ELSEVIER

Robotics and Autonomous Systems 31 (2000) 5–15

Robotics and Autonomous Systems

www.elsevier.com/locate/robot

Time to collision from first-order spherical image motion

Carlo Colombo*

Dipartimento di Sistemi e Informatica, Università di Firenze, Via S. Marta 3, I-50139 Firenze, Italy

Abstract

In the absence of contraints on either object motion or surface slant, a narrow field of view constraint has to be assumed to compute time to collision as a scaled depth from first-order image motion measurements. In this work, time to collision and scaled depth are regarded as different visual entities, and it is shown that a bound for time to collision can always be computed regardless of the field of view, thus extending the range of applicability of time to collision based techniques in areas such as mobile robotics and visual surveillance. The method relies on computing in closed form the spherical motion field and the associated parallax from image plane measurements obtained with either conventional cameras or space-variant sensors. An experimental validation of the main theoretical results highlights the difference between time to collision and scaled depth, and addresses a comparison of time to collision approaches using both dense and sparse motion estimates.

Keywords: Computer vision; Scaled depth and time to collision; Visual sensors; First-order image motion; Optical flow; Active contours

1. Introduction

In the last few decades, the analysis of monocular image sequences has proved a powerful tool for the extraction of geometric and kinematic information about a viewed scene (for a review, see, e.g., [20]). It is well known that due to the so-called speed-scale ambiguity, the "structure from motion" problem can only be solved up to an unknown scale factor [12], so that scene structure is usually expressed as a *scaled depth*. Most of the visual motion information is embedded in the first-order local structure of motion fields or motion parallax [11]. Evidence has been presented that the human visual system has specific sensitivities to motion parallax characteristic patterns such as dilation and shear [14], thus supporting the idea that first-order motion fields play a crucial role in biological vision for the execution of tasks such as visual exploration and heading direction control [8]. The motion parallax has been widely exploited in computer vision to carry out a number of tasks (e.g., image segmentation [1], robot navigation and obstacle avoidance [13], and robot visuo-motor control [3]) which do not require a full solution of the difficult structure from motion problem. Specifically, it has been shown [17] that the motion parallax embeds sufficient information for estimating the temporal distance between the observer and the observed object also known as *time to collision*.

Several theories on visual motion analysis have been developed for an ideal spherical sensor with a 360° field of view (FOV), representing the whole optic array of visual directions [11,12]. A wide FOV is no doubt very useful to both the biological and the artificial observer as a means of controlling the largest possible part of the visual environment. The recent advances

* Tel.: +39-055-479-6540; fax: +39-055-479-6363.
E-mail address: columbus@dsi.unifi.it (C. Colombo)

0921-8890/00/$ – see front matter
PII: S0921-8890(99)00077-9

in hardware technology have made it possible to use a wider FOV than in the past, typically combined with the reduced image resolution of space-variant visual sensors [18]. Yet, in practical applications, the assumption of a spherical eye is — often only implicitly — made by limiting the visual analysis about the optical axis of perspective projection, thus assuming a narrow FOV to ensure that, locally, the image plane closely approximates the image sphere [13,17]. Indeed, the wider is the FOV, the larger is, as compared to the corresponding spherical image, the distortion in a planar image due to perspective effects; this induces gross visual parameter estimation errors especially in the image periphery. Leonardo da Vinci was among the first to point out the visual distortions produced by "artificial" perspective in frescos of large dimensions as compared to the projection under "natural perspective" onto a human retina (a hemisphere) [10].

This paper describes a method for computing time to collision from local first-order approximations of planar motion fields with an arbitrarily wide FOV in the presence of general rigid motion and surface orientation (full mathematical details and proofs of all results can be found in [4]). After stating the computational framework (see Section 2), in Section 3 it is shown that, about the optical axis, time to collision can be effectively confused with scaled depth and estimated from local motion field observations 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 4 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 parallax coefficients is obtained 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.

2. Computational framework

2.1. Preliminaries and notation

Let the imaged scene be composed of rigid surfaces in relative motion with respect to the camera. The geometry of image projection is shown in Fig. 1. (For the sake of simplicity, the figure illustrates the special case in which the plane common to all velocity vectors coincides with the plane defined by the visual ray and the optical axis.) The camera frame is $\{\boldsymbol{Oijk}\}$, where $\boldsymbol{O}$ is the center of projection and $\boldsymbol{k}$ is the depth axis. The perspective projection $\boldsymbol{p} = x\boldsymbol{i} + y\boldsymbol{j} + \boldsymbol{k}$ of a visible surface point in space $\boldsymbol{P} = X\boldsymbol{i} + Y\boldsymbol{j} + Z\boldsymbol{k}$ is defined by $\boldsymbol{p} = \boldsymbol{P}/Z$ where, without loss of generality, focal length is set to 1.

