Z_Y(A - \frac{V}{Z_0})\right)$$

There are two cases where we cannot recover the complete set of parameters. If $W = 0$, we can not recover Z_X, Z_Y, U. If $U = 0$, all the motion parameters can be found, but the orientation parameters Z_X and Z_Y can not be recovered [13], [15], [17].

III. Multiresolution Motion Estimation

We need to estimate the first order coefficients of the motion field $u_0, v_0, u_x, v_x, u_y, v_y$, and $\dot{u_0}$. The authors in [15], [17] have assumed that the first order coefficients could be obtained by computing the partial derivatives of the motion field. This approach is expected to yield noise corrupted coefficients. Indeed, estimates of the motion field obtained from standard methods are usually noisy, and taking derivatives of the flow will even further increase the amount of noise. Our approach does not tackle the intricate issue of estimating precise motion fields, but rather directly estimates a first order (or affine) model of the motion field, with a multiresolution approach. We assume that the depth of the object of interest is large compared to its distance from the optical axis, thus we have $\frac{x-x_c}{f_x} \ll 1$, and $\frac{y-y_c}{f_y} \ll 1$. This assumption is realistic for outdoor scenes where objects are not close to the camera. Under the same assumptions, scaled orthographic projection ("weak projection") is a good approximation to perspective projection. It also results in an affine motion field. A first order expansion of the quadratic flow can be derived around (x_c, y_c):

$$\begin{cases} u = a_1 + a_2 x + a_3 y \\ v = a_4 + a_5 x + a_6 y \end{cases} \tag{7}$$

with

$$\begin{aligned} a_1 &= f_x\left[u_0 - u_x\frac{x_c}{f_x} - u_y\frac{y_c}{f_y}\right] \\ a_2 &= u_x \\ a_3 &= \frac{f_x}{f_y}u_y \\ a_4 &= f_x\left[v_0 - v_x\frac{x_c}{f_x} - v_y\frac{y_c}{f_y}\right] \\ a_5 &= \frac{f_y}{f_x}v_x \\ a_6 &= v_y \end{aligned} \tag{8}$$

We estimate the vector of parameters $\mathbf{a} = [a_1, a_2, a_3, a_4, a_5, a_6]$, of this model with a multiresolution scheme. Given the intrinsic parameters of the camera, namely f_x, f_y, x_c, y_c, we can then directly obtain from (8) estimates of the first order coefficients of the flow

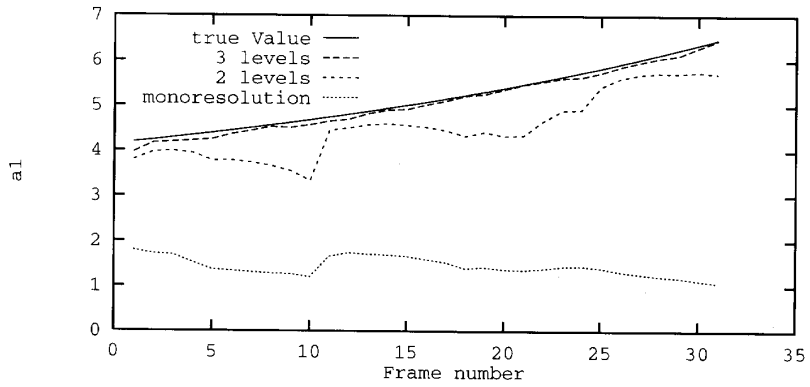
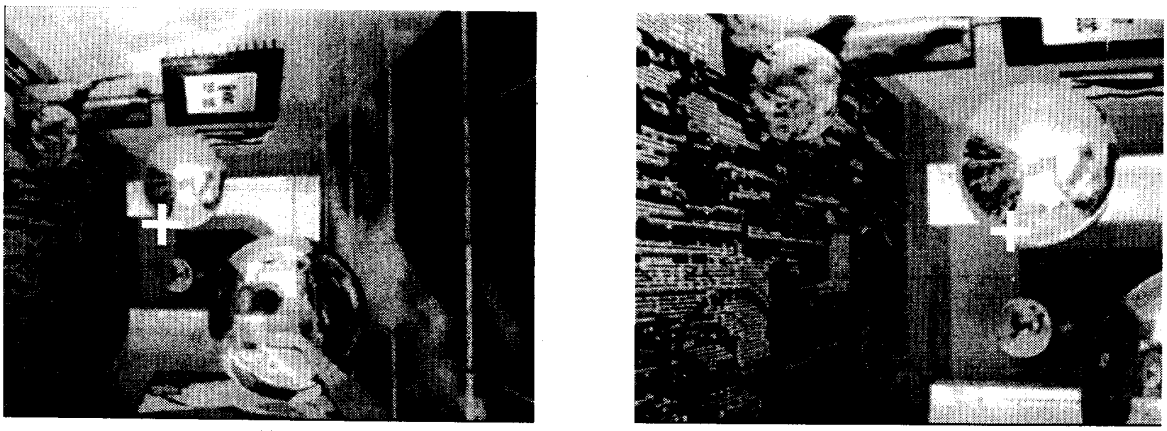


