Analysis and comparison of two methods for the estimation of 3D motion parameters

Nuno Gonçalves, Helder Araújo

Department of Electrical and Computer Engineering, Institute of Systems and Robotics–Coimbra,
University of Coimbra, Pinhal de Marrocos-Polo II, 3030 Coimbra, Portugal

Abstract

This paper analyzes the problem of motion estimation from a sequence of stereo images. Two methods are considered and their differential and discrete approaches are compared. The differential approaches use differential optical flow whereas the discrete approaches use feature correspondences. Both methods are used to compute, first, the 3D velocity in the direction of the optical axis and, next, the complete rigid motion parameters. The uncertainty propagation models for both methods and approaches are derived. These models are analyzed in order to point out the critical variables for the methods. The methods were extensively tested using synthetic images as well as real images and conclusions are drawn from the results. Real images are used without any illumination control of the scene in order to study the behavior of the methods in strongly noisy environments with low resolution depth maps.
© 2003 Elsevier Science B.V. All rights reserved.

Keywords: 3D motion; Ego-motion; Time-to-impact; Rigid body; Uncertainty propagation

1. Introduction

The estimation of the 3D motion is a problem extensively studied in computer vision. It is highly related to the problem of 3D reconstruction. This dependency can be found in *structure from motion* when the 3D recovery is done with the estimated 3D motion parameters, in *motion from structure* when 3D motion parameters are estimated using 3D structure information and in the case when both motion and structure are estimated together. There are several applications of the 3D motion parameters. Some of them are: 3D reconstruction, surveillance, tracking, obstacles de-

tection and avoidance, object pose estimation, facial and gesture recognition and many other.

This problem can be decomposed into several different sub-problems. First of all we can distinguish the cases where motion is rigid, that is, we consider that all 3D points or regions of 3D points move rigidly with each other. Another problem is the case of non-rigid motion where there are more than one independent motion of points or regions of space. The problem of the rigid motion is often equivalent to the problem of ego-motion.

The case we want to analyze is rigid motion. This problem is known to be a highly unstable regression problem. For the solution of this problem several methods and algorithms have been used. In a first step it is possible to classify them into: discrete methods, differential methods and direct methods. All these classes

E-mail addresses: nunogon@isr.uc.pt (N. Gonçalves),
helder@isr.uc.pt (H. Araújo).

of methods use temporal sequences of images. The former class of methods is designated by discrete because they use a set of features and they assume the correspondences of all features through time. On the other hand, differential and direct methods, also called area-correlation methods, use the induced motion in images, the image velocities (often approximated by the optical flow). The distinction between those classes of methods is that the differential methods use directly the optical flow and the direct methods use the temporal and spatial gradients of scalar fields to estimate the parameters of 3D motion without calculating explicitly the optical flow. Some of those scalar fields are intensity images and/or depth fields.

There are some advantages and disadvantages to each class of methods. There are in each class different problems to solve. Regarding the discrete methods there is the important correspondence problem. The computation of correspondence between features is itself a wide field of research. The most used features are corners, lines and edges. Despite the difficulty in the correspondence problem, discrete methods allow two consecutive images to have a higher displacement from each other than differential and direct methods. When this displacement is very small the discrete methods tend to present serious problems in triangulation. On the other hand, differential and direct methods use the optical flow that, as is well known, can be estimated with reasonable accuracy only when the disparities of the sequence of images is small (not more than some pixels in image). Those methods assume in almost all cases that the brightness is constant with respect to time. They are very sensitive to noise and often numerically unstable.

The majority of the methods presented so far to recover the 3D motion use monocular vision. However, some methods use stereo systems to take advantage of the great amount of information provided by a stereo pair of images. Dynamic stereo systems are systems where the pair of images is taken by the same camera in different moments of time.