The relative velocity at $\boldsymbol{P}$ can be expressed as $\dot{\boldsymbol{P}} = \boldsymbol{T} + \boldsymbol{\Omega} \wedge (\boldsymbol{P} - \boldsymbol{Q})$, where $\boldsymbol{T}$ is the translation, $\boldsymbol{\Omega}$ is the rotation, $\wedge$ denotes the usual vector product and the axis of rotation passes through an assigned world point $\boldsymbol{Q}$. The surface-camera relative twist screw is defined as $(\boldsymbol{V}, \boldsymbol{\Omega})$, where $\boldsymbol{V} = \boldsymbol{T} - \boldsymbol{\Omega} \wedge \boldsymbol{Q}$ and rotation is conventionally referred to an axis passing through $\boldsymbol{O}$. As $\boldsymbol{P}$ moves in space, its projection $\boldsymbol{p}$ also moves in the image plane, giving rise to a *planar motion field*

$$\dot{\boldsymbol{p}} = \frac{\dot{\boldsymbol{P}} - \dot{Z}\boldsymbol{p}}{Z}, \tag{1}$$

whose component along $\boldsymbol{k}$ is identically zero: $\dot{\boldsymbol{p}} = u\boldsymbol{i} + v\boldsymbol{j}$. The unit vector $\boldsymbol{r} = \boldsymbol{P}/R$ can be interpreted both as the direction of the visual ray through $\boldsymbol{P}$ and as the central projection of $\boldsymbol{P}$ onto the unit sphere (natural perspective). A new reference frame $\{\boldsymbol{Otsr}\}$ obtained by rotation of the camera frame is associated to each point of the sphere. As $\boldsymbol{P}$ changes, a *spherical motion field* $\dot{\boldsymbol{r}} = u'\boldsymbol{t} + v'\boldsymbol{s}$ is obtained in the plane tangent to the unit sphere at $\boldsymbol{r}$. The planar and spherical motion fields are clearly identical at the planar image origin $\boldsymbol{p} = \boldsymbol{k}$, where the image plane is tangent to the unit sphere and the optical and visual rays are parallel: $\boldsymbol{r} = \boldsymbol{k}$. The fields become more and more dissimilar as the angle ϑ between the two rays increases moving from the image origin towards the periphery. The field of view FOV $\in [0, \pi]$ is defined as twice the maximum attainable ϑ as related to the actual focal length and physical dimensions of the imaging sensor. For any given $\boldsymbol{r}$, coordinate representations in the camera

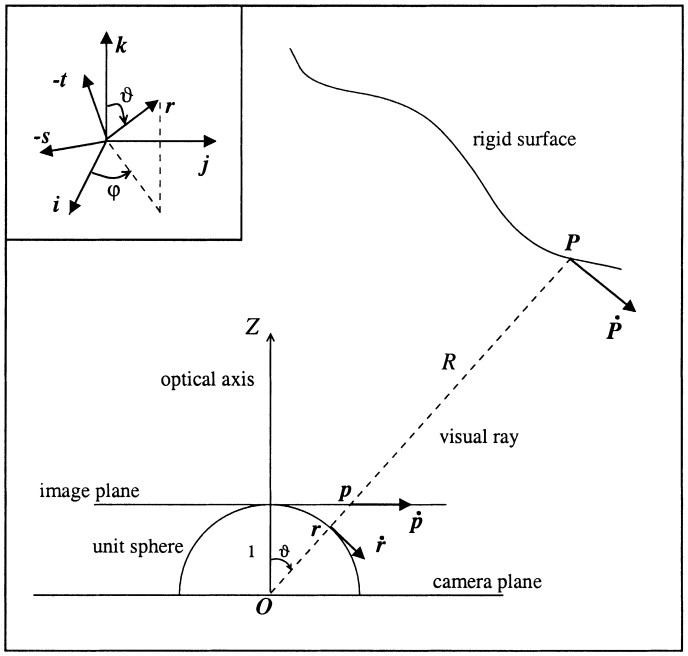


Fig. 1. Geometry of image projection onto camera plane (artificial perspective) and spherical optical array (natural perspective).

and spherical frames are related to each other by a rotation matrix whose entries depend on co-latitude ϑ and longitude φ. Specifically, pairs of image coordinates and visual directions correspond one-to-one, according to the transformation

$$[x \quad y]^{\mathrm{T}} = \tan\vartheta[\cos\varphi \quad \sin\varphi]^{\mathrm{T}}.$$

2.2. First-order planar motion field

The computational framework requires that the planar motion field and its first spatial derivatives be available at a given number of image points. Such local measurements can either be dense (e.g., the result of an optic flow image sequence analysis [2]) or sparse (as resulting from the tracking of visual features from frame to frame [3]) in the image. The planar motion field $\dot{\boldsymbol{p}}$ is expressible as a function of image position $[x \quad y]^{\mathrm{T}}$ and depth Z. In particular, in a suitable neighborhood of $\boldsymbol{p}$, the motion field can be approximated as

$$[u(x_n, y_n) \quad v(x_n, y_n)]^{\mathrm{T}} = [u(x, y)v(x, y)]^{\mathrm{T}} + M[x_n - x \quad y_n - y]^{\mathrm{T}},$$

where the Jacobian matrix

$$M = \frac{1}{2}\left\{\begin{bmatrix} \mathrm{div} & 0 \\ 0 & \mathrm{div} \end{bmatrix} + \begin{bmatrix} 0 & -\mathrm{rot} \\ \mathrm{rot} & 0 \end{bmatrix} + \begin{bmatrix} \mathrm{defx} & \mathrm{defy} \\ \mathrm{defy} & -\mathrm{defx} \end{bmatrix}\right\} \tag{2}$$

