FLORE

Repository istituzionale dell'Università degli Studi di Firenze

GENERALIZED BOUNDS FOR TIME TO COLLISION FROM FIRST-ORDER IMAGE MOTION

(Article begins on next page)

Generalized Bounds for Time to Collision from First-Order Image Motion

C. Colombo   A. Del Bimbo

Dipartimento di Sistemi e Informatica
Via S. Marta 3, I-50139 Firenze, ITALY
{columbus, delbimbo}@dsi.unifi.it

Abstract

This paper addresses the problem of estimating time to collision from local motion field measurements in the case of unconstrained relative rigid motion and surface orientation. It is first observed that, as long as time to collision is regarded as a scaled depth, the above problem does not admit a solution unless a narrow camera field of view is assumed. By a careful generalization of the time to collision concept, it is then expounded how to compute novel solutions which hold however wide the field of view. The formulation, which reduces to known literature approaches in the narrow field of view case, extends the applicability range of time to collision based techniques in areas such as mobile robotics and visual surveillance. The experimental validation of the main theoretical results includes a comparison of narrow- and wide-field of view time to collision approaches using both dense and sparse motion estimates.

1 Introduction

It is well known that the first-order local structure of motion fields (or motion parallax) embeds a great deal of geometric and kinematic information about a visual scene [9]. Evidence exists that biological visual systems exploit their specific sensitivities to motion parallax characteristic patterns such as dilatation and shear to support the execution of tasks such as visual exploration and heading direction control [15]. In computer vision, motion parallax extraction and analysis is often used in the place of more elaborate structure from motion techniques to achieve real-time performance in tasks such as frame-rate image segmentation [2], visual tracking and pose estimation [6], free space exploration [18], visual surveillance and obstacle avoidance [14], and vision-based robot control [1]. The *time to collision* (i.e., informally, the temporal distance between any scene point and the camera) is an important scalar visual field, theoretically obtainable from direct motion parallax measurements in the case of a spherical image surface [10]. In practice though, visual analysis is based on planar cameras, and specific criteria are to be devised so as to ensure that meaningful time to collision estimates are extracted from planar image motion observations. In this respect, time to collision approaches developed so far can be roughly divided in two classes, namely *exterospecific* and *propriospecific*. Exterospecific approaches are based on a partial a priori knowledge of either camera-scene relative geometry (e.g. frontoparallel surfaces [18]) or motion (e.g. dominant translation [3]); the main limitation of these approaches is that they only work for carefully controlled operating scenarios. Propriospecific approaches rely instead on limiting the visual analysis about the optical axis of perspective projection, thus reducing the camera field of view (FOV) to ensure that the image plane closely approximates locally the image sphere [17], [12], [4]. The main advantage of the simple narrow FOV constraint is that it can be applied whatever the external environment; however, if the constraint fails to be met, gross time to collision estimation errors have to be expected in the image periphery. This is most undesirable, since modern hardware technology allows the processing of large images in real-time, and a number of applications (e.g., surveillance) may indeed benefit from using a wide FOV.

This paper addresses a revisitation of the concept of time to collision, and describes a method for computing this important parameter from local first-order approximations of planar motion fields given an arbitrarily wide FOV and unknown relative motion and orientation. After some mathematical preliminaries, in Section 2 it is shown that, about the optical axis, time to collision can be effectively confused with scaled depth and estimated from local motion field observa-