We are particularly interested in methods that use stereo systems. In what concerns the discrete methods, Roach and Aggarwal [29] showed that using the perspective projection model two images are sufficient to recover 3D motion of a camera if a set of correspondent points are given. Huang and Blostein [12] applied an iterative technique based on least squares to recover the 3D motion parameters using correspondence of points in two images taken in different moments. Kim and Aggarwal [17] used depth maps to take features correspondences between lines to recover motion parameters. Weng et al. [39] constructed a locally constant angular momentum model to the same objective. Matthies and Shafer [23] used the relation between the 3D structure in two different moments of time. They developed the uncertainty propagation models for the estimation of the 3D structure and motion parameters, from stereo correspondences. Another important work was developed by Young and Chellappa [41] that used a kinematic model to approximate the 3D motion parameters. They also used correspondences between image features. Lee and Kay [19] started from correspondences between a static stereo pair (left and right) and from correspondences in time to recover the pose of objects. Zhang and Faugeras [42,43] use correspondences between 3D lines in space to recover 3D motion parameters (using a Kalman filter). Kanatani [14] and Kanatani and Takeda [15] construct the essential matrix to recover the 3D motion parameters using renormalization. Recently, Demirdjian and Horaud [6] used the projective geometry formalism to split the image points into static points and dynamic points and calculate at the same time the ego-motion parameters. Those parameters are then used to recover all independent objects motions.

Regarding the differential and direct methods, as they are essentially the same, we will mention them indistinctly. Richards [28] proposed in the early 1980s the use of the differential flow (difference between the left and right flow induced by the same 3D point in both images) and the disparity to calculate the 3D motion parameters, using an orthographic projection model. This method was later proposed using a perspective projection model by Waxman and Duncan [36], pointing the importance of the ratio between the derivative of the disparity and the disparity itself to establish stereo correspondences. Waxman and Sinha [37] proposed another method that takes two dynamic stereo images and the image flows to recover not only the 3D motion parameters but also depth. This method works well when the temporal disparity is negligible. Sudhakar et al. [34] and Shieh et al. [32] propose a direct method based on the gradients of the intensity images to estimate the same 3D motion parameters using long sequences of stereo images. Wang and Duncan

[35] extended those concepts to recover motion parameters of multiple objects. Stein and Shashua [33] used another differential method to recover both the 3D motion and 3D structure based on the optical flow. Harville et al. [10] used a brightness constraint and a depth constraint to recover the 3D motion parameters showing that often the depth constraint gives more accurate results than the former since there is no sensitivity to illumination problems. The depth fields were known at starting point. Recently, Molton and Brady [26] presented a method that tries to take all possible information from stereo and motion using several combinations of stereo pairs: left and right cameras at the same time, left camera with right camera in different moments, etc. This method recovers the 3D motion and 3D structure. Other differential and direct works in stereo vision are [2,13,16,20,24,25,38,40].

Motion estimation has been studied mainly within the framework of rigid body motion. However, in robotics literature it is easy to find the motion estimation problem also stated in a different way: the estimation of the time-to-impact (TTI) or time-to-collision.

This quantity yields the time needed to impact with the nearest obstacle if motion remains unchanged. It can be computed with the expression $\text{TTI} = Z/V_Z$ [5], where Z is the depth of the nearest obstacle and V_Z the 3D velocity of the vehicle in the depth (Z) direction. Given the depth information the problem becomes the estimation of V_Z. Notice, however, that the TTI can be calculated without the depth information, using the rate of expansion of the shape of an object [22].

In robotics applications it is very important to avoid the collision with obstacles and TTI performs an important role in that matter. Physiological researchers [4,27] stated that in the human (and animal, in general) visual system the speed of self-motion cannot be determined visually using only the optical flow pattern. TTI, however, can be directly measured from the optical flow. There is, nevertheless, no general agreement if human uses this strategy in avoiding collisions.

Colombo and Del Bimbo [5] point out that often the TTI is confused with scaled depth (which considers only the translational motion). This approximation is reasonable when a narrow field of view is used but at the image periphery gross estimation errors should be expected. To avoid this model error, both translational and rotational components of rigid body motion should be considered.

