Time to Collision
from a Natural Perspective

Carlo Colombo

Dipartimento di Elettronica per l'Automazione

Università di Brescia

Via Branze 38, I-25123 Brescia, ITALY

E-mail columbus@bsing.ing.unibs.it

Abstract

In this paper, the problem of estimating time to collision from local (dense or sparse) motion field measurements using an arbitrarily wide field of view and without imposing constraints on either relative motion or surface orientation is addressed. Two different definitions of wide field of view time to collision are provided, referring respectively to a planar and spherical sensor geometry, and both reducing to scaled depth in the special case of narrow field of view. It is shown that, if no exterospecific assumption about relative motion or surface orientation can be made, then virtually all literature approaches proposed so far assume a narrow field of view (propriospecific) constraint to estimate time to collision from local measurements of first-order motion field structure. The main contribution of this work is to use elementary tools from differential geometry so as to relax such a constraint, and obtain a closed form solution to both the planar and spherical expressions of time to collision. It is also found that a limit for the narrow field of view constraint and scaled depth computation is forty degrees. A solution to the specific case of radial imaging geometry is finally derived, which provides an insight into the computational advantages of using polar and log-polar sensors to compute time to collision.

Keywords: Motion Analysis, Active and Robot Vision, Motion Parallax, Time to Collision and Relative Depth, Polar and Log-Polar Sensors, Imaging Geometry

Contents

List of Figures

Time to Collision from a Natural Perspective

Carlo Colombo

Dipartimento di Elettronica per l'Automazione

Via Branze 38, I-25123 Brescia, Italy

e-mail columbus@bsing.ing.unibs.it

April 30, 1997

Abstract

In this paper, the problem of estimating time to collision from local (dense or sparse) motion field measurements using an arbitrarily wide field of view and without imposing constraints on either relative motion or surface orientation is addressed. Two different definitions of wide field of view time to collision are provided, referring respectively to a planar and spherical sensor geometry, and both reducing to scaled depth in the special case of narrow field of view. It is shown that, if no exterospecific assumption about relative motion or surface orientation can be made, then virtually all literature approaches proposed so far assume a narrow field of view (propriospecific) constraint to estimate time to collision from local measurements of first-order motion field structure. The main contribution of this work is to use elementary tools from differential geometry so as to relax such a constraint, and obtain a closed form solution to both the planar and spherical expressions of time to collision. It is also found that a limit for the narrow field of view constraint and scaled depth computation is forty degrees. A solution to the specific case of radial imaging geometry is finally derived, which provides an insight into the computational advantages of using polar and log-polar sensors to compute time to collision.

1 Introduction

In the last few decades, the motion field analysis of time-varying image sequences has proved to be a powerful tool for the extraction of both geometric and kinematic information about a viewed scene [2], [38], [36], [14]. Specifically, it has been shown that most part of such visual information is embedded in the first-order local structure of motion fields, or motion parallax [17], [15], [7]. Although not sufficient to solve completely the structure from motion problem by means of local image observations [22], [40], motion parallax has been succesfully applied in computer vision to address difficult problems and applications such as image segmentation [5], [7], visual surveillance, robot navigation and obstacle avoidance [30], [34] and robot visuo-motor control [8], [9]. Quite recently, evidence has been presented that also the human visual system has specific sensitivities to motion parallax characteristic patterns such as dilatation and shear [31], thus supporting the idea that first-order motion fields play a crucial role in the execution of tasks such as visual exploration and heading direction control [13].

Having a wide field of view (FOV) is no doubt useful both to the biological and to the artificial observer, as a means of controlling the largest possible part of the visual environment:

1

this is the reason why most animals have a remarkably wide (up to nearly 360 degrees) FOV. In particular, motion information from the far sensor periphery can be used by an observer to redirect gaze to interesting regions of the environment, and to point out the presence of moving obstacles. The recent advances in hardware technology make it feasible to use a wider FOV than in the past at the same image resolution. Beside being widely present in nature, space-variant sensors are also particularly suited to computer vision applications allowing to have, at the cost of some loss of resolution in the image periphery, a wider field of view than traditional cameras for a fixed amount of image computation [32]. Yet, the wider is the camera FOV, the larger is, as compared to the corresponding image in a human retina, the distortion in the image periphery due to perspective effects [11]. In fact, while the photoreceptors in the human eye can be assumed to be arranged according to a roughly spherical layout, the surface of projection of electronic imaging sensors has — as the most part of the surfaces created by human technology, such as walls, tables, paintings etc. — a planar shape. Leonardo da Vinci was among the first to point out the lack of likelihood produced by perspective in frescos of large dimensions as compared to what he called *natural perspective* [16].

Several theories on visual motion analysis have been developed for an ideal spherical sensor as representing the whole "optic array" of visual directions [13], [27], [18], [29]. Indeed, a spherical sensor model strongly appeals to theorists of both human and computer vision, since natural perspective is biologically plausible and also greatly simplifies mathematical expressions. Several computer vision approaches exist which use spherical theories for planar implementations. The assumption of a spherical eye is — very 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 [22], [35]. Using a wide FOV would lead, in these cases, to gross visual parameters estimation errors especially in the far image periphery.

This paper describes a method for computing time to collision from a planar motion field with an arbitrarily wide FOV, in the presence of general rigid motion and surface orientation. In Section 2.2 it is shown that, about the optical axis, time to collision — i.e., informally, the time before the collision between the observer and the observed object — 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 do differ from one another. Two distinct definitions of time to collision for an arbitrarily wide FOV are provided, including scaled depth as a particular case (narrow FOV), and referring respectively to a planar and spherical sensor geometry.

A comparison of the most popular constraints used for estimating scaled depth from local observations, carried out in Section 3.2 using a uniform notational framework to describe the first-order motion field structure, shows that virtually all approaches estimate time to collision from a linear combination of motion field invariants. How these invariants are distorted by perspective at large visual angles, and how to recover the invariants corresponding to natural projection in a closed form by projecting — by the use of elementary tools from differential geometry — the planar motion field structure onto the sphere is the subject of Section 4.1.

Finally, Section 4.2 offers a discussion of the results, addressing *(i)* the question of up to

2

which FOV scaled depth and time to collision can be used interchangeably, *(ii)* aspects related to the algorithmic implementation of the closed form result and to the difference between the two most common field structure representations, and *(iii)* which are the computational advantages intrinsic to polar and log-polar sensor geometry. The Appendices provide mathematical details about *(A)* how the first-order structure of a field can be characterized using either the topological or the tensor calculus approach, and *(B)* how to compute the planar motion field invariants using polar and log-polar sensors.
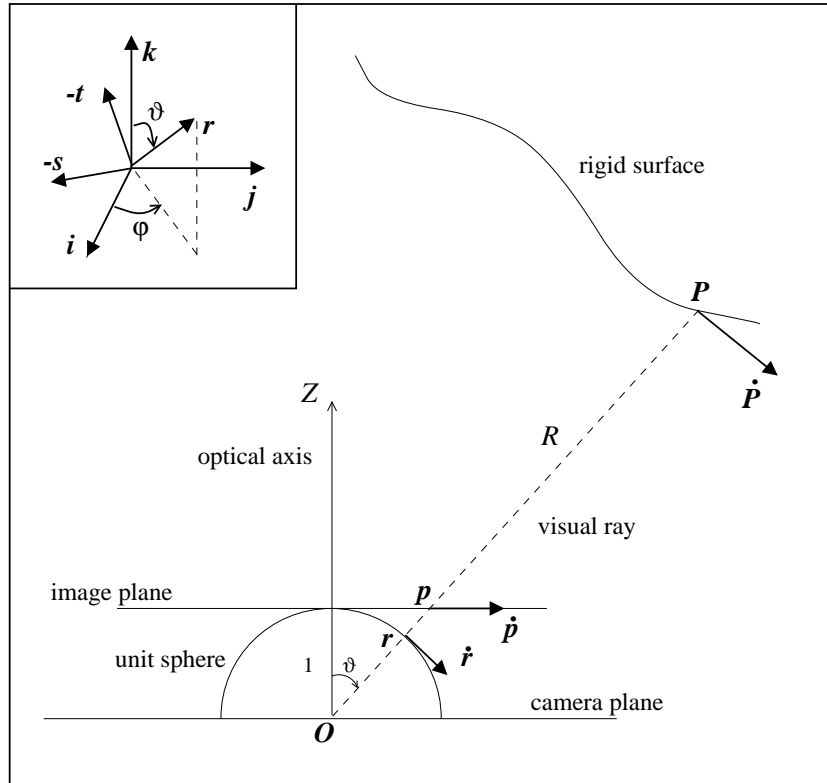
Figure 1: Image projection and relative motion. For the sake of simplicity, the plane common to all velocity vectors is assumed equal to the plane of the figure, which contains the optical axis and the visual ray.

2 Time to Collision Revisited

2.1 Preliminaries and Notation

Let the imaged scene be composed of rigid surfaces in relative motion w.r.t. the camera. Geometry of image projection and camera-surface relative kinematics are illustrated in Fig. 1.