tions around the image origin. Yet, at larger visual angles, a natural definition of time to collision should include both the translational and rotational components of rigid motion and, as a result, time to collision and scaled depth should be regarded as different visual entities. Section 3 provides two novel definitions of time to collision for an arbitrarily wide FOV, referring respectively to a planar and spherical sensor geometry, and converging to scaled depth in the particular case of narrow FOV. A closed-form solution using linear combinations of planar motion field invariants is obtained in Section 4 for the two times to collision by applying elementary differential geometry and projecting the planar motion field structure onto the unit sphere. In an experimental validation of the theoretical framework (Section 5), results of tests featuring both dense (optical flow) and sparse (active contours) affine motion estimates are presented and discussed. Finally, in Section 6 conclusions are drawn and future work is outlined.
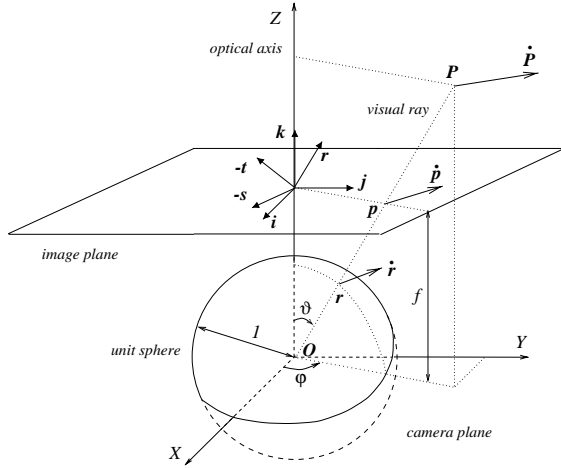

Figure 1: Geometry of image formation under planar and spherical projection.

2 Scaled Depth vs Time to Collision

Preliminaries. Let the imaged scene be composed of rigid surfaces in relative motion w.r.t. the camera. The geometry of image projection is illustrated in Fig. 1. The camera frame is $\{Oijk\}$, where O the center of projection and k the depth axis. The perspective projection $p = x\,i + y\,j + k$ of a visible surface point in space $P = R\,r$ is defined by $p = P/Z$ where, without loss of generality, focal length f is set to 1. The planar motion field $\dot{p} = u\,i + v\,j$ is expressable as a function of image position p and surface depth Z. The local motion field structure, or motion parallax,

is encoded in the Jacobian matrix $\partial \dot{p}/\partial p$ evaluated at p, i.e. in the four differential invariants of the motion field: divergence (div), curl (rot), and (two components of) deformation (defx, defy) [17]. It is not difficult to show [5] that the four invariants depend on image position p, depth Z, surface orientation (gradient ∇Z), and relative twist screw (rigid motion vectors V and Ω). Specifically, the deformation vector $\mathbf{def} =$ defx i + defy j can be expressed as the sum of two terms, taking into account respectively translations and rotations: $\mathbf{def} = \mathbf{def}_V(p; \nabla Z; Z; V) + \mathbf{def}_\Omega(p; \Omega)$. The term $\mathbf{def}_\Omega$ vanishes in the case of pure translation ($\Omega = 0$) or, whatever Ω, at the image origin ($p = 0$), while $\mathbf{def}_V$ vanishes either in the case of pure rotation ($V = 0$) or, whatever V, if the tangent plane at P is parallel to the image plane (frontoparallel condition, $\nabla Z = 0$).

Time to Collision as a Scaled Depth. In the recent computer vision literature, the terms "time to collision" and "scaled depth" are used interchangeably [17], [18], and referred to the scalar field

$$t_z = -\frac{Z}{V \cdot k} \tag{1}$$

giving at each image location p the ratio between surface depth at P and the camera-surface *translational velocity* component directed towards the image plane (a positive quantity for camera and surface approaching each other).

Due to the well known speed-scale ambiguity, the structure from motion problem can only be solved up to an unknown scale factor, so that scene structure is usually expressed in terms of scaled depth t_z. Hence, in principle, computing time to collision as a scaled depth implies solving in advance the structure from motion problem, and specifically separating the translational and rotational components of relative rigid motion [11], [8]. However, an approximation of scaled depth can be obtained from the first-order structure of the planar motion field without solving explicitly for rigid motion, provided that some constraints are set on relative motion or viewing angles.