This paper presents a study on two of the cited methods. One of the methods, that was proposed by Harville et al. [10], a differential method, uses a linear depth change constraint equation (DCCE), that is, assumes a model for the change of the depth fields. If depth measurements are available this method can be applied to a sequence of monocular images, using the temporal and spatial derivatives of the depth. Otherwise a sequence of stereo images can be used to estimate both depth and ego-motion.

The second method analyzed in this paper was proposed by Waxman and Duncan [36], and uses stereo sequences to recover the 3D motion parameters. This method uses the differential image flow between left and right images to compute the motion parameters. Both methods presented in a differential/direct formulation are extended to a discrete approach.

The main goal is to compare the accuracy of those methods to recover two quantities: the total tridimensional velocity in the Z-direction (V_Z) and the full set of 3D motion parameters (ϕ). The estimation of the 3D velocity in the Z-direction is a relevant problem for the computation of time-to-collision [5], which is useful for robotics navigation (although other methods that do not need the computation of Z can be used [22]).

This paper also analyzes those methods within the scope of uncertainty propagation. Matthies and Shafer [23] have also considered similar issues. Errors in the variables used to compute V_Z will inevitably introduce uncertainty in their results. Even very small errors in the optical flow and disparity information can produce a high level of uncertainty in the values of V_Z and ϕ. The aim of this analysis is to quantify the variance of the computed values for V_Z which provides a mean to point out the critical input variables in the methods. Those critical factors indicate which measurements should be carefully done.

The robustness of the methods was also analyzed as a function of the resolution in the depth estimates, in structured indoor scenarios, without a priori knowledge.

In the next section the motion estimation problem is stated. Then the differential and discrete formulations of both methods, for the computation of V_Z and ϕ, are derived in Sections 3 and 4. The uncertainty analysis is presented in Section 5 and in Section 7 the experiments and results obtained are reported. The main conclusions drawn are presented in Section 8.

2. Motion estimation problem

The notations and geometry used throughout this paper shall be first introduced, before the description of the methods used to compute the partial and total 3D velocity.

A 3D point in space is represented by its coordinate vector $\mathbf{P} = [X, Y, Z]^T$ and the world coordinate system is coincident with the cyclopean coordinate system. The cameras (with focal f) are parallel to each other, separated by the baseline b.

Fig. 1 shows the geometry of the stereo vision system and the world coordinate system.

Rigid body motion is used. Let $\mathbf{V}$ be the total 3D velocity of point $\mathbf{P}$. Any rigid body motion can be expressed by a translational component given by $\mathbf{t} = [t_X, t_Y, t_Z]^T$ and a rotational component given by $\mathbf{\Omega} = [\Omega_X, \Omega_Y, \Omega_Z]^T$. The 3D velocity is then $\mathbf{V} = \mathbf{t} + \mathbf{\Omega} \times \mathbf{P}$.

Computing the components of the total 3D velocity $\mathbf{V}$ by expanding its equation, the following expression is obtained:

$$\mathbf{V} = \begin{bmatrix} t_X + \Omega_Y Z - \Omega_Z Y \\ t_Y + \Omega_Z X - \Omega_X Z \\ t_Z + \Omega_X Y - \Omega_Y X \end{bmatrix} = \begin{bmatrix} V_X \\ V_Y \\ V_Z \end{bmatrix} = \begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix}. \tag{1}$$

In the differential approach, the image velocities, induced in the image plane by a 3D point with motion, are given by $v^l = (v^l_x, v^l_y)$ for the left image and by $v^r = (v^r_x, v^r_y)$ for the right image. Using the perspective projection model ($x = f(X/Z)$, $y = f(Y/Z)$) to project the total 3D velocity in the image plane, one obtains