evaluated at $[x \quad y]^{\mathrm{T}}$ encodes the first-order local spatial structure of the planar field (motion parallax) in the four numbers div, rot, defx and defy, referred to respectively as divergence, curl, and (two components of) deformation. Regarded as image transformations, divergence accounts for isotropic expansions, curl for rigid rotations, and deformation for a simultaneous expansion and compression along two orthogonal image directions [11]. Concerning the projective properties, the planar motion parallax depends on image position, visible surface depth and orientation, and relative twist screw [4]. In particular, the deformation vector $\mathbf{def} = \mathrm{defx}\boldsymbol{i} + \mathrm{defy}\boldsymbol{j}$ can be expressed as the sum of two terms, taking into account respectively 3D translations and rotations: $\mathbf{def} = \mathbf{def}_V(\boldsymbol{p}; \nabla Z; Z; \boldsymbol{V}) + \mathbf{def}_\Omega(\boldsymbol{p}; \boldsymbol{\Omega})$. The term $\mathbf{def}_\Omega$ vanishes in the case of

Table 1
Constraints for scaled depth approximation

Approximation	Constraint	Validity	Formula	Refs.
Narrow field of view	$\|\boldsymbol{p}\| \approx 1$	Near image origin	$t_z^{-1} = \frac{1}{2}(\mathrm{div} \pm \mathrm{def})$	[13,17]
Dominant translation	$\|\boldsymbol{\Omega}\| \approx 0$	Any image point	$t_z^{-1} = \frac{1}{2}(\mathrm{div} \pm \mathrm{def})$	[2,7]
Frontoparallel surface	$\|\nabla Z\| \approx 0$	Any image point	$t_z^{-1} = \frac{1}{2}(\mathrm{div} - 3\mathrm{def}_\varphi)$	[16,19]

pure translation ($\boldsymbol{\Omega} = \mathbf{0}$) or, whatever $\boldsymbol{\Omega}$, at the image origin ($\boldsymbol{p} = \boldsymbol{k}$), while $\mathbf{def}_V$ vanishes either in the case of pure rotation ($\boldsymbol{V} = \mathbf{0}$) or, whatever $\boldsymbol{V}$, if the surface gradient ∇Z is zero, i.e., if the tangent plane at $\boldsymbol{P}$ is parallel to the image plane (frontoparallel surface condition).

3. Narrow field of view time to collision

3.1. Time to collision as a scaled depth

In the recent computer vision literature, the terms "time to collision" and "scaled depth" are used interchangeably [17,19], and referred to the scalar field

$$t_z = -\frac{Z}{\boldsymbol{V} \cdot \boldsymbol{k}}, \tag{3}$$

giving at each image location $\boldsymbol{p}$ the ratio between surface depth at $\boldsymbol{P}$ and the camera-surface *translational velocity* component directed towards the image plane (a positive quantity for camera and surface getting closer one to the other).

In principle, computing time to collision as a scaled depth implies solving in advance the structure from motion problem, and specifically separating (decoupling) the translational and rotational components of relative rigid motion [9,12]. However, an approximation of scaled depth (and time to collision) can be obtained without explicitly solving for rigid motion, provided that some constraints are set on relative motion and/or viewing angles [4]. Widely used approximated formulas for scaled depth are reported in Table 1, where $\mathrm{def} \doteq \|\mathbf{def}\|$ and $\mathrm{def}_\varphi \doteq \mathrm{defx}\cos 2\varphi + \mathrm{defy}\sin 2\varphi$.

As it clearly emerges from the table, the stronger the operational constraints are, the easier is the scaled depth estimation process, of course at the expense of a 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. (In the specific case of frontoparallel condition, notice how, in principle, scaled depth can be not only bounded, but even exactly computed at any image point.) Hence, when nothing is known about camera-scene relative geometry and motion (*exterospecific* conditions), only the narrow FOV constraint — which is a *propriospecific* condition, depending only on the observer — can be used. 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. The $\pm$ sign in the associated formula, namely,

$$t_z^{-1} = \frac{\mathrm{div} \pm \mathrm{def}}{2}, \tag{4}$$

is due to the fact that the orientation (tilt) of the plane tangent to the visible surface at $\boldsymbol{P}$ is generally unknown. Several variations to the basic narrow FOV bound have been proposed so far, attempting to overcome the tilt ambiguity by coupling the narrow FOV constraint with additional contraints such as fixation, motion on a ground plane, partially known motion, etc. [3,19].