Approximation	Constraint	Refs
Dominant Translation	$\|\Omega\| \approx 0$	[3], [7]
Frontoparallel Surface	$\|\nabla Z\| \approx 0$	[16], [18]
Narrow Field of View	$\|p\| \approx 0$	[14], [17]

Table 1: Constraints for scaled depth approximation.

Table 1 shows some popular constraints used for scaled depth approximation.

A general expression can be easily derived involving scaled depth at a generic image point [5]:

$$t_z^{-1} = \frac{\text{div} - \text{def}_V^\tau - 3\,\text{def}_\Omega^\varphi}{2} \quad , \qquad (2)$$

where τ is the (unknown) surface tilt angle, $\text{def} = \|\mathbf{def}\|$, and $\text{def}^\alpha = \cos 2\alpha\,\text{defx} + \sin 2\alpha\,\text{defy}$ is defined as *directional deformation* (in the image plane direction α), being $\text{defx} = \text{def}\cos 2\mu$ and $\text{defy} = \text{def}\sin 2\mu$. The approximating formulas for scaled depth can be obtained by using the constraint formulas of Table 1 in eq. (2). Of course, the stronger the operational constraints are, the easier is the scaled depth estimation process, at the expense of an higher probability of gross systematic errors when the constraints fail to be met perfectly. This is often the case when the dominant translation and the frontoparallel surface constraints are set. The narrow field of view constraint limits the range of visual directions to a small visual angle around the optical axis, for which it is assumed that no significant deformations exist in the first-order motion field structure. In such a case, scaled depth approximation has the form of a bound:

$$t_z^{-1} = \frac{\text{div} \pm \text{def}}{2} \quad . \qquad (3)$$

Several enhancements to the basic narrow FOV bound of eq. (3) have been proposed so far, by introducing, whenever possible, additional contraints such as fixation, partial knowledge of motion, etc. [4], [18].

Criticism. The strict requirement of the narrow FOV condition for approximating time to collision as a scaled depth is illustrated in Fig. 2, showing the reciprocal of time to collision (often referred to as *collision immediacy*) plotted as a function of the co-latitude angle ϑ spanning half of the overall visual field (FOV = 160 deg) in the image direction $\varphi = 0$ (see Fig. 1). The situation described in the figure is geometrically equivalent to the frontoparallel observation of a forehand stroke at tennis, where the player is simultaneously approaching the net and rotating his racket to hit the ball. The figure shows that, in a general case of surface rototranslation like this, the narrow FOV constraint approximation cannot be used to bound scaled depth at co-latitude angles wider than a few (say, fifteen) degrees, since the true value of scaled depth goes out of its bounds. This is because, while eq. (1) only refers to translation and disregards rotation, image divergence and deformation do depend on both translational and rotational velocities. Eq. (2) can be used to show why scaled depth cannot be determined from first-order motion field structure but

under special conditions. Indeed, it can be shown [5] that although the sum $\text{def}_V^\tau + \text{def}_\Omega^\varphi$ in eq. (2) can always be bounded, a bound for $\text{def}_V^\tau - \text{def}_\Omega^\varphi$ does never exist, so that the two directional deformations cannot be individually bounded: again, to have that, translation and rotation should be decoupled.
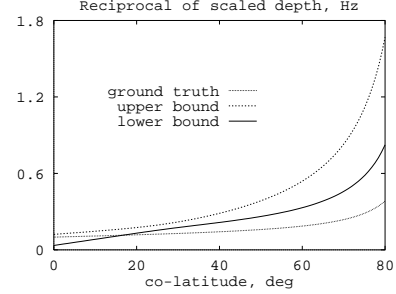

Figure 2: Erroneous bounds for scaled depth immediacy.

Another important observation should now be made about the general non appropriateness of interpreting scaled depth as a time to collision. Indeed, consider the case of a tennis player rotating his racket in place just in front of a camera (or, equivalently, consider someone trying to slap you!): as evident from Fig. 3, while such a pure rotation about an axis external and parallel to the camera plane corresponds no doubt to a dangerous situation for the observer, yet the scaled depth collision immediacy happens to be *identically zero everywhere*.