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = f \begin{bmatrix} \dfrac{\mathrm{d}}{\mathrm{d}t}\left(\dfrac{X}{Z}\right) \\ \dfrac{\mathrm{d}}{\mathrm{d}t}\left(\dfrac{Y}{Z}\right) \end{bmatrix} = \begin{bmatrix} f\left(\dfrac{\dot{X}}{Z} - X\dfrac{\dot{Z}}{Z^2}\right) \\ f\left(\dfrac{\dot{Y}}{Z} - Y\dfrac{\dot{Z}}{Z^2}\right) \end{bmatrix}. \tag{2}$$

Substituting Eq. (1) into Eq. (2) the image flow for the cyclopean coordinate system is obtained:

$$v_x = \left\{ f\frac{t_X}{Z} - x\frac{t_Z}{Z} \right\}$$
$$\quad + \left\{ -\frac{xy}{f}\Omega_X + \left(f + \frac{x^2}{f} \right)\Omega_Y - y\Omega_Z \right\},$$
$$v_y = \left\{ f\frac{t_Y}{Z} - y\frac{t_Z}{Z} \right\} \tag{3}$$
$$\quad + \left\{ -\left(f + \frac{y^2}{f} \right)\Omega_X + \frac{xy}{f}\Omega_Y - x\Omega_Z \right\}.$$

To compute the image flow equations for left and right cameras (x_l, y_l) and (x_r, y_r) are used, instead of (x, y) and the motion parameters for each camera. Those are related to the cyclopean system motion parameters by the following expressions:

$$\mathbf{\Omega}_l = \mathbf{\Omega}_r = \mathbf{\Omega},$$
$$\mathbf{t}_l = \mathbf{t} + \mathbf{\Omega} \times \tfrac{1}{2}(b)\hat{\mathbf{i}}, \qquad \mathbf{t}_r = \mathbf{t} - \mathbf{\Omega} \times \tfrac{1}{2}(b)\hat{\mathbf{i}}. \tag{4}$$
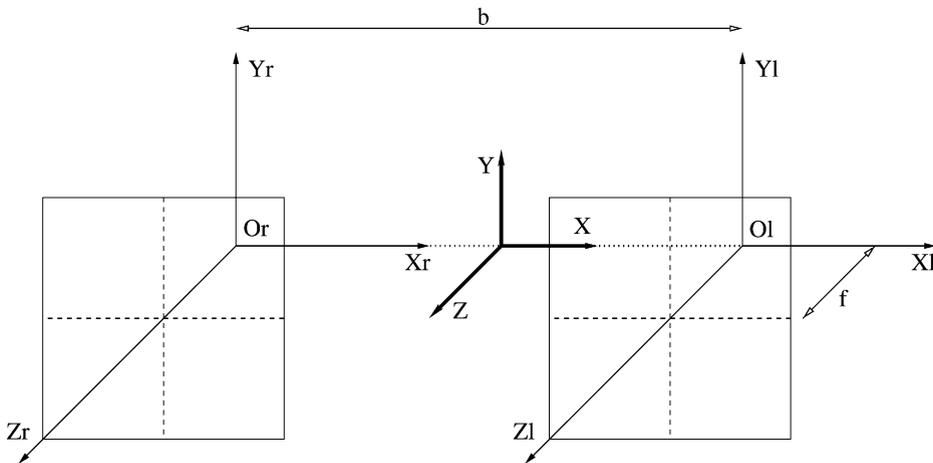

Fig. 1. World and stereo coordinate system.

From the discrete standpoint, however, two points are related in space by a linear transformation composed by a rotation matrix (**R**) and a translation vector (**T**) such that

$$\mathbf{P}' = \mathbf{R} \cdot \mathbf{P} + \mathbf{T}$$
$$\Leftrightarrow \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix},$$
$$(5)$$

where **P** is a 3D point at time t and **P**′ the same point at time t'. In the discrete formulation the velocity of a 3D point is approximated by the finite differences between the point coordinates in time t' and t, that is, $V_Z = Z' - Z$.

In the next two sections, two methods to compute the total 3D velocity along the optical axis (V_Z) as well as the estimation of the translational and rotational components of motion are presented in two different approaches: the differential and the discrete one.