A perspective projection model with camera frame $\{O\boldsymbol{ijk}\}$ is assumed, with $\boldsymbol{O}$ the center of projection and $\boldsymbol{k}$ the depth axis. The image 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

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

where, without loss of generality, focal length is set to 1. The relative velocity at $\boldsymbol{P}$ can be expressed as

$$\dot{\boldsymbol{P}} = \boldsymbol{V} + \boldsymbol{\Omega} \times \boldsymbol{P} \quad , \tag{2}$$

where $(\boldsymbol{V}, \boldsymbol{\Omega})$ is the surface-camera relative twist screw, $\times$ denotes the usual vector product, 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*

$$\dot{\boldsymbol{p}} = \boldsymbol{k} \times \left[\left(\frac{\boldsymbol{V}}{Z} + \boldsymbol{\Omega} \times \boldsymbol{p} \right) \times \boldsymbol{p} \right] \tag{3}$$

whose component along $\boldsymbol{k}$ is identically zero: $\dot{\boldsymbol{p}} = u\,\boldsymbol{i} + v\boldsymbol{j}$. The unit vector

$$r = \frac{\boldsymbol{P}}{R} \tag{4}$$

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* $\dot{\boldsymbol{r}} = u'\,\boldsymbol{t} + v'\boldsymbol{s}$ s.t.

$$\dot{\boldsymbol{r}} = \boldsymbol{r} \times \left[\left(\frac{\boldsymbol{V}}{R} + \boldsymbol{\Omega} \times \boldsymbol{r} \right) \times \boldsymbol{r} \right] \tag{5}$$

is obtained in the plane tangent to the unit sphere at $\boldsymbol{r}$. The planar and spherical motion fields are clearly identical at the 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 $\vartheta(\boldsymbol{r}) = \arccos(\boldsymbol{k} \cdot \boldsymbol{r})$ between the two rays increases moving from image origin towards image periphery. The camera's *field of view* FOV $\in [0, \pi]$ is defined as twice the maximum angle between the optical and visual rays

$$\mathrm{FOV} = 2 \max_{\boldsymbol{r}} \vartheta(\boldsymbol{r}) \quad , \tag{6}$$

as related to the actual focal length and physical dimensions of the imaging sensor.

Let $\overline{\boldsymbol{v}} = [\,\boldsymbol{v} \cdot \boldsymbol{i} \;\; \boldsymbol{v} \cdot \boldsymbol{j} \;\; \boldsymbol{v} \cdot \boldsymbol{k}\,]^{\mathrm{T}}$ and $\widetilde{\boldsymbol{v}} = [\,\boldsymbol{v} \cdot \boldsymbol{t} \;\; \boldsymbol{v} \cdot \boldsymbol{s} \;\; \boldsymbol{v} \cdot \boldsymbol{r}\,]^{\mathrm{T}}$ be the coordinate-dependent representations of a vector $\boldsymbol{v}$: these are one-one related by $\overline{\boldsymbol{v}} = \mathrm{C}\,\widetilde{\boldsymbol{v}}$, where C is the orthogonal matrix

$$\mathrm{C} = \begin{bmatrix} \cos\vartheta\cos\varphi & -\sin\varphi & \sin\vartheta\cos\varphi \\ \cos\vartheta\sin\varphi & \cos\varphi & \sin\vartheta\sin\varphi \\ -\sin\vartheta & 0 & \cos\vartheta \end{bmatrix} \quad . \tag{7}$$

Co-latitude $\vartheta(\boldsymbol{r})$ and longitude $\varphi(\boldsymbol{r})$ angles are obtained directly from image data, as

$$[\,x \;\; y\,]^{\mathrm{T}} = \tan\vartheta\,[\,\cos\varphi \;\; \sin\varphi\,]^{\mathrm{T}} \quad . \tag{8}$$

2.2 Scaled Depth vs Time(s) to Collision

In the computer vision literature, the terms "time to collision" and "scaled depth" are used interchangeably [3], [20], [33], and referred to the scalar field

$$t_z = -\frac{Z}{\boldsymbol{V}\cdot\boldsymbol{k}} \tag{9}$$

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).

A viewpoint expressed in this paper is that *as eq. (9) only refers to translation and disregards rotation, it provides an exact definition of scaled depth, while it cannot be considered but an approximation of time to collision.* This makes the task of recovering scaled depth more difficult than determining time to collision, as we prove below. Indeed, recovering scaled depth is tantamount to solving completely the structure from motion problem since, as evident from eq. (3), structure (expressed as a depth) and relative translation can only be recovered from the motion field up to a common scale factor — the well known speed-scale ambiguity [6], [4]. Two major difficulties arise in the structure from motion problem: *(i)* the segmentation of the image in a set of regions corresponding to single rigid surfaces in the scene [1]; *(ii)* the recovery of rigid motion for each surface, by separating the translational and rotational parts [22], [14].

Contrariwise, being it quite natural to define time to collision as a scalar field indicating the time for each surface point to meet the observer, and since the velocity of a surface point is due to both rigid translation and rotation, there should be actually no need for separating rotation from translation — point *(ii)* above — to obtain time to collision. Better still, rather than requiring a segmentation step — point *(i)* above —, time to collision should be of help as a powerful source of information to carry out the segmentation task needed to address the structure from motion problem.

Several authors (e.g. [28], [10]) have recently argued against attempting to measure scaled depth without solving explicitly for rigid motion. However, this can be done within a certain degree of approximation provided that some constraints are put on relative motion and/or viewing angle (see Section 3.2).

There are, of course, diverse possible definitions of time to collision matching the above criteria, 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, they will be stated two definitions of time to collision, which rely on two different observer models: the planar and the spherical observer.

Definition 1 (Planar Time to Collision)   *The Planar Time to Collision is the time t_p it would take a point $\boldsymbol{P}$ to reach the camera plane by traveling at a uniform velocity $\dot{\boldsymbol{P}}\cdot(-\boldsymbol{k})\,\boldsymbol{k}$, i.e.,*

$$t_p = -\frac{Z}{\dot{\boldsymbol{P}}\cdot\boldsymbol{k}} \quad . \tag{10}$$

The concept of planar observer directly refers to the perspective model of imaging projection, with its privileged direction in space (the optical axis) providing the normal to the surface
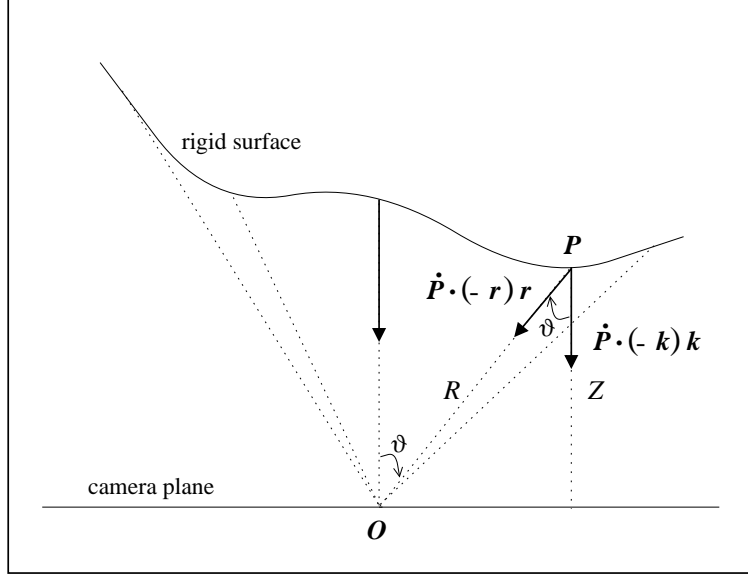
Figure 2: Geometry for the definition of planar and central time to collision.

of collision (the 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 (see also Fig. 2). Yet, in a robotic application planar time to collision could be appropriate in the case of a robot with dominant transversal dimensions.

Eq. (10) is the natural extension of eq. (9) to the general case of relative rototranslation. The rotational component of relative velocity comes into play anywhere far from image origin $\vartheta = 0$, thus differentiating t_p from scaled depth as the denominator

$$\dot{\boldsymbol{P}} \cdot \boldsymbol{k} = \boldsymbol{V} \cdot \boldsymbol{k} - R \sin \vartheta \, \boldsymbol{\Omega} \cdot \boldsymbol{s} \quad , \tag{11}$$

where $\sin \vartheta \, \boldsymbol{s} = \boldsymbol{k} \times \boldsymbol{r}$, is no longer constant over the rigid surface. Using eqs. (9), (10) and (11), the mismatch between scaled depth and planar time to collision can be expressed in terms of their reciprocals as

$$t_p^{-1} - t_z^{-1} = \tan \vartheta \, \boldsymbol{\Omega} \cdot \boldsymbol{s} \quad , \tag{12}$$

where $Z = R \cos \vartheta$. Ideally, the mismatch is zero if $\boldsymbol{\Omega}$ vanishes identically (pure translation) or is coplanar with $\boldsymbol{k}$ and $\boldsymbol{r}$, or in any case at the image origin. In practice, planar time to collision approximates scaled depth in the case of dominant translation $\|\boldsymbol{V}\| \gg \|R\,\boldsymbol{\Omega}\|$ or, when nothing can be said about $\boldsymbol{\Omega}$, when the FOV is sufficiently small to assume that rotation is uneffective.