From the discussion above, we can state that: (i) scaled depth and time to collision should be regarded as distinct concepts; (ii) recovering scaled depth is a more difficult task than determining time to collision, since there is in general the need to separate translations from rotations. The rest of the paper is devoted to (iii) give alternative definitions of time to collision which can still hold when scaled depth fails; (iv) show how to compute the newly defined times to collision using the motion field and its first-order structure.

3 Time to Collision Revisited

There are, of course, diverse possible definitions of wide FOV time to collision extending eq. (1), each referring to a precise application context and geometry of the observer. In the following, two distinct definitions of time to collision are given, relying respectively on a spherical and a planar observer model.

The *spherical time to collision* is the time t_r it would take a point $\boldsymbol{P}$ to reach the camera center by traveling at a uniform velocity $\boldsymbol{P}\cdot(-\boldsymbol{r})\,\boldsymbol{r}$ along the line

Figure 4: Spherical *(a)* and planar *(b)* times to collision for the case of Fig. 3 (brighter means closer).
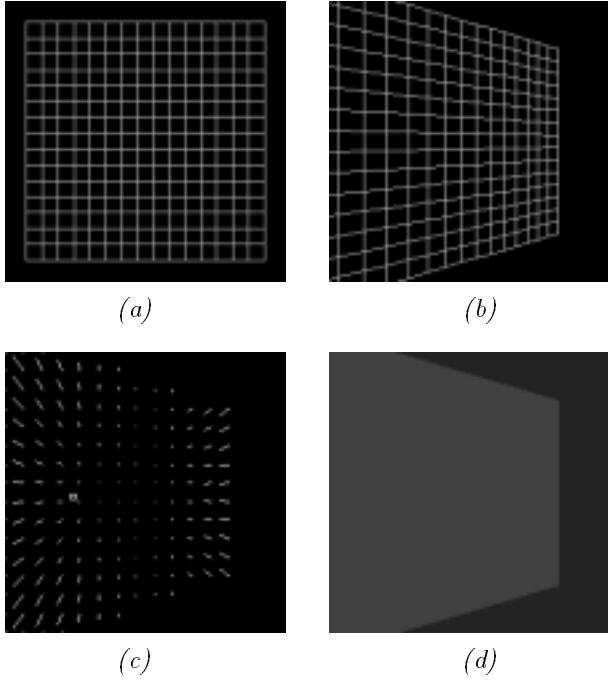

Figure 3: *(a),(b)*: two frames of a rotating plane; associated motion field *(c)* and scaled depth *(d)*. Scaled depth immediacy vanishes identically.

of sight, i.e.,

$$t_r = -\frac{R}{\boldsymbol{P} \cdot \boldsymbol{r}} \quad . \tag{4}$$

Eq. (4) provides a convenient way of defining time to collision having a sphere as the imaging surface. Such a definition is best suited to mobile robotics applications involving a robot with no dominant dimensions. The *planar time to collision* is defined as the time t_p it would take a point $\boldsymbol{P}$ to reach the camera plane by traveling at a uniform velocity $\boldsymbol{P} \cdot (-\boldsymbol{k})\,\boldsymbol{k}$ along the optical axis, i.e.,

$$t_p = -\frac{Z}{\boldsymbol{P} \cdot \boldsymbol{k}} \quad . \tag{5}$$

Eq. (5) is the natural extension of eq. (1) to the general case of planar observer and relative rototranslation, the optical axis of perspective providing the normal to the surface of collision (camera plane). In a driving application context, planar time to collision could be appropriate in the case of a vehicle with dominant transversal dimensions (think also of an aircraft and its wings).