3.2. A case study

In Fig. 2, the reciprocal of time to collision (often referred to as *collision immediacy*) is plotted as a function of the co-latitude angle ϑ spanning half of the overall visual field (FOV $= 160^\circ$) in the image direction $\varphi = 0$. The case depicted in the figure is quite general, as it involves both 3D translation and rotation. Specifically, the axis of rotation passes through point $\boldsymbol{Q} = c\boldsymbol{k}$ in space, with $c = 100$ flu (focal length units). Hence, relative translation $\boldsymbol{V}$ is partially due to rotation, being $\boldsymbol{T} = -10\boldsymbol{k}$ flu $\cdot$ frames^{-1} the translation of the axis of rotation, and $\boldsymbol{\Omega} = 5\boldsymbol{j}^\circ \cdot \mathrm{frames}^{-1}$. The planar surface has a constant depth gradient equal to $\nabla Z = -\frac{1}{2}(\boldsymbol{i} + \sqrt{3}\boldsymbol{j})$. The figure shows that, in

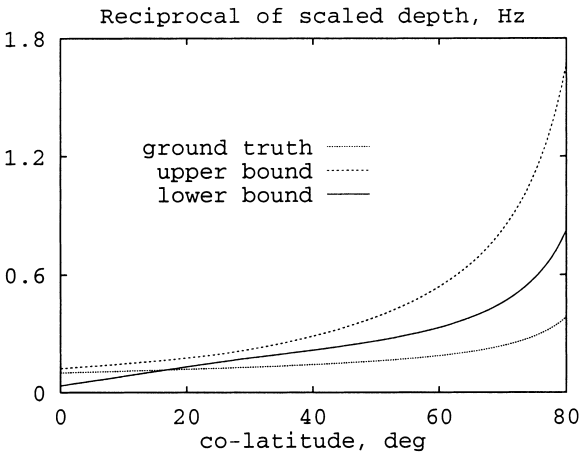


Fig. 2. Erroneous bound estimate of scaled depth immediacy.

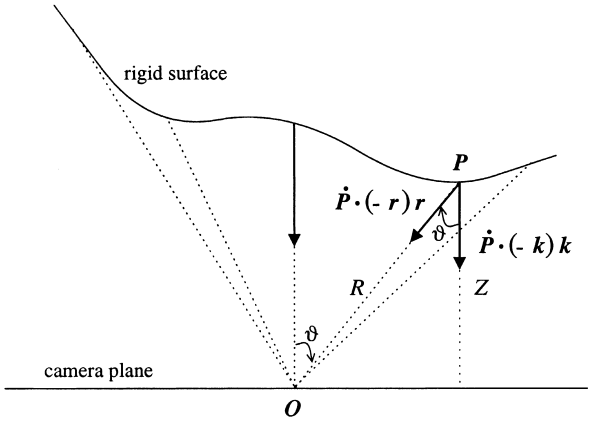


Fig. 3. Geometric meaning of planar and spherical time to collision.

the general case of surface rototranslation, the narrow FOV constraint approximation cannot be used to bound scaled depth at co-latitude angles wider than a few (say, 15) degrees, since the true value of scaled depth goes out of its bounds. This is because, while Eq. (3) only refers to translation and disregards rotation, image divergence and deformation do depend on both translational and rotational velocities. As a result, *recovering scaled depth is a more difficult task than determining time to collision*. Indeed, unlike scaled depth, wide FOV time to collision can always be bounded by first-order structure coefficients of either the planar or the spherical motion field (see Section 4), since it encodes translation and rotation in an undecoupled way.

4. Wide field of view time to collision

4.1. Time to collision revisited

There are, of course, diverse possible definitions of wide FOV time to collision extending Eq. (3), each referring to a precise operating context, application and geometry of the observer. A reference framework of application is mobile robotics, where time to collision can be used to carry out free space exploration, obstacle detection and surveillance. In the following, two distinct definitions of time to collision are given, relying, respectively, on a spherical and a planar observer model (see also Fig. 3).

Spherical time to collision. The spherical time to collision is defined as the time t_r it would take a point $\boldsymbol{P}$ to reach the camera center by traveling at a uniform *total velocity* $\dot{\boldsymbol{P}} \cdot (-\boldsymbol{r})\boldsymbol{r}$ along the line of sight, i.e.,

$$t_r = -\frac{R}{\dot{\boldsymbol{P}} \cdot \boldsymbol{r}}. \tag{5}$$

Planar time to collision. 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 *total velocity* $\dot{\boldsymbol{P}} \cdot (-\boldsymbol{k})\boldsymbol{k}$ along the optical axis, i.e.,

$$t_p = -\frac{Z}{\dot{\boldsymbol{P}} \cdot \boldsymbol{k}}. \tag{6}$$

Eq. (5) provides a convenient way of defining time to collision having a sphere as the imaging surface. Such a definition has a strict relationship with the "time to flight" of sonar-based robotic systems, being best suited to robotic applications involving a robot with no dominant dimensions (e.g., typical mobile robots). Eq. (6) is the natural extension of Eq. (3) 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). Since the actual imaging sensors' dimensions are quite small, collision with the (infinite) camera plane has more the nature of a virtual event than that of a physical event. Yet, in a robotic application planar time to collision could be appropriate in the case of a robot with dominant transverse dimensions (think also of an aircraft and its wings).