$\square$

Definition 2 (Central Time to Collision) *The Central (or 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 $\dot{\boldsymbol{P}} \cdot (-\boldsymbol{r}) \, \boldsymbol{r}$, i.e.,*

$$t_r = -\frac{R}{\dot{\boldsymbol{P}} \cdot \boldsymbol{r}} \quad . \tag{13}$$

To the spherical observer using the natural perspective projection of eq. (4), all visual directions look alike. Eq. (13) provides a convenient way of defining time to collision having a sphere as the imaging surface (see again Fig. 2). Such a definition has a strict relationship with the *time to flight* used with sonar-based systems [19], and is best suited to robotic applications involving a robot with no dominant dimensions. A slight change to eq. (13) could be made to define time to collision w.r.t. the *minimum enclosing sphere* around the robot. Notice that, differently from the planar time to collision, t_r is apparently not affected by relative rotation, as $\dot{\boldsymbol{P}} \cdot \boldsymbol{r} = \boldsymbol{V} \cdot \boldsymbol{r}$. However, this term is not a constant, as it depends on the direction of the visual ray, and when we try to express it in a ray-independent fashion using the camera reference frame, rotation reappears:

$$\dot{\boldsymbol{P}} \cdot \boldsymbol{r} = \sec \vartheta \, \dot{\boldsymbol{P}} \cdot \boldsymbol{k} + \tan \vartheta \, \boldsymbol{V} \cdot \boldsymbol{t} + R \tan \vartheta \, \boldsymbol{\Omega} \cdot \boldsymbol{s} \quad , \tag{14}$$

where $\boldsymbol{t} = \boldsymbol{s} \times \boldsymbol{r}$. As a result, as with t_p, t_r cannot be considered as a scaled depth (or, better still, a scaled range, R being used as depth): its mismatch with scaled depth evaluates as

$$t_r^{-1} - t_z^{-1} = - \tan \vartheta \, \frac{\boldsymbol{V} \cdot \boldsymbol{t}}{R} \tag{15}$$

and, assuming that nothing is known about $\boldsymbol{V}$, is approximately zero in the case of a narrow FOV.

$\square$

Eqs. (10) and (13) offer a generalization to the time to collision of eq. (9) given an arbitrary motion and field of view. Specifically, all three equations become identical in the degenerate case FOV $= \vartheta = 0$. Section 3.2 investigates on the operating conditions under which scaled depth and time to collision can be (albeit approximately) confused with one another — a quantitative result being also derived at Section 4.2. This will prepare for our main result of Section 4.1, explaining how to get an estimate of both the planar and spherical time to collision without making any assumption on the FOV.

3 Scaled Depth Approximations

3.1 Computational Framework

Our computational framework assumes that the planar motion field and its first spatial derivatives are available at a given number of image points. Such *local* measurements can be either dense (e.g., the result of an optic flow image sequence analysis [2], [24] or sparse (as resulting from the tracking of point, line or curve features from frame to frame [8], [23]) in the image. A dense field allows to obtain a dense scaled depth map and also to perform segmentation [7], [5]; yet, it is often computationally more expensive and less robust than a sparse field.

Since our main task is to get a reasonable estimate of time to collision without solving for motion and scene structure, it is not required that a segmentation step be performed so as to associate each single measurement to a specific world surface. For the same reason, our framework is limited to using spatial derivatives of the field up to the first order. Indeed, theoretically speaking, a precise estimate of time to collision (together with an estimate of

7

motion and structure) can be derived using a second-order motion field [18], [22], [40]. However, in practice, second spatial derivatives of image motion (corresponding to third spatial derivative of image intensity!) are quite difficult to compute in a robust fashion, and also require a support image neighborhood so large that it is likely to cross the boundaries of discontinuity between any two different imaged surfaces, thus proving grossly inaccurate time to collision estimates at most of the evaluation points.

First-Order Planar Motion Field. Using the coordinate-dependent vector notation introduced in Section 2.1, the planar motion field of eq. (3) is expressable as a function of image position p as

$$\begin{bmatrix} u(x,y) \\ v(x,y) \end{bmatrix} = \frac{1}{z} \begin{bmatrix} 1 & 0 & -x \\ 0 & 1 & -y \end{bmatrix} \overline{V} + \begin{bmatrix} -xy & (1+x^2) & -y \\ -(1+y^2) & xy & x \end{bmatrix} \overline{\Omega} \quad , \qquad (16)$$

with $z(x,y)$ s.t. $z(X/Z, Y/Z) = Z$.

The motion field can be approximated, in a convenient neighborhood of an image point p, by the motion field vector at p and the first-order field local spatial structure, encoded in the four differential invariants divergence, curl, and (two components of) deformation

$$\text{div} = u_x + v_y \ ; \quad \text{curl} = -u_y + v_x \ ; \quad \text{defx} = u_x - v_y \ ; \quad \text{defy} = u_y + v_x \quad , \qquad (17)$$

where subscripts denote partial differentiation (see Appendix A). The differential invariants of the planar motion field at p can be compactly represented through the 4×1 matrix $\mathrm{D} = [\,\text{div} \ \text{curl} \ \text{defx} \ \text{defy}\,]^{\mathrm{T}}$: they depend on image position, visible surface depth and orientation, and relative twist screw. Surface depth and orientation are related to the parameters characterizing the first-order differential structure of the visual surface in the neighborhood of point P (the tangent plane at P): these are the depth gradient components $p = \partial Z/\partial X$ and $q = \partial Z/\partial Y$, and the intercept $c = z\,(1 - p\,x - q\,y)$ with the optical axis. The gradient components can be expressed in terms of the slant $\sigma \in [0, \pi/2]$ and tilt $\tau \in [0, 2\pi]$ angles of the tangent plane, as

$$[\,p \ q\,]^{\mathrm{T}} = -\tan\sigma\,[\,\cos\tau \ \sin\tau\,]^{\mathrm{T}} \quad , \qquad (18)$$

being $n = -p\,i - q\,j + k$ a vector normal to the surface at P. Observing that $\partial(z^{-1})/\partial x = -p/c$ and $\partial(z^{-1})/\partial y = -q/c$, we get from eqs. (16) and (17)

$$\mathrm{D} = \Pi_V(x,y;p,q,c)\,\overline{V} + \Pi_\Omega(x,y)\,\overline{\Omega} \quad , \qquad (19)$$

where

$$\Pi_V(x,y;p,q,c) = \frac{1}{c} \begin{bmatrix} -p & -q & -2+3p\,x+3q\,y \\ q & -p & -q\,x + p\,y \\ -p & q & p\,x - q\,y \\ -q & -p & q\,x + p\,y \end{bmatrix} \ ; \quad \Pi_\Omega(x,y) = \begin{bmatrix} -3\,y & 3\,x & 0 \\ x & y & 2 \\ y & x & 0 \\ -x & y & 0 \end{bmatrix} \quad . \qquad (20)$$

3.2 Analysis of Constraints for Scaled Depth

This Section is devoted to analyze in some detail the most common assumptions used in the recent literature to estimate time to collision as a scaled depth from local observations.

Since several approaches are discussed, a uniform notation based on first-order motion field invariants has been devised so as to facilitate comparisons.

Let us introduce the *directional deformation* (in the image plane direction α) as the quantity

$$\mathrm{def}^{\alpha} = \cos 2\alpha\, \mathrm{defx} + \sin 2\alpha\, \mathrm{defy} = \mathrm{def} \cos(2\mu - 2\alpha) \quad , \tag{21}$$

being $\mathrm{defx} = \mathrm{def} \cos 2\mu$ and $\mathrm{defy} = \mathrm{def} \sin 2\mu$. The quantity $\mathrm{def} = \mathrm{def}^{\mu} \geq 0$, usually referred to as deformation, is — together with div and curl — a scalar invariant, as it does not change if image coordinates are rotated (see Appendix A).

A general expression for scaled depth at a generic image point is obtained from eq. (19) in terms of directional deformations:

$$t_z^{-1} = \frac{\mathrm{div} - \mathrm{def}_V^{\tau} - 3\,\mathrm{def}_{\Omega}^{\varphi}}{2} \quad . \tag{22}$$

In eq. (22), def_V and def_{Ω} are due respectively to the translational ($\boldsymbol{V}$) and rotational ($\boldsymbol{\Omega}$) components of relative rigid motion, being $\mathrm{def}^{\alpha} = \mathrm{def}_V^{\alpha} + \mathrm{def}_{\Omega}^{\alpha}$. We observe from eq. (20) that the rotation-related deformation depends only on $\boldsymbol{\Omega}$ and image coordinates, vanishing in the case of pure translation or, whichever $\boldsymbol{\Omega}$, at the image origin. The translation-related deformation depends instead on $\boldsymbol{V}/Z$ and $[p\ q]^{\mathrm{T}}$, vanishing identically either if $\|\boldsymbol{V}\| = 0$ or if the tangent plane at $\boldsymbol{P}$ and the image plane are mutually parallel ($p=q=0$).