Fig. 3. Poster sequence: true value, multiresolution estimate for different pyramid levels, and monoresolution estimate of the motion parameter $a_1(t)$.



(a) (b)

Fig. 4. Poster sequence, with the tracked point displayed as a plus sign, at time t_1 (a) and t_{32} (b).

$u_0, v_0, u_x, v_x, u_y, v_y$. We recall the multiresolution scheme that is described in [11], [12]. The idea is to estimate the motion parameters with a "coarse-to-fine" strategy. We build two low pass Gaussian pyramids [4] for each image at time t and $t + \delta t$. A rough estimate of the vector of parameters $\hat{\mathbf{a}}^L$ is obtained at the lowest resolution level L. We use the well-known optical flow constraint equation [8] that relates the motion field $\mathbf{v}$ with the spatial gradient of the image intensity $\nabla \mathbf{I}$, and with the partial derivative with respect to time of the image intensity I_t

$$\nabla \mathbf{I} \cdot \mathbf{v} + I_t = 0 \tag{9}$$

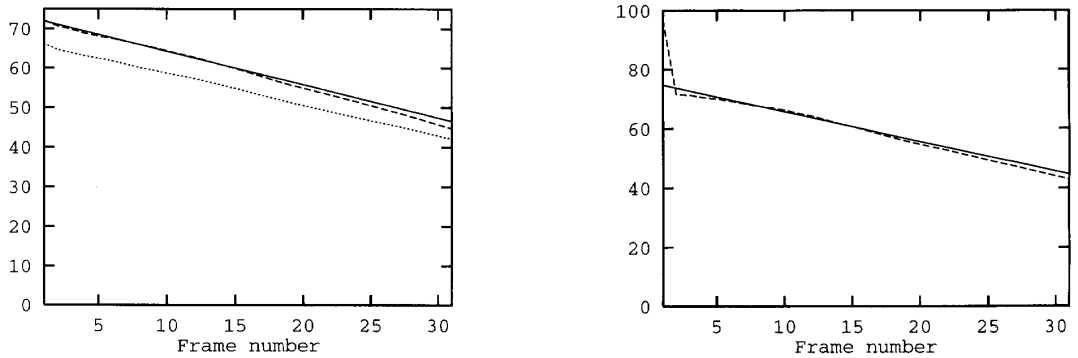
Let $\mathbf{v}_{\mathbf{a}^L}(\mathbf{p}^L)$ be the affine model of the 2-Dmotion field at location $\mathbf{p}^L$, given by (7). We substitute this model of the motion field into (9), and obtain

$$\nabla \mathbf{I}(\mathbf{p}^L, t) \cdot \mathbf{v}_{\mathbf{a}^L}(\mathbf{p}^L) + I_t(\mathbf{p}^L, t) = 0 \tag{10}$$