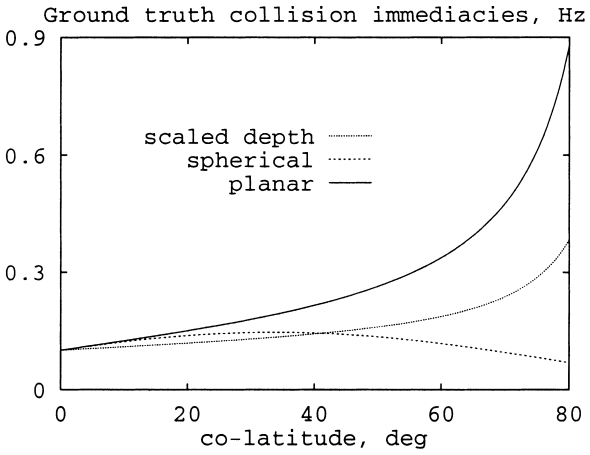


Fig. 4. Scaled depth and wide FOV collision immediacies.

The novel definitions of Eqs. (5) and (6) hold *whatever be the relative motion and the FOV*, thus generalizing Eq. (3) to the case of arbitrarily wide FOV. It is easy to prove that both the equations yield scaled depth when the narrow FOV constraint is met: this fact is clearly visible in Fig. 4, reporting collision immediacies as defined in Eqs. (3), (5) and (6), respectively, for the same motion and surface parameters of Fig. 2. It appears that these three quantities are approximately equal only for small values of ϑ, becoming significantly different for FOV of 30° or larger. This confirms the fact that scaled depth is well approximated as a time to collision only at small co-latitude angles.

4.2. Computing wide FOV time to collision

A closed form bound is derived for the two times to collision, 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. (5) and (6) and the statements of Section 2, the planar time to collision can be obtained at any co-latitude from the spherical time to collision and the spherical motion field component along $\boldsymbol{t}$, as

$$t_{\mathrm{p}}^{-1} = t_{\mathrm{r}}^{-1} + \tan\vartheta u'. \tag{7}$$

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.,

$$t_{\mathrm{r}}^{-1} = \frac{\mathrm{div}' \pm \mathrm{def}'}{2}. \tag{8}$$

Eq. (8) 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. (4) holds at the origin of the virtual image plane (see also [17]).

Hence, to compute both the spherical and planar times to collision by planar motion field estimates, there remains to show how to compute spherical quantities from image plane observations. The following result, proved in [4], is used, allowing to project the planar motion field and its first-order structure onto the unit sphere.

Proposition (Correspondence of planar and spherical motion fields).

1. *The planar and spherical motion fields are related one-to-one by*

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = H \begin{bmatrix} u \\ v \end{bmatrix} \tag{9}$$

where

$$H = \begin{bmatrix} \cos^2\vartheta & 0 \\ 0 & \cos\vartheta \end{bmatrix} \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}.$$

2. *The planar and spherical motion parallaxes, encoded respectively in the Jacobian matrices M and M', are related one-to-one by*

$$M' = HMH^{-1} + K' \tag{10}$$

where

$$K' = -\tan\vartheta \begin{bmatrix} 2u' & 0 \\ v' & u' \end{bmatrix}.$$

To summarize, once the planar field and parallax have been estimated, Eqs. (9) and (10) can be used to compute in closed form their spherical counterparts; then, an equation analogous to Eq. (2) allows extracting the spherical divergence and deformation from M', and eventually computing the planar and spherical times to collision according to Eqs. (7) and (8).

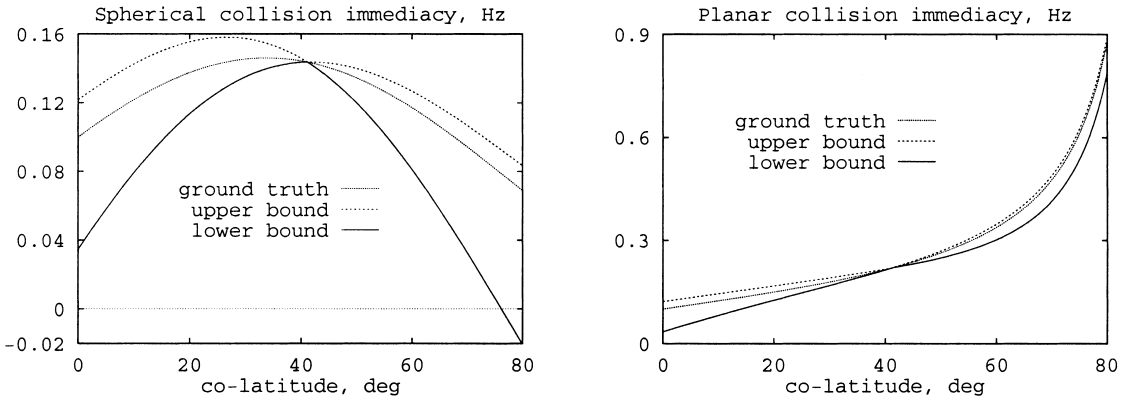


Fig. 5. Spherical (*left*) and planar (*right*) collision immediacies, and their estimates.

Fig. 5 illustrates the results for the case of a planar visible surface rototranslating rigidly with respect to 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.

4.3. Computational analysis

A further insight into the framework is gained in the following discussion of the computational advantages of computing wide FOV time to collision in closed form, and of using log-polar sensors in the place of conventional cameras.