3. Differential approach

In this section 3D motion estimation is considered from a differential standpoint. The correspondences across time are not known and the differential optical flow is available (approximation of image velocities).

The computation of the third component of the total 3D velocity, V_Z, instead of the computation of the total 3D velocity, is important since V_Z is used in the computation of time-to-impact (TTI $= Z/V_Z$). This quantity is very important in robotics applications and in navigation in particular. It is used mainly for obstacle avoidance. Furthermore, it is easy to compute V_Z for each image point with one single equation.

In this section two methods to estimate V_Z are presented. Those methods are later extended for the computation of the rigid motion parameters (ϕ). This paper reviews those methods from previous papers. For further details, see [7,8,10,36].

3.1. Estimation of V_Z—depth constraint

The depth change of a point or rigid body over time is directly related to its velocity in 3D space.

This principle can be used to relate the velocity in the optical axis with depth.

Consider a point $\mathbf{P} = [X, Y, Z]^{\mathrm{T}}$, which projects into a point with coordinates (x, y) in the image plane at a time t and in point $(x + v_x, y + v_y)$ at a time $t + 1$. The depth at instant $t + 1$ should be the depth at the instant t plus the amount of space that the point moved along the optical axis, V_Z. This relationship is given by the following expression, the linear depth change constraint equation—DCCE (see [7,8,10]):

$$Z(x, y, t) + V_Z(x, y, t)$$
$$= Z(x + v_x(x, y, t), y + v_y(x, y, t), t + 1), \quad (6)$$

where $Z(x, y, t)$ is the depth of the point **P** at a given time t and $V_Z(x, y, t)$ the total 3D velocity in the optical axis. $v_x(x, y, t)$ and $v_y(x, y, t)$ are the components of the optical flow.

Approximating Eq. (6) by a first-order Taylor series expansion, the DCCE equation then reduces to

$$V_Z = Z_x v_x + Z_y v_y + Z_t. \quad (7)$$

As mentioned by Harville et al. [10], often motion recovered with depth information is more accurate than that recovered from the intensity images because it is less sensitive to illumination and shading problems.

3.2. Estimation of V_Z—binocular flow constraint

The second method to compute the V_Z is now introduced. This is based on the differences between the flows induced by the movement of a point in a stereo pair of images [7,8,36]. The parallel stereo system is again used and is considered to move rigidly with the scene.

Consider a point $\mathbf{P} = [X, Y, T]^{\mathrm{T}}$ that projects in both image planes as shown in Fig. 2.

Point **P** in Fig. 2, its projection in each image plane $((x_l, y_l)$ and $(x_r, y_r))$ and the optical centers ($\mathbf{O}_l$ and $\mathbf{O}_r$) define two similar triangles, so that we can write the relationship

$$\frac{Z}{b} = \frac{Z - f}{b - (x^r - x^l)}. \quad (8)$$

Computing its temporal derivative and rearranging the terms, one obtains

$$V_Z = -\frac{Z^2}{bf} \Delta v_x, \quad (9)$$
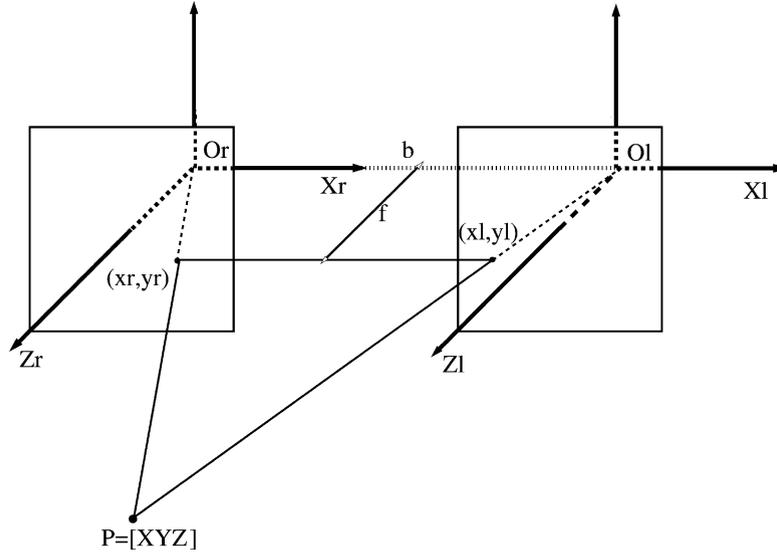
Fig. 2. Binocular triangulation.