(10) is linear with respect to the parameters of the affine model ${a_i}^L$, $i = 1, .., 6$. If we consider this equation for all the points within the region, we obtain an over-constrained system of linear equations. Least-squares estimates of $\mathbf{a}^L$ can thus be obtained. The estimate $\hat{\mathbf{a}}^L$ is then subsequently refined using the higher resolution images, (see Fig. 2). Fig. 3 illustrates the behavior of the multiresolution algorithm with a sequence of real data, the sequence *poster* (see Fig. 4). Fig. 3 shows the true value of the parameter $a_1(t)$ and the estimate obtained with different pyramid levels. The magnitude of the motion is significant (4 to 6 pixels per frame), and we need three levels to accurately estimate the motion parameters. Rather than relying on the precise location of the points on the contour, the method exploits the spatio-temporal variation of the intensity function inside the region. Therefore the boundary of the region need not be recovered accurately, as we have demonstrated with sequences of real images. A temporal filtering of each coefficient is then performed, yielding smooth estimates of the motion parameters and their temporal derivatives over time. A Taylor-series expansion of each parameter $a_i(t), i = 1, \cdots, 6$ yields a local approximation of the temporal evolution of the parameter:

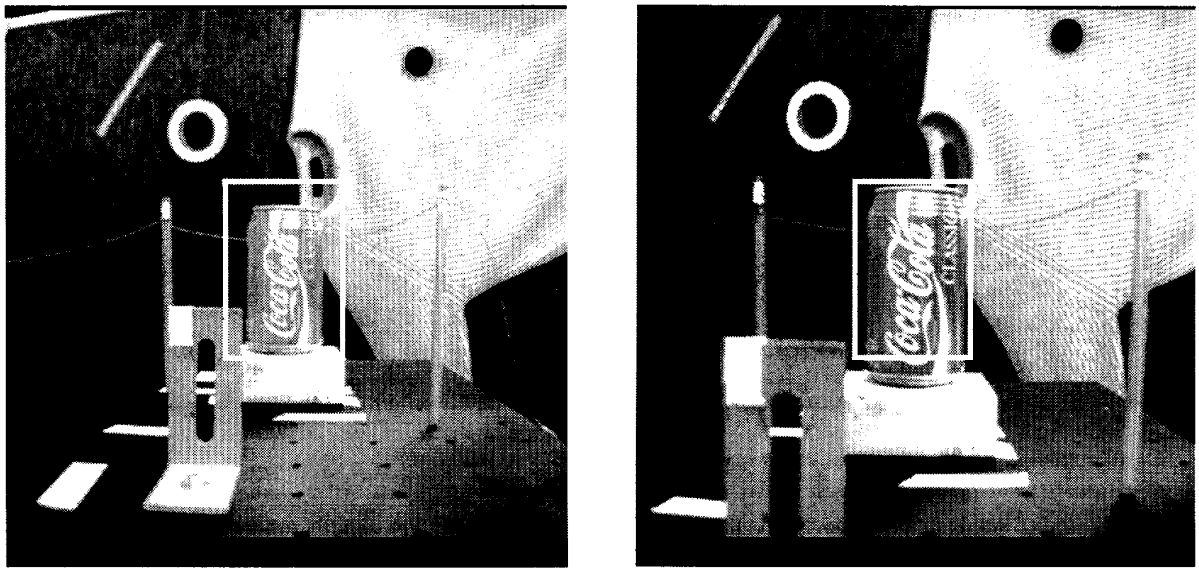
$$\begin{cases} \begin{bmatrix} a_i \\ \dot{a}_i \end{bmatrix}(t+1) = \begin{bmatrix} 1 & \delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_i \\ \dot{a}_i \end{bmatrix}(t) + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}(t) \\ \tilde{a}_i(t) = a_i(t) + \beta(t) \end{cases}$$

where $\tilde{a}_i(t)$ is the instantaneous measurement estimated by the multiresolution scheme, and $\beta(t)$, $\varepsilon_1(t)$ and $\varepsilon_2(t)$ are sequences of zero-mean Gaussian white noise. A standard Kalman filter [7] generates recursive estimates of each motion parameter and its temporal derivative. Fig. 9 illustrates the importance of the Kalman filter with a sequence of real data, the sequence *van* (see Fig. 8). We have plotted the estimate of time-to-collision (in seconds) directly computed from the coefficients estimated with the multiresolution scheme (i.e.,



(a) (b)

Fig. 5. Poster sequence. (a) Estimated time-to-collision (dashed line) and true time-to-collision (solid line) at the intersection of the optical axis and the image plane, and 2/div (dotted line). (b) Estimated time-to-collision (dashed line) and true time-to-collision (solid line) of the tracked plus sign.