Fig. 4 shows how, differently from scaled depth, both the spherical and planar times to collision defined above succeed to providing collision information in the "slap" sequence of Fig. 3. The novel definitions
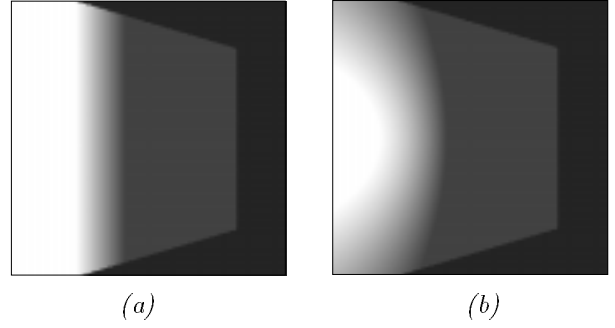
of eqs. (4) and (5) hold *whatever the relative motion and the FOV*, thus generalizing eq. (1) to the case of arbitrarily wide FOV. Indeed, the definitions do not refer to rigid motion vectors in a separate way, but regard the global effect of a given rigid motion at a given point. It is easy to prove [5] that both the equations yield scaled depth when the narrow FOV constraint is met; in particular, all given time to collision definitions are approximately equivalent to each other for small values of ϑ, while they become significantly different for FOVs of 30 degrees or larger. This confirms the fact that scaled depth is well approximated as a time to collision only at small co-latitude angles.

4   Generalizing the Bounds

In this Section, a closed form bound is derived for the two times to collision defined before, and an operational way to compute each bound from local image plane observations is provided. First of all, notice that, as a direct consequence of eqs. (4) and (5), the planar time to collision can be obtained at any co-latitude from the spherical time to collision and the *spherical motion field* $\dot{\boldsymbol{r}} = u'\,\boldsymbol{t} + v'\,\boldsymbol{s}$ (refer again to Fig. 1) as

$$t_p^{-1} = t_r^{-1} + \tan\vartheta\,u' \quad . \tag{6}$$

It is also easy to show that the spherical time to collision is bounded, at any co-latitude angle ϑ, by the divergence and deformation of the spherical motion field, i.e.,

$$t_r^{-1} = \frac{\mathrm{div}' \pm \mathrm{def}'}{2} \quad . \tag{7}$$

Eq. (7) is proved by regarding the plane tangent to the unit sphere at $\boldsymbol{r}$ as the image plane of a *virtual camera* with optical axis $\boldsymbol{r}$, and noting that an equation akin to eq. (3) holds at the origin of the virtual image plane—see also [17].

In order to compute both the spherical and planar times to collision by planar motion field estimates, there remains to show how to obtain spherical quantities from image plane observations. The following result can be proved [5], allowing to project the planar motion field and its first-order structure onto the unit sphere.

Lemma (Correspondence of Planar and Spherical Motion Fields) *The spherical divergence and deformation can be obtained by projection onto the unit sphere of the planar motion field linear structure, as*

$$\text{div}' = \text{div} - 3\tan\vartheta\, u'\ ;$$

$$
\text{def}' = \Big[\big(\text{def}^{\varphi} - \tan\vartheta\, u'\big)^2 \tag{8}
$$

$$
+ \ \big(m_\vartheta\, \text{def}^{\varphi+\pi/4} + n_\vartheta\, \text{rot} - \tan\vartheta\, v'\big)^2\Big]^{1/2}\ ,
$$

where $m_\vartheta = \frac{1}{2}\left(\sec\vartheta + \cos\vartheta\right)$, $n_\vartheta = \frac{1}{2}\left(\sec\vartheta - \cos\vartheta\right)$, *and the planar and spherical motion fields are related one-to-one by*

$$
\begin{bmatrix} u' \\ v' \end{bmatrix} = \begin{bmatrix} \cos^2\vartheta\cos\varphi & \cos^2\vartheta\sin\varphi \\ -\cos\vartheta\sin\varphi & \cos\vartheta\cos\varphi \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \ . \tag{9}
$$