As mentioned earlier, any attempt to recover scaled depth from local motion field computations must rely on some simplifying assumption about surface-camera relative motion and/or viewing angle. In the following, the popular *narrow field of view* [30], [35], [8], [34], *dominant translation* [7], [5], [12] and *frontoparallel surface* [34], [37] constraints are discussed.

Narrow Field of View. 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, i.e.,

$$\mathrm{D} \approx \mathrm{D}_o \quad , \tag{23}$$

where $D_o = [\mathrm{div}_o\ \mathrm{curl}_o\ \mathrm{defx}_o\ \mathrm{defy}_o]^{\mathrm{T}}$ is the matrix of motion field invariants at the image origin. Recalling eq. (19), we have:

$$\mathrm{D}_o = \Pi_V(0, 0; p_o, q_o, c_o)\, \overline{\boldsymbol{V}} + \Pi_{\Omega}(0, 0)\, \overline{\boldsymbol{\Omega}} \quad . \tag{24}$$

From eqs. (18) through (21) we see that, at the origin, $2\mu_o = \psi + \tau_o$, where $\psi = \arctan(\boldsymbol{V} \cdot \boldsymbol{j}/\boldsymbol{V} \cdot \boldsymbol{i})$ is the angle between the translational velocity component parallel to the image plane and the x-axis. We also have, from eq. (22),

$$t_{z,o}^{-1} = \frac{\mathrm{div}_o - \mathrm{def}_o^{\tau_o}}{2} \quad , \tag{25}$$

where from eq. (21) it follows that $\mathrm{def}_o^{\tau_o} = \mathrm{def}_o \cos(\psi - \tau_o)$. Assuming that the orientation of the visible surface at $\boldsymbol{p}$ is unknown, a bound on scaled depth is obtained

$$t_{z,o}^{-1} = \frac{\mathrm{div}_o \pm \mathrm{def}_o}{2} \quad , \tag{26}$$

9

reflecting the fact that the *difference* between surface tilt and direction of relative translation parallel to the image is, differently from their sum, undecidable if only local first-order information is available [35], [8].

The bound of eq. (26) generalizes the *directional divergence* algorithm, proposed earlier in [30], based on an iterative search and exploiting the invariant properties of the divergence to image coordinate rotations. According to such an algorithm, an approaching motion can be detected, and its scaled depth estimated, by iteratively searching, if it exists, for the image plane direction $\phi_{\max}$ s.t. the directional divergence operator $\mathrm{div}_\phi = \cos^2\phi\,u_x + \sin^2\phi\,v_y + \sin\phi\cos\phi\,[u_y + v_x]$ is maximum and positive. Since it is easy to show that

$$\mathrm{div}_\phi = \frac{\mathrm{div} + \mathrm{def}^\phi}{2} \quad , \tag{27}$$

it is clear that the directional divergence algorithm consists basically in progressively rotating the image coordinate frame in eq. (25) until finding the value of ϕ by which $\max_\phi(\mathrm{div}+\mathrm{def}^\phi) = \mathrm{div} + \mathrm{def} \geq 0$, thus being equivalent to using only the upper bound in eq. (26). Notice that, if a positive maximum exists, then it is attained of course when the directional deformation is maximum and equal to the deformation, i.e. at $\phi_{\max} = \mu$ $(\approx \mu_o)$: such an angle can be computed in a noniterative fashion from the two deformation components defx and defy.

Several variations to the basic bound of eq. (26) have been proposed so far, attempting to overcome the orientation ambiguity by coupling the narrow FOV constraint of eq. (23) with additional contraints such as fixation, partial knowledge of motion, etc. [8], [37].

Dominant Translation.   In the case of dominant translation it is assumed that rotation be negligible w.r.t. translation, i.e. $\mathrm{def}_\Omega^\varphi = 0$ so that $\mathrm{def}_V = \mathrm{def}$. As a result, eq. (22) reduces to

$$t_z^{-1} = \frac{\mathrm{div} - \mathrm{def}^\tau}{2} \quad , \tag{28}$$

thus allowing to bound scaled depth *at any image point* as

$$t_z^{-1} = \frac{\mathrm{div} \pm \mathrm{def}}{2} \tag{29}$$

if the surface orientation τ is unknown. Besides, as mentioned before in the case of pure translation the structure from motion problem is greatly simplified, being it easy to estimate the *epipole* $\boldsymbol{p}_e = \boldsymbol{V}/(\boldsymbol{V}\cdot\boldsymbol{k})$ and thus to recover scaled depth exactly [22], [6], [20]. As a special case, we consider now estimating scaled depth at the epipole [7], [12]: as easily seen from eq. (3), this is the unique image point at which the motion field annihilates (singular point) when $\boldsymbol{\Omega} = \boldsymbol{0}$. A rapid inspection to eq. (20) allows us to check that at the epipole the curl and deformation terms vanish identically whichever the surface orientation: curl = defx = defy = 0. Hence, being

$$\mathrm{M} = \frac{1}{2}\begin{bmatrix} \mathrm{div} + \mathrm{defx} & -\mathrm{curl} + \mathrm{defy} \\ \mathrm{curl} + \mathrm{defy} & \mathrm{div} - \mathrm{defx} \end{bmatrix} \tag{30}$$

the Jacobian matrix describing the linear part of the motion field — see also Appendix A —, under dominant translation scaled depth at the epipole can be computed from

$$t_{z,e}^{-1} = \frac{\mathrm{div}_e}{2} = \lambda_e \quad , \tag{31}$$

where the latter equality arises from matrix M of eq. (30) being at the epipole a multiple of the identity.

Frontoparallel Surface. A very special case is that of a frontoparallel surface, i.e. when at some image point $\boldsymbol{p}$ a vanishing depth gradient occurs: $p = q = 0$. At such a point, the narrow FOV constraint can be relaxed, and scaled depth can be computed exactly *whichever the value of* ϑ. With the help of eq. (20) it is possible to verify that in the frontoparallel condition it holds $\mathrm{def}_V = 0$ and, by consequence, $\mathrm{def}_\Omega = \mathrm{def}$, so that

$$t_z^{-1} = \frac{\mathrm{div} - 3\,\mathrm{def}^\varphi}{2} \quad . \tag{32}$$

This is exactly the formula proposed in [37] to compute scaled depth using polar (ρ, φ) and log-polar (ξ, γ) coordinates:

$$t_z^{-1} = 2\left(\frac{\dot{\rho}}{\rho} + \frac{\partial \dot{\varphi}}{\partial \varphi}\right) - \frac{\partial \dot{\rho}}{\partial \rho} = \dot{\xi}\ln a - \frac{\partial \dot{\xi}}{\partial \xi} + 2\frac{\partial \dot{\gamma}}{\partial \gamma} \tag{33}$$

(for a proof, see Appendix B). It is worth noting that, although the frontoparallel surface constraint is a very powerful one (allowing to compute scaled depth without using bounds, and above all even with a wide FOV), yet it is also a very strong one, so that it is likely to produce large systematic errors in practical applications.

4 Estimating Time to Collision

4.1 Time to Collision with a Wide Field of View

From the previous Section it appears that the stronger the operational constraints are, the easier is the scaled depth estimation process, and the higher the probability of gross systematic errors when the constraints fail to be met perfectly.

As already evident from Section 2.2, in the presence of a narrow FOV scaled depth, planar and central time to collision are approximately equal:

$$t_r \approx t_p \approx t_z \qquad \text{if } \sin\vartheta \ll 1 \quad , \tag{34}$$

being $t_r = t_p = t_z$ at the image origin. Among the constraints used to compute time to collision from local observations discussed before, only the narrow FOV constraint applies to conditions related to the viewer's conditions only. It is clear that only such kind of *propriospecific* constraints should be used when nothing can be said about camera-scene relative geometry and motion (*exterospecific* conditions). As this is the most general case, in the following we will not be making any assumption on relative motion or surface orientation, while we will show how to relax the narrow FOV constraint and obtain estimates of planar and central time to collision which apply at any co-latitude.

Let us begin with the following result, establishing a correspondence between planar and spherical motion fields in terms of field vectors and first-order structures.

Theorem 1 (Connection of Planar and Spherical Motion Fields) *The planar and spherical motion fields are one-one related by*

$$[u'\ v']^{\mathrm{T}} = \mathrm{H}\,[u\ v]^{\mathrm{T}} \quad , \tag{35}$$

11

where

$$H = \begin{bmatrix} \cos^2 \vartheta & 0 \\ 0 & \cos \vartheta \end{bmatrix} \begin{bmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{bmatrix} \quad . \tag{36}$$

The first-order structures of the two fields, encoded in the Jacobian matrices M *and* M', *are one-one related by*

$$M' = H\,M\,H^{-1} + K' \quad , \tag{37}$$

where

$$K' = -\tan \vartheta \begin{bmatrix} 2\,u' & 0 \\ v' & u' \end{bmatrix} \quad . \tag{38}$$

Proof:

Combining eqs. (1) and (4), and recalling that $Z = R \cos \vartheta$, we obtain

$$\boldsymbol{r} = \cos \vartheta \, \boldsymbol{p} \quad . \tag{39}$$

Substituting this into eq. (5), by direct comparison with eq. (3) we get $\dot{\boldsymbol{r}} \times \boldsymbol{r} = \cos^2 \vartheta \, \dot{\boldsymbol{p}} \times \boldsymbol{p}$. Eq. (35) is finally obtained by expressing all vectors in the same coordinates.