(a) (b)

Fig. 6. NASA sequence with the selected window (of fixed size throughout the sequence), at time t_0 (a) and at time t_{150} (b).

without using the Kalman filter). We also plotted the estimates of time-to-collision computed from the smoothed coefficients after the filtering stage. The temporal filtering of the first order coefficients of the flow provides some immunity to the spurious variations of the motion field, due to shocks or vibrations of the camera.

IV. Experiments

The estimation of orientation and time-to-collision has been tested on a number of synthetic and real images. We present here four experiments, conducted with real images, that best illustrate the approach.

A. Quantitative Experiments

For each of the experiments that we present now, we have compared the estimate obtained by our method with the ground truth, determined by independent means, to judge the accuracy of the estimates.

1) Poster Sequence: The sequence has been acquired with an experimental cell consisting of a camera mounted on the end effector of a 6 DOF robot. The camera is undergoing a planar motion toward a poster pinned on a wall. Fig. 4 shows the first and the last frames of the sequence. The constant velocity of the camera, in the camera coordinate system (see Fig. 1), 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 ($17\,\mathrm{deg}$), i.e., the image plane is not parallel to the wall. We have performed two experiments with the sequence.

In the first experiment time-to-collision has been estimated at the intersection of the optical axis and the image plane. We note in Fig. 5(a) that the estimate of time-to-collision is in close agreement with the ground truth. We have also plotted the quantity $\frac{2}{div}$, where $div = u_x + v_y$ is the divergence of the motion field, to illustrate the improvement gained with our method when the surface of the object is not parallel to the image plane. The divergence has an obvious connection with time-to-collision, and it has been exploited in [14] to accomplish obstacle avoidance in a qualitative way. However it

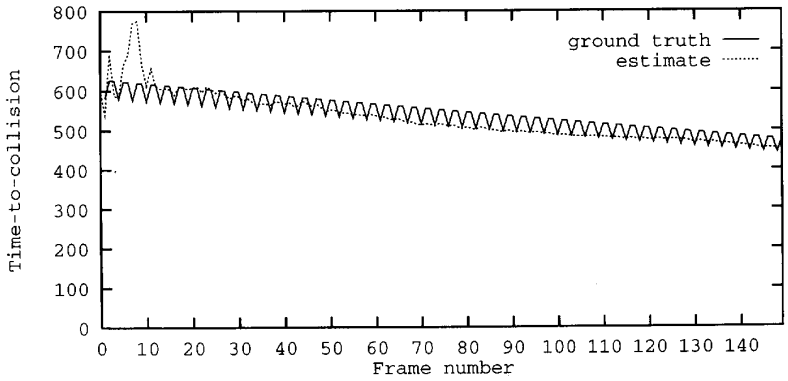


Fig. 7. NASA sequence. Time-to-collision: ground truth and estimate.

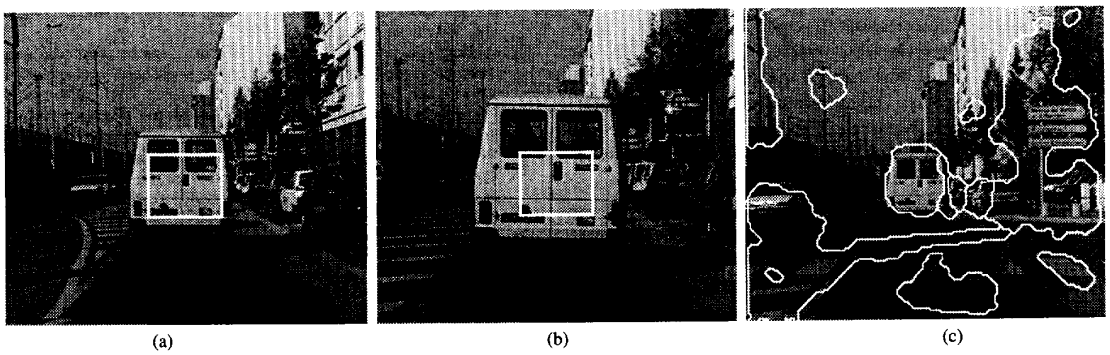


Fig. 8. From left to right: van sequence, at times t_5, t_{18}, and t_{60}. The selected window (of fixed size throughout the sequence) is superimposed at times t_5 (a) and t_{18} (b). The contours of the regions obtained by the segmentation are superimposed onto the original image at time t_{60} (c).