Sketch of proof. First, the virtual camera concept introduced above is exploited to derive expressions for $\dot{r}$ and $\dot{p}$ as functions of r, V and Ω. Eq. (9) follows then easily by noting that $r = \cos\vartheta\, p$ and expressing $\dot{r}$ and $\dot{p}$ in the same coordinate system. To prove eq. (8), a basic result from differential geometry is used, stating that, if a smooth map exists between two manifolds, then the tangent spaces at corresponding points in the two manifolds are linearly related by the derivative of the same map [13]. In this case, the map is perspective, and the manifolds are the planar and spherical image surfaces. In particular, the first-order structures of the planar and spherical fields, encoded respectively in the Jacobian matrices of the planar and spherical motion fields, are related to each other linearly, via two matrices depending on ϑ, φ, u' and v'. Such matrices are finally used to get the divergence and deformation components in the spherical case, thus obtaining the desired result. □

Fig. 5 illustrates the results for the case of a planar visible surface rototranslating rigidly w.r.t. the camera in the same motion and surface conditions of Fig. 2. A glance to Fig. 5 shows that, differently from Fig. 2, *the true values of the spherical and planar times to collision always remain inside the bounds* obtained in this Section.
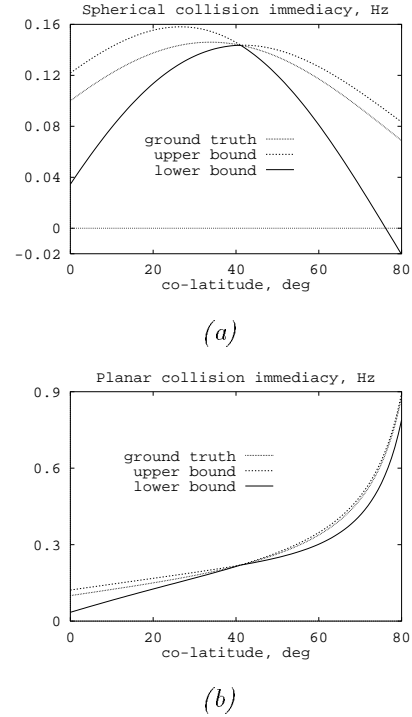

(a)

(b)

Figure 5: Spherical *(a)* and planar *(b)* collision immediacies, and their estimates.

5 Experimental Results

A number of experiments were conducted using the framework above to estimate time to collision from both videotape and live image sequences.

Time to collision from dense optic flow. Figs. 6*(a),(b)* show two subsequent frames of a videotape sequence featuring the rotation of a rigid flat panel in front of the camera. This situation is geometrically equivalent to having a camera mounted on a mobile robot rotating about a given point of the ground floor in proximity of a wall (also compare with Fig. 3). Fig. 6*(c)* shows the computed optic flow for a specific frame of the sequence. Optic flow computation was done by tracking image corners, and then interpolating linearly the obtained sparse image motion so as to get a smooth and dense motion field approximation (corner tracking also automatically provides an estimate of motion parallax). Due to the specific kind of 3D motion of the panel, the resulting image motion features a positive divergence in the left part of the image (where the panel is coming closer to the camera), and a negative divergence in the right part of the image—a zero divergence being obtained in correspondence of the vertical axis of rotation. Figs. 6*(d)–(f)*