To prove the second part of the result, we look for the relationship between the matrices D and D', encoding the spherical and planar first-order motion field structures at $\boldsymbol{r}$ and $\boldsymbol{p}$, respectively. Let us refer to the plane tangent to the unit sphere at $\boldsymbol{r}$ as the image plane of a *virtual camera* with optical axis $\boldsymbol{r}$: an equation similar to eq. (24) can then be written at the origin of the virtual image plane. Explicitly:

$$D' = \Sigma_V(p', q', R)\,\widetilde{\boldsymbol{V}} + \Sigma_\Omega\,\widetilde{\boldsymbol{\Omega}} \quad , \tag{40}$$

where $\Sigma_V(p', q', R) = \Pi_V(0,0;p',q',R)$ and $\Sigma_\Omega = \Pi_\Omega(0,0)$. Eq. (40) was first introduced in a different form in [18]. The values p' and q' are the visible surface orientation parameters at $\boldsymbol{P}$ as measured from the virtual camera:

$$\begin{bmatrix} p' \\ q' \end{bmatrix} = -\frac{1}{\boldsymbol{n} \cdot \boldsymbol{r}} \begin{bmatrix} \boldsymbol{n} \cdot \boldsymbol{t} \\ \boldsymbol{n} \cdot \boldsymbol{s} \end{bmatrix} \quad . \tag{41}$$

Let us now observe that the point on the unit (hemi-)sphere $\boldsymbol{r}$ can be thought of as the virtual perspective projection of $\boldsymbol{p}$ according to eq. (39), which is nothing but another way to express natural perspective. A basic result from differential geometry states that, if a smooth map exists between two manifolds — in our case, the manifolds are the image plane and the unit sphere, and the map is natural perspective —, then the tangent space at corresponding points in the two manifolds are related by the *derivative* of the same map, which is always a linear map [25]. As the planar and spherical motion fields belong respectively to the tangent space at $\boldsymbol{p}$ (the camera plane) and $\boldsymbol{r}$ (the virtual camera plane), we are allowed to express the relationship between the two differential invariant sets as a linear map of the kind

$$D' = A\,D + B_V\,\widetilde{\boldsymbol{V}} + B_\Omega\,\widetilde{\boldsymbol{\Omega}} \quad , \tag{42}$$

with A and the B's some 4×4 and 4×3 matrices, respectively, depending exclusively on the relative orientation (ϑ, φ) between the planar and spherical frames. Using eqs. (19) and (40), and taking into account that eq. (42) must hold whichever the rigid motion, we obtain

$$\Sigma_V(p', q', R) = A\,\Pi_V(x, y; p, q, c)\,C + B_V \quad ; \quad \Sigma_\Omega = A\,\Pi_\Omega(x, y)\,C + B_\Omega \quad , \tag{43}$$

which can be used to solve for A, B_V and B_Ω. After some algebraic manipulation we finally come up with

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & m_\vartheta & -n_\vartheta \sin 2\varphi & n_\vartheta \cos 2\varphi \\ 0 & 0 & \cos 2\varphi & \sin 2\varphi \\ 0 & n_\vartheta & -m_\vartheta \sin 2\varphi & m_\vartheta \cos 2\varphi \end{bmatrix} \tag{44}$$

and

$$B_V \, \widetilde{\boldsymbol{V}} + B_\Omega \, \widetilde{\boldsymbol{\Omega}} = -\tan \vartheta \left[\, 3 \, u' \;\; v' \;\; u' \;\; v' \, \right]^{\mathrm{T}} \quad , \tag{45}$$

with

$$m_\vartheta = \tfrac{1}{2} \left(\sec \vartheta + \cos \vartheta \right) \;\; ; \quad n_\vartheta = \tfrac{1}{2} \left(\sec \vartheta - \cos \vartheta \right) \quad . \tag{46}$$

The final result of eq. (37) easily follows by rewriting eq. (42) in the matrix form of eq. (30).

$\square$

The result above allows us to *project the planar motion field and its first-order structure onto the unit sphere*. This is essential to computing the planar and central time to collision with a wide field of view, as we show below.

Theorem 2 (Bounds on Wide FOV Time to Collision) *The central time to collision is bounded, at any co-latitude angle ϑ, by the divergence and the deformation of the spherical motion field, i.e.,*

$$t_r^{-1} = \frac{\mathrm{div}' \pm \mathrm{def}'}{2} \quad , \tag{47}$$

where

$$\mathrm{div}' = \mathrm{div} - 3 \tan \vartheta \, u' \tag{48}$$

and

$$\mathrm{def}' = \left[\left(\mathrm{def}^\varphi - \tan \vartheta \, u' \right)^2 + \left(m_\vartheta \, \mathrm{def}^{\varphi + \pi/4} + n_\vartheta \, \mathrm{curl} - \tan \vartheta \, v' \right)^2 \right]^{1/2} \quad . \tag{49}$$

The planar time to collision can be obtained at any co-latitude from the knowledge of the central time to collision and of the spherical motion field component along $\boldsymbol{t}$, i.e.,

$$t_p^{-1} = t_r^{-1} + \tan \vartheta \, u' \quad . \tag{50}$$

Proof:

We prove the second part of the result first. Subtracting eq. (15) from eq. (12) we obtain

$$t_p^{-1} - t_r^{-1} = \tan \vartheta \left[\frac{\boldsymbol{V} \cdot \boldsymbol{t}}{R} + \boldsymbol{\Omega} \cdot \boldsymbol{s} \right] \quad . \tag{51}$$

Recalling that $u' = \dot{\boldsymbol{r}} \cdot \boldsymbol{t}$ and using eq. (5), eq. (50) follows.

That t_r can be estimated as in eq. (47) is easily proved using the concept of virtual camera introduced earlier. Notice first that eq. (47) has the same form of eq. (26), which holds at the image origin $\boldsymbol{p} = \boldsymbol{k}$ where $t_r = t_z$. Indeed, eq. (47) refers to the image origin $\boldsymbol{p} = \boldsymbol{r}$ of the virtual camera having $\boldsymbol{r}$ as optical axis, thus including eq. (26) as the special case $\boldsymbol{r} = \boldsymbol{k}$. The invariants appearing in eq. (47) are those of the spherical motion field: these coincide with the planar invariants at the image origin. A close look to eqs. (44) and (45) confirms this last fact, since at the image origin ($\vartheta = \varphi = 0$) matrix A is the identity and the B's are zero.

$\square$

13

The results of this Section also clarify why, as anticipated in Section 2.2, scaled depth cannot be estimated simply as a combination of first-order invariants of either the planar or the spherical motion field. Referring again to eq. (22), we notice in fact that, although the bound on the sum

$$\text{def}_V^\tau + \text{def}_\Omega^\varphi = \text{def}'^{\tau'} + \tan\vartheta\, u' \tag{52}$$

allows to express eq. (22) in terms of only one of def_V^τ and $\text{def}_\Omega^\varphi$, it is not possible to obtain a similar bound for their difference. As a result, the two directional divergencies in eq. (22) cannot be individually bounded: to do that, translation and rotation should be decoupled.

4.2   Discussion of Results

Computer Simulation.   A simulation environment provided with an XWindow graphic interface was developed to test the feasibility of time to collision approaches and constraints under a wide range of operating conditions, such as relative motion, surface orientation, and field of view.

Figs. 3 *(a)–(d)* illustrate the results for the case of a planar visible surface rototranslating rigidly w.r.t. the camera. The axis of rotation passes through point $\boldsymbol{Q} = c\boldsymbol{k}$ in space, with $c = 100$ flu (focal length units). As a result, relative translation is partially due to rotation: $\boldsymbol{V} = \boldsymbol{T} - \boldsymbol{\Omega} \times \boldsymbol{Q}$, being $\boldsymbol{T} = -10\,\boldsymbol{k}$ flu · frames^{-1} the translation of the axis of rotation, and $\boldsymbol{\Omega} = 5\,\boldsymbol{j}$ deg · frames^{-1}. The planar surface has an orientation of $\sigma = 45$ deg and $\tau = 60$ deg at each of its points. The reciprocal of the time to collision value (often referred to as *collision immediacy*, frames^{-1}) is plotted as a function of the co-latitude angle ϑ spanning half of the overall visual field (FOV = 160 deg) in the image direction $\varphi = 0$.