4.3.1. The importance of a closed form

The closed form formulation of Eqs. (9) and (10) has a clear computational advantage over the corresponding algorithmic formulation, i.e., first remapping the planar field onto the unit sphere using Eq. (9), and then obtaining the spherical parallax by direct differentiation of the spherical field. Indeed, apart from being impractical, using the algorithmic formulation could be even impossible, as it happens when only a sparse planar field is available.

A delicate point about spherical field differentiation is related to the geometric distortion of the neighborhood of support for the computation of derivatives. In fact, not only the motion field in the neighborhood of an image point is deformed when passing from the plane to the sphere, but the neighborhood itself undergoes the deformation governed by Eq. (9), since it can obviously be regarded as the particular motion field with the identity matrix as Jacobian (see Fig. 6). In conclusion, both the motion field and its support neighborhood must be projected onto the sphere, where for field differentiation the nonregular sampling of the support must be taken into account.

4.3.2. Advantages of using log-polar sensors

Log-polar imaging sensors are a special kind of sensors provided with a radial disposition of photosensitive elements instead of the *xy* organization of conventional raster cameras. In the log-polar case the ratio between consecutive radii is equal to a constant $a > 1$ so that $\xi\gamma$ log-polar pixels s.t. $\xi = \log_a \sqrt{x^2 + y^2} - h$ and $\gamma = k \arctan(y/x)$, h and k real positive, grow exponentially in size as the radial coordinate increases from the uniform resolution fovea towards sensor periphery [15]. As a result, while in traditional cameras the tangent of the FOV angle grows only linearly with sensor dimensions, in a log-polar sensor such growth is exponential, and a trade-off exists between image resolution and FOV width for a fixed number of sensor pixels (see Fig. 7). It is thus clear that using a log-polar sensor to estimate time to collision proves particularly convenient if a wide FOV has to be monitored, and a dense time to collision map has to be computed [6,19].

Some authors have argued recently that radial sensors have further specific advantages over traditional

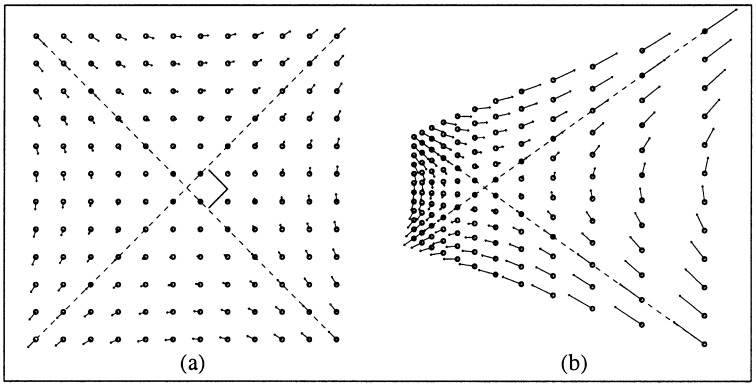


Fig. 6. Two image projections of a spherical motion field of pure deformation: (a) $\vartheta = 0$ (undistorted field); (b) $\vartheta = 45^\circ$.

cameras related to the computation of time to collision. For example, in [19], a formula is proposed to compute exactly time to collision for an arbitrarily large FOV. Yet, as shown in Table 1, such a result appears not to be due to intrinsic advantages of radial sensors but to the strong assumption that the surface be frontoparallel at each imaged point. Anyway, as proved in [4], the log-polar organization of sensor shape does present a computational advantage over traditional sensors in the wide FOV computation scheme expounded in this paper. The advantage is related to the fact that the spherical field can be obtained from the log-polar representation of the corresponding planar field as (compare to Eq. (10))

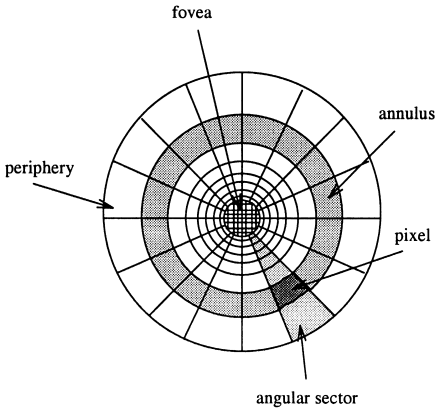


Fig. 7. Log-polar sensor geometry.

$$M' = \begin{bmatrix} \cos\vartheta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\partial\dot{\xi}}{\partial\xi} + \ln a\,\dot{\xi} & k\ln a\,\frac{\partial\dot{\xi}}{\partial\gamma} - \frac{\dot{\gamma}}{k} \\ \frac{1}{k\ln a}\,\frac{\partial\dot{\gamma}}{\partial\xi} + \frac{\dot{\gamma}}{k} & \frac{\partial\dot{\gamma}}{\partial\gamma} + \ln a\,\dot{\xi} \end{bmatrix} \times \begin{bmatrix} \sec\vartheta & 0 \\ 0 & 1 \end{bmatrix} + K', \tag{11}$$

hence without the need (as with raster cameras) to rotate, at each image pixel, the planar coordinate frame by an angle φ.