show in order average scaled depth, planar and spherical collision immediacies: negative or zero immediacies are indicated in black. As evident from a qualitative comparison of the three results, the spherical and planar times to collision appear to be more appropriate than scaled depth to monitor rotations w.r.t. a planar object (it is evident, in Figs. 6(e),(f), the image area corresponding to approaching motion). In fact, unlike eqs. (6)–(7), the scaled depth immediacy bound of eq. (3) fails completely to take into account the effect of 3D rotations on motion vectors, and only uses planar motion parallax to get a time to collision estimate. A quantitative analysis of the results confirms this point (with an error of within 5% w.r.t. the ground truth for the planar and spherical times to collision).
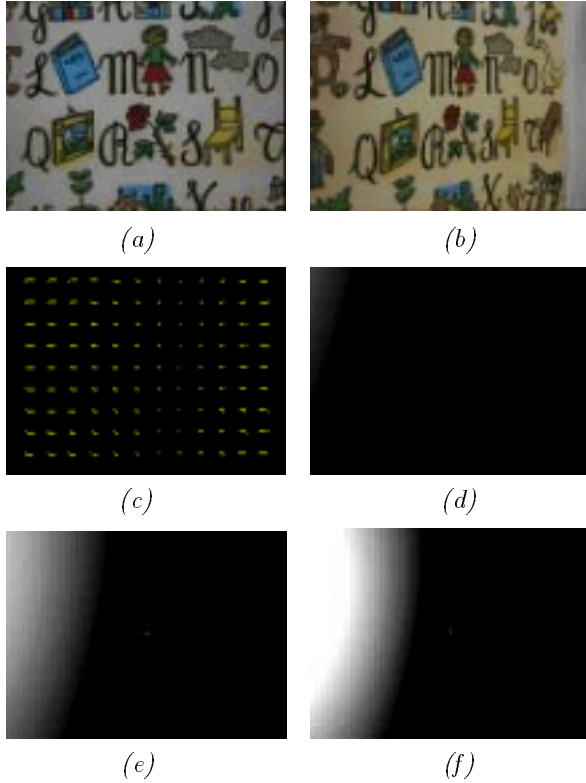
the average first-order parameters of image motion as in [4]. The tracker is initialized at startup using the computer mouse, and deforms at run-time in an affine way. The tracker includes a Kalman filter, ensuring a stable and robust behavior even in the presence of modeling uncertainties and distractors. The computed scaled depth and planar collision immediacies are reported in Figs. 7(b),(c) respectively. The figures show that, after a short transition time due to contour prediction inertia, while the average value of the planar time to collision keeps quite close to the ground truth value (specifically, around 4 s to collision), the time to collision bound based on scaled depth is practically useless, due to the fact that the slapping action involves more a hand rotation than a translation.
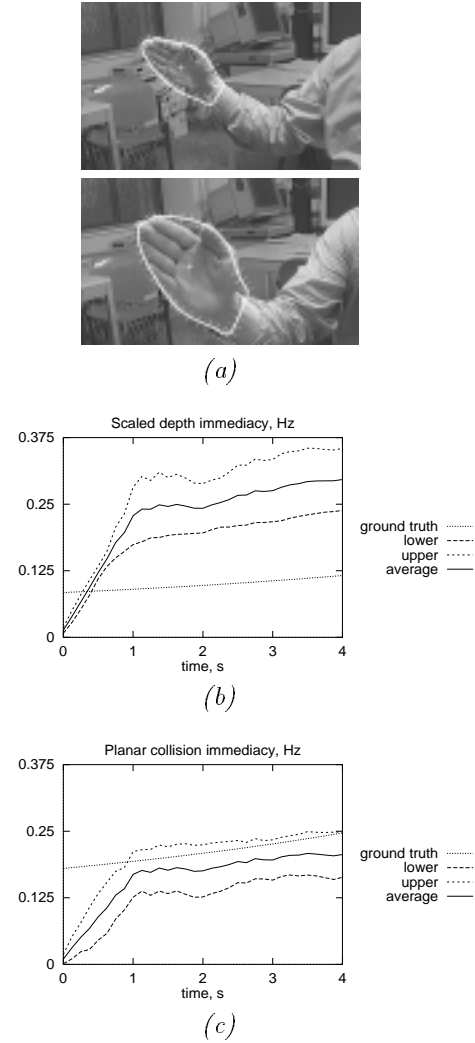

(a) *(b)*

(c) *(d)*

(e) *(f)*

Figure 6: A rotating plane. *(a)*, *(b)*: Two subsequent frames of the original sequence. *(c)*: Computed optic flow. *(d)–(f)*: Scaled depth, planar and spherical immediacies (brighter means closer).