Fig. 3 *(a)* reports collision immediacies as defined in eqs. (9), (10) and (13) respectively. It appears that these three quantities are approximately equal only for small values of ϑ, becoming significantly different for FOV of 40 degrees or larger: hence, this is a validity range for the narrow FOV constraint. Fig. 3 *(b)* confirms in fact that scaled depth is well approximated using eq. (23) plus eq. (26) at co-latitudes no larger than 20 deg. Notice that as the narrow FOV fails to be met, the true scalar depth value goes out of the bounds. Conversely, Figs. 3 *(c),(d)* show that *the true values of the central and planar time to collision always remains inside the bounds* obtained in eqs. (47) and (50) respectively. Specifically, there is a value of ϑ at which the spherical deformation vanishes, so that the bounds and the true value are equal: at that value the bounds have a discontinuity, since the derivative of the spherical deformation changes its sign.

Mapping Planar Fields onto the Unit Sphere: Geometric Deformations.   As shown in Section 4.1, the result of Theorem 1 can be used in Theorem 2 to compute time to collision directly from the planar motion field and its invariants as a linear combination of spherical motion field invariants. Such a closed form formulation has a clear computational advantage over the corresponding algorithmic formulation, i.e. first remapping the planar field onto the unit sphere using eq. (35), and then obtaining the spherical invariants by direct differentiation of the spherical field with the analogous of eq. (17). Indeed, apart from being impractical, using the algorithmic formulation could even be impossible, as it happens when only a sparse planar field is available. A delicate point concerning spherical field differentiation is related to
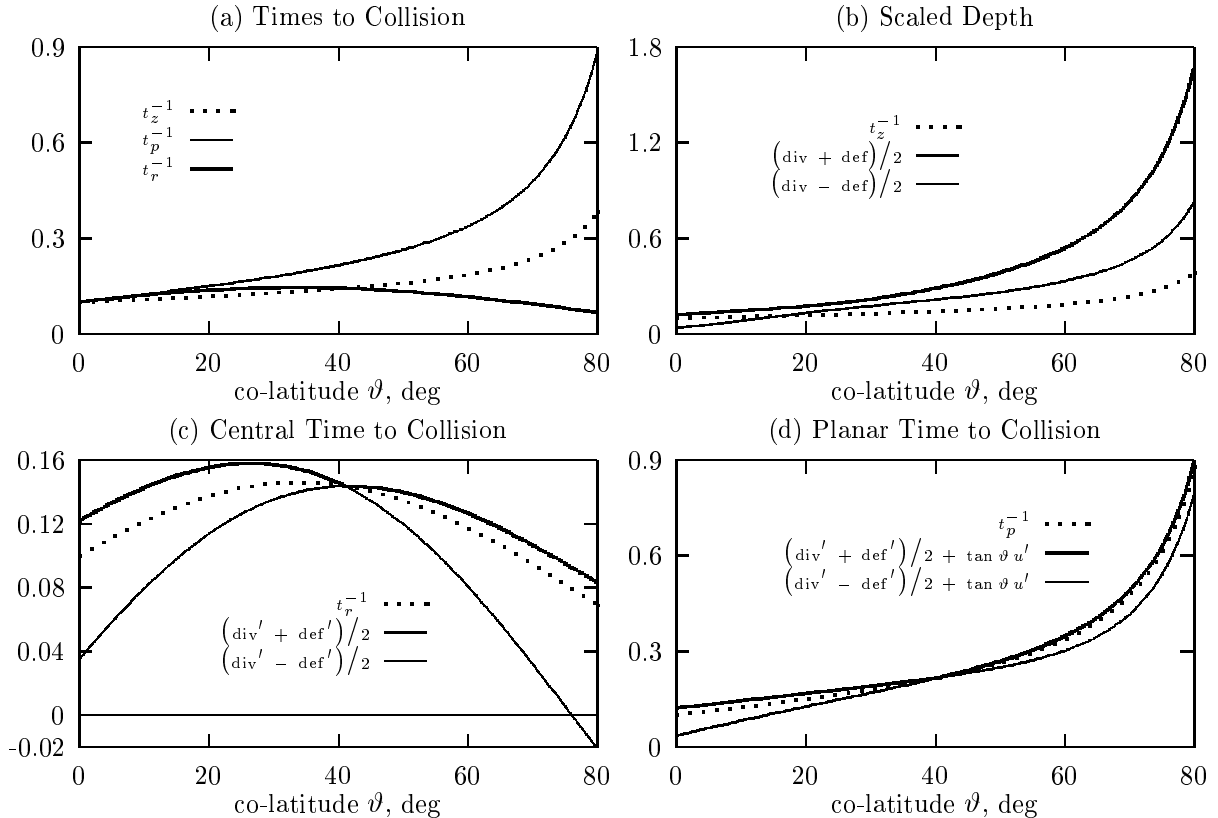
Figure 3: Times to collision and their estimates.

the geometric distortion of the neighborhood of support for the computation of derivatives. In fact, not only the motion field in the neighborhood $\{\boldsymbol{p}_n : |x_n - x| \le \epsilon_x; \ |y_n - y| \le \epsilon_y\}$ of an image point $\boldsymbol{p}$ is deformed when passing from the plane to the sphere, but the neighborhood itself undergoes the deformation governed by eq. (35), since it can obviously be regarded as the particular motion field $[\, x_n - x \ \ y_n - y \,]^{\mathrm{T}}$ with M as the identity matrix. 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.

The planar motion field deformation can be better investigated in terms of eq. (37), describing the mapping which transforms the planar field structure into the spherical one. In the general case, both the eigenvalues and eigenvectors of the matrices M and M$'$ will differ. At a singular point $\boldsymbol{p}$ s.t. $\dot{\boldsymbol{p}} = \boldsymbol{0}$ though, K$'$ vanishes identically, and M and M$'$ *have the same eigenvalues*, since the mapping is a similarity transform. This result is in accordance with the claim that singular points carry topological information about the motion field which is invariant on imaging sensor shape [39]. However, *the similarity mapping does not preserve the eigenvectors of* M, which are nonlinearly deformed due to the multiplication with the nonorthogonal matrix H. Eq. (63) of Appendix A, together with the above observation, provides an insight into the relationship between the geometric deformation of the planar motion field and the way the motion field invariants change after projection onto the sphere. Specifically, a change in the angle β between the eigenvectors has to be expected, as a consequence of the fact that perspective projection does not preserve angles [26]: indeed, it holds

$$\cos \beta = -\frac{\mathrm{curl}}{\mathrm{def}} \quad . \tag{53}$$

15

As clear from eq. (43), at a singular point the divergence — which is equal to the sum of the eigenvalues of M, see Appendix A — is also an invariant of the mapping: this is the reason why it does not appear in eq. (53).
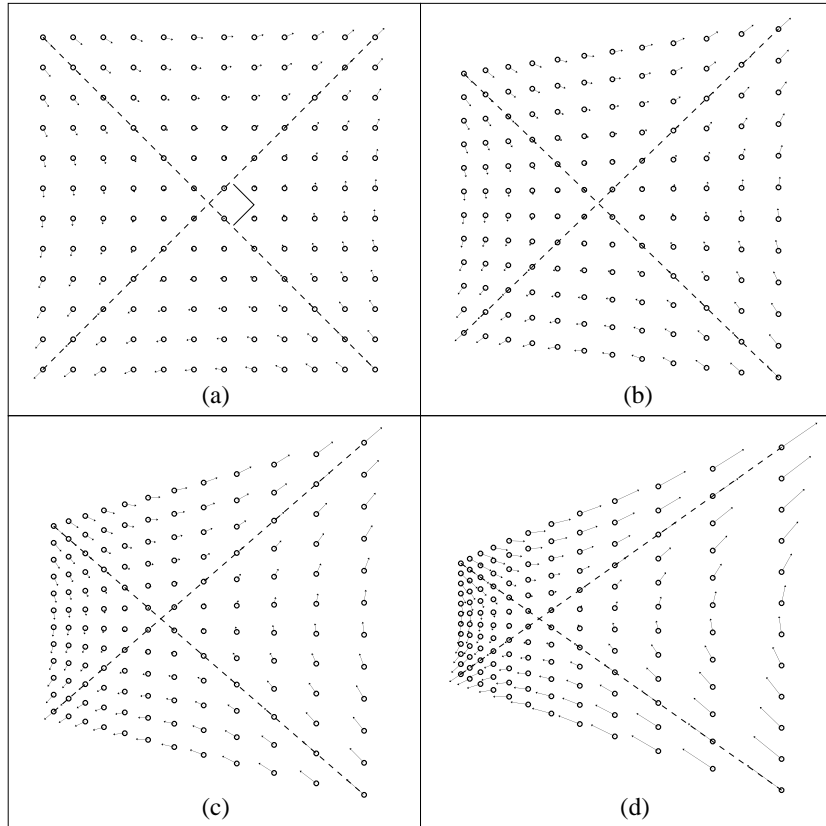

Figure 4: Projection onto a spherical motion field of pure deformation at a singular point. Notice the geometric distortion of both the field and the neighborhood (see text).