5. Experimental results

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

5.1. Time to collision from dense optic flow

Figs. 8(a) and (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 kinematically equivalent to having a camera mounted on a mobile robot rotating about a given point of the ground floor in proximity of a wall. Fig. 8(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

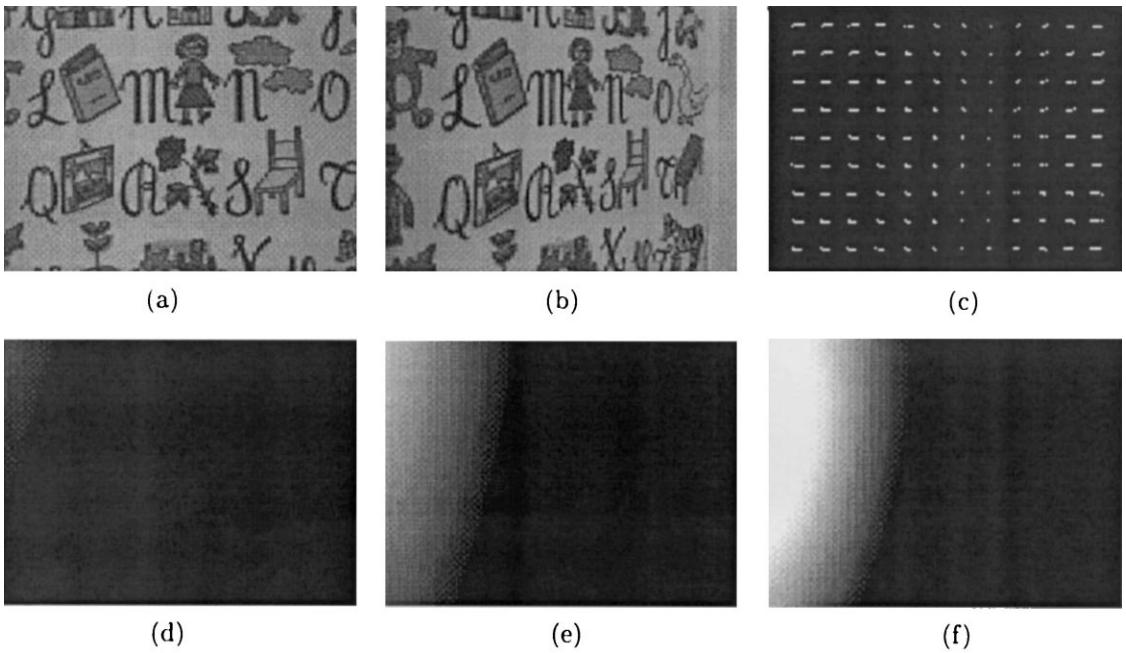


Fig. 8. A rotating plane. (a), (b): Two subsequent frames of the original sequence. (c): Computed optic flow. (d)–(f): Scaled depth, planar and spherical immediacies (brighter means closer).

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. The second row of Figs. 8(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 with respect to a planar fobject (it is evident, in Figs. 8(e) and (f), the image area corresponding to approaching motion). In fact, unlike Eqs. (7) and (8), the scaled depth immediacy bound of Eq. (4) fails to take into account the effect of 3D rotations on motion vectors, and uses only 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% with respect to the ground truth for the planar and spherical times to collision).

5.2. Time to collision from active contour deformations

Fig. 9 (left) 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 the average first-order parameters of image motion as in [5]. 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 hand starts moving at time $t = 0$ and stops after 4 s. The computed average collision immediacies are reported in Fig. 9 (right), showing that after a short transition time due to contour inertia, the value of the spherical and planar times to collision is quite close to the ground truth value (specifically, around 7 s to collision in the spherical case, and 20 s in the planar case during hand motion). Again, due to the fact that the slapping action involves more a hand rotation than

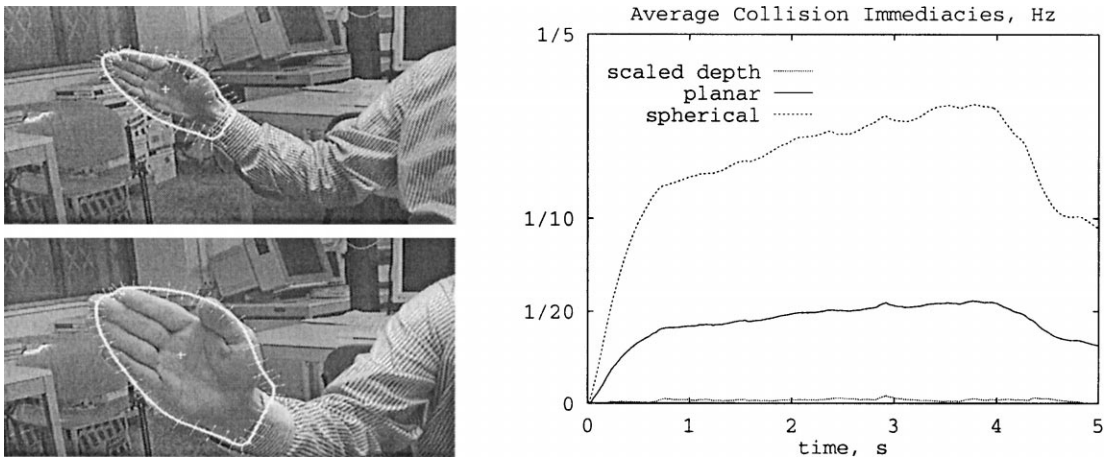


Fig. 9. A slapping hand. *Left*: two frames of the original sequence with superimposed active contour tracker. *Right*: spherical, planar and scaled depth collision immediacies.

a translation, the collision immediacy estimate based on scaled depth has at any time a very low value (hence of limited practical use) in all parts of the image.