Time to collision from active contour deformations. Fig. 7(a) shows two frames of a real-time sequence featuring a rototranslating hand as if "slapping" the camera. The hand is tracked using an active contour, whose deformations are used to compute

(a)

(b)

(c)

Figure 7: A slapping hand. *(a)*: Two frames of the original sequence, with superimposed active contour tracker. Scaled depth *(b)*, and planar *(c)* collision immediacies.

6 Conclusions and Future Work

The main contribution of the paper is to show that, although scaled depth and time to collision are usually considered as the same visual entity, they generally differ, especially in the far visual periphery. Better still, time to collision can be reliably estimated whatever large the field of view by suitable combinations of first-order motion field parameters, while scaled depth cannot.

The work can be expanded in several directions. A more general differential geometry formulation can be introduced to study the problem of visual parameters estimation in the case of general sensor shape and to prove the feasibility of the computational framework to applications involving non rigid motions. Currently, a new model of space-variant sensor is being experimented with, which was explicitly designed to carry out the computations required by the approach with a minimum of computational effort.

References

[1] B. Allotta and C. Colombo. On the use of linear camera-object interaction models in visual servoing. *IEEE Transactions on Robotics and Automation*, 15(2):350-357, 1999.

[2] P. Bouthemy and E. François. Motion segmentation and qualitative dynamic scene analysis from an image sequence. *International Journal of Computer Vision*, 10(2):157-182, 1993.

[3] M. Campani and A. Verri. Motion analysis from first-order properties of optical flow. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 56(1):90-107, 1992.

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

[5] C. Colombo. Time to collision from a natural perspective. Technical Report No. RT-199703-11, Dipartimento di Elettronica per l'Automazione, Università di Brescia, Italy, April 1997.

[6] C. Colombo and A. Del Bimbo. Real-time head tracking from the deformation of eye contours using a piecewise affine camera. *Pattern Recognition Letters*, 20(7):721-730, 1999.

[7] A. Giachetti and V. Torre. The use of optical flow for the analysis of non-rigid motions. *International Journal of Computer Vision*, 18(3):255-279, 1996.

[8] D.J. Heeger and A.D. Jepson. Subspace methods for recovering rigid motion I: Algorithm and implementation. *International Journal of Computer Vision*, 7(2):95-117, 1992.

[9] 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.

[10] J.J. Koenderink. Optic flow. *Vision Research*, 26(1):161-180, 1986.

[11] H.C. Longuet-Higgins and K. Prazdny. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B*, 208:385-397, 1980.

[12] F. Meyer and P. Bouthemy. Estimation of time-to-collision maps from first order motion models and normal flows", In *Proceedings of the 1992 IEEE International Conference on Pattern Recognition ICPR'92*, pages (I):78-82, 1992.

[13] J.W. Milnor. *Topology from the Differentiable Viewpoint*. The University Press of Virginia, Charlottesville, VI, 1965.

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

[15] D. Regan. Visual processing of four kinds of relative motion. *Vision Research*, 26:127-145, 1986.

[16] R. Sharma. Active vision in robot navigation: Monitoring time-to-collision while tracking. In *Proceedings of the 1992 IEEE/RSJ International Conference on Intelligent Robots and Systems IROS'92*, pages 2203-2208, 1992.

[17] M. Subbarao. Bounds on time-to-collision and rotational component from first-order derivatives of image flow. *Computer Vision, Graphics, and Image Processing*, 50:329-341, 1990.

[18] 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 Transactions on Pattern Analysis and Machine Intelligence*, 15(4):401-410, 1993.