Figs. 4 *(a)–(d)* show four planar fields producing, in the neighborhood of singular points located at different ϑ angles (0, 15, 30 and 45 deg respectively) along the $\varphi = 0$ direction the same spherical field of pure deformation. We can think of obtaining these fields by making the deformation field of Fig. 4 *(a)* to "slide" on the spherical surface along the meridian $\varphi = 0$, each time backprojecting it onto the image plane by inverting eq. (35). The spherical field is characterized by a traceless symmetric M′ and orthogonal eigenvectors (the axes of expansion and contraction, respectively), rotated in the sphere tangent plane by an angle $\mu' = \pi/4$. It is evident that all the fields have the same topological characterization (a saddle). However, as ϑ increases, a curl component arises, and the angle between the asymptotic directions (i.e., the eigenvectors) of the planar field decreases. Using eq. (37) together with eq. (53), it is not difficult to show that such an angle depends both on ϑ and μ' as $\cos \beta(\vartheta, \mu') = [n_\vartheta^2/(m_\vartheta^2 + \cot^2 2\mu')]^{1/2}$, where m_ϑ and n_ϑ are defined in eq. (46).

To summarize, there is a strict relationship between the geometric deformation induced by the projection of the planar field onto the sphere and the quantitative changes of the motion field invariants. At singular points, while qualitative field structure information as encoded in the eigenvalues of M is unchanged, all motion field invariants but divergence undergo a quantitative change. Therefore, the local field structure representation in terms of the eigen-

values of M is more stable than the one based on motion field invariants. This encourages the use of eigenvalues rather than invariants to the purpose of doing motion field based image segmentation.

Computational Advantages in the Polar and Log-Polar Planes. A last remark is devoted to the use of polar and log-polar imaging sensors for time to collision estimation. Such sensors are provided with a radial disposition of photosensitive elements, instead than the usual x-y organization of conventional cameras. Specifically, in the log-polar case the ratio between consecutive radii is equal to a constant $a > 1$, so that log-polar pixels grow exponentially in size as the radial coordinate increases from the uniform resolution fovea towards sensor periphery [32]. As a result, while in traditional cameras the tangent of the field of view 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.

It is thus clear that using a log-polar sensor to estimate time to collision proves particularly convenient whenever a wide FOV has to be monitored, and a dense time to collision map has to be computed (these are very common requirements in robot vision). In [37] it is claimed that radial sensors have further specific advantages over traditional cameras, namely, that time to collision can be computed for an arbitrarily large FOV. Yet, as shown in Section 3.2, this is not due to intrinsic advantages of radial sensors (which, basically, offer a new representation of the same image content through a change of coordinates) but to the very strong assumption that the surface be frontoparallel at each imaged point.

However, thanks to their polar organization of image information, both polar and log-polar sensors *do* have a computational advantage over traditional sensors in the wide FOV computation scheme presented in this paper. The advantage is related to the fact that the spherical field can be obtained from the radial representation of the corresponding planar field without the need to rotate by φ the planar coordinate frame:

$$\mathrm{M}' = \begin{bmatrix} \cos\vartheta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\partial\dot\rho}{\partial\rho} & \frac{1}{\rho}\frac{\partial\dot\rho}{\partial\varphi} - \dot\varphi \\ \rho\frac{\partial\dot\varphi}{\partial\rho} + \dot\varphi & \frac{\partial\dot\varphi}{\partial\varphi} + \frac{\dot\rho}{\rho} \end{bmatrix} \begin{bmatrix} \sec\vartheta & 0 \\ 0 & 1 \end{bmatrix} + \mathrm{K}' \quad ; \tag{54}$$

$$\mathrm{M}' = \begin{bmatrix} \cos\vartheta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\partial\dot\xi}{\partial\xi} + \ln a\,\dot\xi & k\ln a\frac{\partial\dot\xi}{\partial\gamma} - \frac{\dot\gamma}{k} \\ \frac{1}{k\ln a}\frac{\partial\dot\gamma}{\partial\xi} + \frac{\dot\gamma}{k} & \frac{\partial\dot\gamma}{\partial\gamma} + \ln a\,\dot\xi \end{bmatrix} \begin{bmatrix} \sec\vartheta & 0 \\ 0 & 1 \end{bmatrix} + \mathrm{K}' \quad . \tag{55}$$

The result above is easily obtained by substituting the polar and log-polar representations of the matrix M derived in Appendix B into eq. (37).

Acknowledgement

References

[1] G. Adiv. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(4):384–401, 1985.

[2] J.K. Aggarwal and N. Nandhakumar. On the computation of motion from sequences of images—a review. *Proceedings of the IEEE*, 76(8):917–935, 1988.

[3] D.H. Ballard and C.M. Brown. *Computer Vision*. Prentice-Hall, Engelwood Cliffs, NJ, 1982.

[4] J.L. Barron, A.D. Jepson, and J.K. Tsotsos. The feasibility of motion and structure from noisy time-varying image velocity information. *International Journal of Computer Vision*, 5(3):239–269, 1990.

[5] 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.

[6] A.R. Bruss and B.K.P. Horn. Passive navigation. *Computer Vision, Graphics, and Image Processing*, pages 3–20, 1983.

[7] 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.

[8] R. Cipolla and A. Blake. Surface orientation and time to contact from image divergence and deformation. In *Proceedings of the 2nd European Conference on Computer Vision ECCV'92*, pages 187–202, 1992.

[9] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 8(3):313–326, 1992.

[10] C. Fermüller and Y. Aloimonos. The role of fixation in visual motion analysis. *International Journal of Computer Vision*, 11(2):165–186, 1993.

[11] M.M. Fleck. Perspective projection: The wrong imaging model. Technical Report TR 95-01, University of Iowa, Computer Science Department, 1994.

[12] 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.

[13] J.J. Gibson. *The Perception of the Visual World*. Houghton-Mifflin, Boston, MA, 1950.

[14] 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.

[15] K. Kanatani. *Group-Theoretical Methods in Image Understanding*. Springer, Berlin, Germany, 1990.

[16] M. Kemp. *The Science of Art. Optical themes in western art from Brunelleschi to Seurat.* Yale University Press, New Haven and London, 1990.

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

[18] 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.

[19] R. Kuc. A spatial sampling criterion for sonar obstacle detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):686–690, 1990.

[20] D.T. Lawton. Processing translational motion sequences. *Computer Vision, Graphics, and Image Processing*, 22:116–144, 1983.

[21] S. Lefschetz. *Differential Equations: Geometric Theory.* Interscience, New York, NY, 1957.

[22] 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.

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

[24] E. De Micheli, S. Uras, and V. Torre. The accuracy of the computation of the optical flow and the recovery of motion parameters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(5):434–447, 1993.

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

[26] J.L. Mundy and A. Zisserman. Projective geometry for machine vision. In J.L. Mundy and A. Zisserman, editors, *Geometric Invariance in Computer Vision*. MIT Press, 1992.

[27] K. Nakayama and J.M. Loomis. Optical velocity patterns, velocity-sensitive neurons, and space perception: a hypothesis. *Perception*, 3:63–80, 1974.

[28] S. Negahdaripour and S. Lee. Motion recovery from image sequences using only first order optical flow information. *International Journal of Computer Vision*, 9(3):163–184, 1992.

[29] R.C. Nelson and J. Aloimonos. Finding motion parameters from spherical motion fields (or the advantages of having eyes in the back of your head). *Biological Cybernetics*, 58:261–273, 1988.

[30] 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.

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

[32] G. Sandini and V. Tagliasco. An anthropomorphic retina-like structure for scene analysis. *Computer Graphics and Image Processing*, 14(3):365–372, 1980.

[33] G. Sandini and M. Tistarelli. Active tracking strategy for monocular depth inference over multiple frames. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):13–27, 1990.

[34] 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.

[35] 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.

[36] M. Tistarelli and G. Sandini. Dynamic aspects in active vision. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 56(1):108–129, 1992.

[37] 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.

[38] A. Verri, F. Girosi, and V. Torre. Mathematical properties of the two-dimentional motion field: from singular points to motion parameters. *Journal of the Optical Society of America A*, 6(5):698–712, 1989.

[39] A. Verri and T. Poggio. Motion field and optical flow: Qualitative properties. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5):490–498, 1989.

[40] K. Wohn and A.M. Waxman. The analytic structure of image flows: Deformation and segmentation. *Computer Vision, Graphics, and Image Processing*, 49:127–151, 1990.

A   Motion Field Invariants: Tensor Calculus and Topological Viewpoints

The properties of a matrix M which remain unaltered after a coordinate system rotation are obtained with the two different viewpoints of tensor calculus and topology.

The tensor calculus viewpoint relies on the Chauchy-Schwartz decomposition theorem, according to which any tensor (matrix) can be split into the sum of three elementary parts: a multiple of the identity matrix, a traceless symmetric and an antisymmetric matrix [15]:

$$
M = \frac{1}{2} \operatorname{div} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{2} \operatorname{curl} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + \frac{1}{2} \operatorname{def} R_\mu^{-1} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} R_\mu \quad, \tag{56}
$$

where R_μ is the orthogonal matrix

$$
R_\mu = \begin{bmatrix} \cos\mu & \sin\mu \\ -\sin\mu & \cos\mu \end{bmatrix} \quad. \tag{57}
$$

The three numbers div, curl and def are the *scalar invariants* of the decomposition, in that they do not change after a rotation of the coordinate axes.