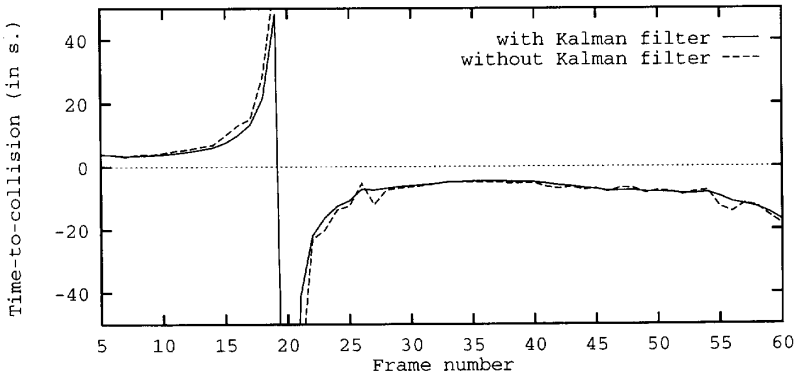


Fig. 9. Van sequence. Estimate of time-to-collision in seconds.

is clear from (6) and (4), that time-to-collision can be derived from the divergence alone only if the object is parallel to the image plane ($Z_X = 0$ and $Z_Y = 0$).

In the second experiment we have estimated time-to-collision of a fixed point on the poster. Since the camera is moving with respect to the poster, time-to-collision is estimated at different locations in the image plane. Rather than tracking a characteristic point on the poster (an edge corner for instance) we considered the 3-D coordinates of a point on the poster, and we projected the point in the image, using (1). We have superimposed in each image a plus sign to illustrate the trajectory of the point over time, as shown in Fig. 4. Fig. 5(b) shows the estimated and the true time-to-collision. Two or three frames are necessary for the Kalman filter to provide reliable estimates of $\dot{w}_0$. After a few frames the estimates of τ_c almost coincide with the true values, as shown in Fig. 5(b). These two experiments demonstrate with real images that the method can accurately estimate time-to-collision when the surface of the object is not parallel to the image plane.

2) NASA Sequence: The sequence was collected at NASA Ames Research Center and is provided courtesy of Dr. Banavar Sridhar.

Fig. 6 shows the first and the last images of the sequence, at time t_0 and t_{150}. The scene is static and the focus of expansion is centered on the can. The incremental displacement between two frames along the optical axis is constant. (In fact, every four frames the displacement is slightly smaller than expected. This explains the "saw-tooth" shape of the true time-to-collision in Fig. 7.) We have selected a window (of fixed size throughout the sequence) located in the center of the image. We estimated the time-to-collision of the camera with respect to the can appearing in the window. We note in Fig. 7 that after a few time steps (necessary for the Kalman filter to reach its steady state) the temporal smoothing performed by the filter allows us to estimate a smoothed version of the oscillating true time-to-collision.

B. A Qualitative Experiment: The Van Sequence

The sequence is provided courtesy of Thomson LER, Cesson-Sevigné, France. It was acquired with a camera attached to a moving car. Because we did not have access to the intrinsic parameters of the camera, we could only estimate τ_c at the center of the image (assuming that (x_c, y_c) coincides with the center of the image). Fig. 8 shows three frames of the sequence at time t_5, t_{18}, and t_{60}. The car drives behind a van. In the first part of the sequence (from t_1 to t_{18}) the van and the car slow down because of a red traffic light, and their relative velocity decreases. The background and the van undergo a similar motion relative to the camera, and the motion-based segmentation [3] is not able to separate the van from the background. Then the traffic light turns green. The van accelerates faster than the car and moves away from it. The van and the background moves in opposite direction with respect to the camera, and thus are much easier to separate, as shown in Fig. 8. We decided not to use the segmentation, but to select a window (of fixed size throughout the sequence) centered in the image. The van was always the dominant object appearing in the window. Fig. 9 shows the estimate of time-to-collision (in seconds). From t_1 to t_{18} the relative velocity of the car and the van decreases, thus time-to-collision increases. At t_{18} the relative velocity is equal to zero and the time-to-collision is infinite. When the van moves away from the car time-to-collision becomes negative. In this case τ_c can not be interpreted as time-to-collision, since there will not be any collision, nevertheless we obtain information about the rate at which the van is moving away from the car. Despite the crude segmentation given by the window (which almost never corresponds to the contour of the van) the approach enables us to recover a smooth and realistic time-to-collision. The results illustrate the ability of the approach to deliver reliable and relevant information for outdoor visual navigation.