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 visual periphery. Better still, time to collision can be estimated however wide be the sensor's FOV by suitable combinations of first-order motion field parameters, while scaled depth cannot. Two different wide FOV times to collision were introduced, each referring to a specific model of geometric layout of the observer; the two definitions converge to the usual scaled depth definition of time to collision in the presence of a narrow FOV. A closed form solution was also discussed to the problem of computing wide FOV time to collision with both traditional and space-variant sensors. Results of experiments with both dense and sparse visual measurements were finally described.

The work can be expanded in several directions. For example, a more general differential geometry formulation can be introduced to study the problem of visual parameter estimation in the case of general sensor shape and to prove the feasibility of the computational framework to applications involving nonrigid motions. Currently, a new model of space-variant sensor is being experimented with, which was explicitly designed so as to carry out the computations required by the approach with a minimum of computational effort. Also, an extensive experimentation of the approach is being carried out to demonstrate its suitability for applications in robotics, human–computer interaction and multimedia technology.

Acknowledgements

The author would like to thank Dr. Gabriele Baldi for his precious help in the videotaping experiments.

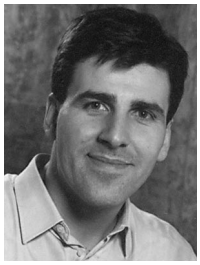
References

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

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

[3] R. Cipolla, A. Blake, Surface orientation and time to contact from image divergence and deformation, in: Proceedings of the Second European Conference on Computer Vision ECCV'92, 1992, pp. 187–202.

[4] C. Colombo, Time to collision from a natural perspective, Technical Report RT-199703-11, Dipartimento di Elettronica per l'Automazione, Brescia, Italy, April 1997.
[5] C. Colombo, E. Kruse, P. Dario, Control of camera motions from the planning of image contours, Robotics and Autonomous Systems 16 (1995) 29–38.
[6] C. Colombo, M. Rucci, P. Dario, Attentive behavior in an anthropomorphic robot vision system, Robotics and Autonomous Systems 12 (1994) 121–131.
[7] A. Giachetti, V. Torre, The use of optical flow for the analysis of non-rigid motions, International Journal of Computer Vision 18 (3) (1996) 255–279.
[8] J.J. Gibson, The Perception of the Visual World, Houghton Mifflin, Boston, MA, 1950.
[9] D.J. Heeger, A.D. Jepson, Subspace methods for recovering rigid motion I: algorithm and implementation, International Journal of Computer Vision 7 (2) (1992) 95–117.
[10] M. Kemp, The Science of Art. Optical Themes in Western Art from Brunelleschi to Seurat, Yale University Press, New Haven, CT, 1990.
[11] J.J. Koenderink, 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 (1975) 773–791.
[12] H.C. Longuet-Higgins, K. Prazdny, The interpretation of a moving retinal image, Proceedings of the Royal Society of London B 208 (1980) 385–397.
[13] R.C. Nelson, J. Aloimonos, Obstacle avoidance using flow field divergence, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (10) (1989) 1102–1106.
[14] D. Regan, Visual processing of four kinds of relative motion, Vision Research 26 (1986) 127–145.
[15] G. Sandini, V. Tagliasco, An anthropomorphic retina-like structure for scene analysis, Computer Graphics and Image Processing 14 (3) (1980) 365–372.
[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, 1992, pp. 2203–2208.
[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 (1990) 329–341.
[18] M. Tistarelli, G. Sandini, Dynamic aspects in active vision, Computer Vision, Graphics and Image Processing: Image Understanding 56 (1) (1992) 108–129.
[19] M. Tistarelli, 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) (1993) 401–410.
[20] E. Trucco, A. Verri, Introductory Techniques to 3D Computer Vision, Prentice-Hall, Upper Saddle River, NJ, 1998.



Carlo Colombo received an M.S. in Electronic Engineering from the University of Florence, Italy (1992) and a Ph.D. in Robotics from the Scuola Superiore di Studi Universitari e di Perfezionamento Sant'Anna, Pisa, Italy (1996). From 1996 to 1998 he was an Assistant Professor at the Faculty of Engineering, University of Brescia, Italy. In 1999 he joined the Faculty of Engineering of the University of Florence, Italy. Dr. Colombo is a member of IEEE and IAPR, and presently serves as the secretary of the IAPR Italian Chapter. Since 1995, Dr. Colombo has served both as organizer and program committee member for several conferences in the area of information technology. In 1997, he was guest editor of a special issue of *Robotics and Autonomous Systems* devoted to the Third International Symposium of Intelligent Robotic Systems SIRS'95, Pisa, Italy. His main research interests are in theoretical and applied computer vision, semi-autonomous robotics, advanced human–machine interfaces and multimedia systems design.