The Jacobian matrix

$$M = \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \tag{58}$$

such that

$$[\, u(x_n, y_n) \;\; v(x_n, y_n) \,]^{\mathrm{T}} \approx [\, u(x, y) \;\; v(x, y) \,]^{\mathrm{T}} + M\,[\, x_n - x \;\; y_n - y \,]^{\mathrm{T}} \tag{59}$$

encodes the first-order structure of a 2D vector field $[\, u(x, y) \;\; v(x, y) \,]^{\mathrm{T}}$ at $[\, x \;\; y \,]^{\mathrm{T}}$. The first-order field approximation is the sum of three elementary field components, a divergence, a curl, and a deformation field, which correspond to simple field configurations. For a pure divergence field, the integral curves — i.e., the curves tangent to the vector field at each point — are straight lines, for a curl field they are circles, and for a deformation field they have two orthogonal asymptotical directions, a direction of expansion and a direction of contraction. The direction of the expansion axis forms an angle μ with the x axis; as such, it does depend on the current reference system. We can regard the deformation component as characterized by a vector $[\, \mathrm{defx} \;\; \mathrm{defy} \,]^{\mathrm{T}} = \mathrm{def}[\, \cos 2\mu \;\; \sin 2\mu \,]^{\mathrm{T}}$, whose magnitude, def, does not depend on the reference frame.

From the topological viewpoint, we qualitatively study the behavior of the field as a whole, based on the matrix eigenvalues. It can be shown that a great number of different planar vector fields can occur, according to the value of the real and the imaginary parts of the eigenvalues. These are referred to as "center," "spiral," "focus," "node," "saddle," and "improper node" [21].

A connection between the two viewpoints of course exists, and relies on the invariant properties of $\mathrm{tr}\,M$ (the trace) and $\det M$ (the determinant). The matrix eigenvalues λ_1 and λ_2 ($\mathrm{Re}\{\lambda_1\} \geq \mathrm{Re}\{\lambda_2\}$) are, indeed, the solutions of the characteristic equation

$$\lambda^2 - \mathrm{tr}\,M\,\lambda + \det M = 0 \quad . \tag{60}$$

Since $\mathrm{tr}\,M = \lambda_1 + \lambda_2$ and $\det M = \lambda_1\,\lambda_2$, we have

$$\lambda_1 + \lambda_2 = \mathrm{div} \;\; ; \quad \lambda_1 - \lambda_2 = \sqrt{\mathrm{def}^2 - \mathrm{curl}^2} \quad . \tag{61}$$

Better still, it is easy to show that if $\lambda_1 - \lambda_2 \neq 0$ and if $\boldsymbol{v}_k$ is the eigenvector relative to the eigenvalue λ_k such that

$$M\,\boldsymbol{v}_k = \lambda_k\,\boldsymbol{v}_k \quad , \tag{62}$$

then

$$\mathrm{div} = \lambda_1 + \lambda_2 \;\; ; \quad \mathrm{curl} = -(\lambda_1 - \lambda_2)\cot\beta \;\; ; \quad \mathrm{def} = (\lambda_1 - \lambda_2)\csc\beta \quad , \tag{63}$$

where the angle β between the two eigenvectors, which is also independent of coordinate rotations, is defined by

$$\cos\beta = \frac{\boldsymbol{v}_1 \cdot \boldsymbol{v}_2^*}{\|\boldsymbol{v}_1\|\,\|\boldsymbol{v}_2\|} \quad , \tag{64}$$

the asterisk denoting complex conjugate. The condition $\lambda_1 - \lambda_2 \neq 0$ ensures that matrix M have exactly two eigenvectors, thus excluding the cases of a focus (M multiple of the identity matrix, infinite eigenvectors) and of an improper node (M not diagonalizable, only one eigenvector).

B Planar Field Invariants using Polar and Log-Polar Coordinates

The scope of this Appendix is to find the expression of the planar motion field differential invariants div, curl, def^φ and $\mathrm{def}^{\varphi+\pi/4}$ after the *polar* $(x, y) \to (\rho = \sqrt{x^2 + y^2}, \varphi = \arctan(y/x))$ and *log-polar* $(\rho, \varphi) \to (\xi = \log_a \rho - h, \gamma = k\,\varphi)$ — a, h and k real positive — coordinate transformations. Let us begin with planar invariants in polar coordinates. We rewrite the Jacobian matrix of eq. (58) as $\mathrm{M} = \partial(u, v)/\partial(x, y)$, and express it using the chain rule for partial derivatives as

$$\mathrm{M} = \frac{\partial(u, v)}{\partial(\rho, \varphi)} \frac{\partial(\rho, \varphi)}{\partial(x, y)} \quad . \tag{65}$$

The two right-hand matrices of eq. (65) evaluate as

$$\frac{\partial(\rho, \varphi)}{\partial(x, y)} = \begin{bmatrix} 1 & 0 \\ 0 & \rho^{-1} \end{bmatrix} \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} \tag{66}$$

and

$$\frac{\partial(u, v)}{\partial(\rho, \varphi)} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}^{-1} \left[\frac{\partial(u^\odot, v^\odot)}{\partial(\rho, \varphi)} + \begin{bmatrix} 0 & -v^\odot \\ 0 & u^\odot \end{bmatrix} \right] \tag{67}$$

respectively, where

$$\begin{bmatrix} u^\odot \\ v^\odot \end{bmatrix} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \dot\rho \\ \rho\,\dot\varphi \end{bmatrix} \tag{68}$$

is the motion field in the polar plane. Since $u_\rho^\odot = \partial\dot\rho/\partial\rho$, $u_\varphi^\odot = \rho\,\partial(\dot\rho/\rho)/\partial\varphi$, $v_\rho^\odot = \rho\,\partial\dot\varphi/\partial\rho + \dot\varphi$ and $v_\varphi^\odot = \rho\,\partial\dot\varphi/\partial\varphi$, we finally obtain

$$\mathrm{M} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial\dot\rho}{\partial\rho} & \frac{1}{\rho}\frac{\partial\dot\rho}{\partial\varphi} - \dot\varphi \\ \rho\frac{\partial\dot\varphi}{\partial\rho} + \dot\varphi & \frac{\partial\dot\varphi}{\partial\varphi} + \frac{\dot\rho}{\rho} \end{bmatrix} \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} \quad . \tag{69}$$

The middle factor in eq. (69) is clearly matrix M after an image coordinate rotation by φ. As mentioned in Appendix A, a coordinate rotation does not change divergence and curl, but only affects the deformation terms. Therefore, recalling the definition of directional deformation of eq. (21), the result for polar coordinates is:

$$\mathrm{div} = \frac{\dot\rho}{\rho} + \frac{\partial\dot\rho}{\partial\rho} + \frac{\partial\dot\varphi}{\partial\varphi} \;\; ; \quad \mathrm{curl} = 2\,\dot\varphi + \rho\frac{\partial\dot\varphi}{\partial\rho} - \frac{1}{\rho}\frac{\partial\dot\rho}{\partial\varphi} \;\; ; \tag{70}$$
$$\mathrm{def}^\varphi = -\frac{\dot\rho}{\rho} + \frac{\partial\dot\rho}{\partial\rho} - \frac{\partial\dot\varphi}{\partial\varphi} \;\; ; \quad \mathrm{def}^{\varphi+\pi/4} = \rho\frac{\partial\dot\varphi}{\partial\rho} + \frac{1}{\rho}\frac{\partial\dot\rho}{\partial\varphi} \quad .$$

A similar development can be used starting from the previous result, thus obtaining

$$\mathrm{M} = \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial\dot\xi}{\partial\xi} + \ln a\,\dot\xi & k\ln a\frac{\partial\dot\xi}{\partial\gamma} - \frac{\dot\gamma}{k} \\ \frac{1}{k\ln a}\frac{\partial\dot\gamma}{\partial\xi} + \frac{\dot\gamma}{k} & \frac{\partial\dot\gamma}{\partial\gamma} + \ln a\,\dot\xi \end{bmatrix} \begin{bmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{bmatrix} \quad , \tag{71}$$

and eventually deriving the log-polar encoding of planar motion field invariants:

$$\mathrm{div} = 2\,\ln a\,\dot\xi + \frac{\partial\dot\xi}{\partial\xi} + \frac{\partial\dot\gamma}{\partial\gamma} \;\; ; \quad \mathrm{curl} = 2\,\frac{\dot\gamma}{k} + \frac{1}{k\ln a}\frac{\partial\dot\gamma}{\partial\xi} - k\ln a\frac{\partial\dot\xi}{\partial\gamma} \;\; ; \tag{72}$$
$$\mathrm{def}^\varphi = \frac{\partial\dot\xi}{\partial\xi} - \frac{\partial\dot\gamma}{\partial\gamma} \;\; ; \quad \mathrm{def}^{\varphi+\pi/4} = \frac{1}{k\ln a}\frac{\partial\dot\gamma}{\partial\xi} + k\ln a\frac{\partial\dot\xi}{\partial\gamma} \quad .$$