V. Conclusion

The main contribution of this paper is to demonstrate, with sequences of real images, that time-to-collision, which is vital 3-D information for the outdoor navigation of unmanned vehicles and mobile robots, can be robustly and accurately recovered. A novel method for estimation of the first order coefficients of the motion field and their temporal derivatives has been developed. The method exploits a multiresolution scheme and temporal continuity of motion using a Kalman filter. This approach made it possible to apply the theoretical analysis of Subbarao and Waxman [15], [17] in realistic situations. The estimation of time-to-collision then becomes an attractive option, directly providing 3-D information for the navigation of unmanned vehicles, without requiring a complete reconstruction of the scene. A theoretical analysis of the accuracy and the sensitivity of the method, is left for a future work.

Acknowledgment

The author wishes to thank Patrick Bouthemy for his continued guidance and advice during the course of this work, Rachid Deriche for helpful comments on an earlier version of the paper, Rupert Curwen for checking the English, and the reviewers for their comments, which helped in improving this manuscript.

References

[1] B. Bascle, P. Bouthemy, R. Deriche, and F. Meyer, "Tracking complex primitives in an image sequence," in *Proc. 12th ICPR*, Jerusalem, to appear.

[2] A. Blake, R. Curwen, and A. Zisserman, "A framework for spatiotemporal control in the tracking of visual contours," *Int. J. Comput. Vis.*, vol. 11, no. 2, pp. 127–145, 1993.

[3] P. Bouthemy and E. François, "Motion segmentation and qualitative dynamic scene analysis from an image sequence," *Int. J. Comput. Vis.*, vol. 10, no. 2, pp. 157–182, 1993.

[4] P. J. Burt, "The pyramid as a structure for efficient computation," in *Multiresolution Image Processing and Analysis*, A. Rosenfeld, Ed. New York: Springer-Verlag, 1984, pp. 6–35.

[5] M. Campani and Alessandro Verri, "Motion analysis from first-order properties of optical flow," *CVGIP: Image Understanding*, vol. 56, no. 1, pp. 90–107, 1992.

[6] R. Cipolla and A. Blake, "Surface orientation and time to contact from image divergence and deformation," in *Proc. ECCV-92*, Italy, 1992, pp. 187–202.

[7] Arthur Gelb, *Applied Optimal Estimation*. Cambridge, MA: MIT Press, 1974.

[8] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp 185–203, 1981.

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

[10] H. C. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image," *Proc. R. Soc. Lond.*, vol. B-208, pp. 385–397, 1980.

[11] F. Meyer and P. Bouthemy, "Estimation of time-to-collision maps from first order motion models and normal flows," in *Proc. 11th ICPR*, The Hague, 1992, pp. 78–82.

[12] F. G. Meyer and P. Bouthemy, "Region based tracking using affine motion models in long image sequences," *CVGIP: Image Understanding*, to appear.

[13] S. Negahdaripour and S. Lee, "Motion recovery from image sequences using only first order optical flow information," *Int. J. Comput. Vis.*, vol. 9, no. 3, pp. 163–184, 1992.

[14] C. Nelson and J. Aloimonos, "Obstacle avoidance using flow field divergence," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 10, pp. 1102–1106, 1989.

[15] M. Subbarao, "Interpretation of image flow: A spatio-temporal approach," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 3, pp. 266–278, 1989.

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

[17] M. Subbarao and A. M. Waxman, "Closed form solutions to image flow equations for planar surfaces in motion," *CVGIP*, vol. 36, pp. 208–228, 1986.

[18] M. Tistarelli and G. Sandini, "On the advantages of polar and log-polar mappings for direct estimation of time-to-impact from optical flow," *IEEE Trans. Pattern Anal. Machine Intell.*, pp. 401–410, 1993.

[19] A. M. Waxman and S. Ullman, "Surface structure and three-dimensional motion from image flow kinematics," *Int. J. Robotics Res.*, vol. 4, no. 3, pp. 72–94, 1985.