which is the binocular flow equation, relating the total 3D velocity in the optical axis (V_Z) and the binocular image flow (DV).

3.3. Motion parameters ϕ—depth constraint

The six motion parameters ($\mathbf{t}$ and $\mathbf{\Omega}$) can also be estimated. The DCCE equation (7) can be written in the following form:

$$-Z_t = [Z_x, Z_y] \begin{bmatrix} v_x \\ v_y \end{bmatrix} - V_Z. \tag{10}$$

Taking the derivatives of the equations of the perspective projection with respect to time, substituting it in Eq. (10) and using Eq. (1) it yields

$$-Z_t = \begin{bmatrix} f\dfrac{Z_x}{Z} \\[2mm] f\dfrac{Z_y}{Z} \\[2mm] -\dfrac{Z + xZ_x + yZ_y}{Z} \\[2mm] -fZ_y - \dfrac{y}{f}(Z + xZ_x + yZ_y) \\[2mm] fZ_x + \dfrac{x}{f}(Z + xZ_x + yZ_y) \\[2mm] xZ_y - yZ_x \end{bmatrix}^{\mathrm{T}} \cdot \phi, \tag{11}$$

where ϕ is the vector with the six motion parameters to be estimated.

All values on the left-hand side of Eq. (11) and in the row vector are known or can be measured (focal length, depth, depth derivatives with respect to time and spatial coordinates and the spatial image coordinates themselves). So there is an equation for each point in the image.

Taking several points (more than six) an overdetermined linear system is obtained. This system can be solved for ϕ with any minimization algorithm.

The system can be written by

$$\mathbf{b}_{DCCE} = \mathbf{H}_{\mathbf{DCCE}}\phi_{DCCE}, \tag{12}$$

where each row is given by Eq. (11) for each point.

3.4. Motion parameters ϕ—binocular flow constraint

The total 3D velocity in the Z-direction (V_Z) can be expressed as a linear equation on three of the six parameters (Eq. (1)) and so it can be substituted in the equation of binocular flow giving

$$t_Z + \Omega_X Y - \Omega_Y X = -\frac{Z^2}{bf}\Delta V_x \tag{13}$$

$$\Leftrightarrow \begin{bmatrix} \dfrac{\Delta V_x}{bf} \end{bmatrix} = \begin{bmatrix} -\dfrac{1}{Z^2}, & -\dfrac{y}{fZ}, & \dfrac{x}{Z} \end{bmatrix} \begin{bmatrix} t_Z \\ \Omega_X \\ \Omega_Y \end{bmatrix}, \tag{14}$$

where 3D point coordinates X and Y were replaced by their inverse perspective projection equations.

Eq. (14) allows the construction of a linear system in the translational velocity along the Z-direction (t_Z) and in the rotational velocities on vertical and horizontal axis (Ω_X and Ω_Y—pan and tilt) using only known variables (focal, differential image flow, depth and image coordinates).

Actually, this is the solution for only half problem since all six parameters should be estimated to completely recover the motion. For now the three parameters are the solution of the linear system given by

$$\mathbf{b}_{DELTAV} = \mathbf{H_{DELTAV}}\phi_{DELTAV} \tag{15}$$

To recover the remaining parameters, the use of the optical flow given by Eq. (3) is proposed. For each image point, t_Z^l, Ω_X^l, Ω_Y^l (equal for all points) and the image flow (v_x^l, v_y^l) are known. Another linear system can be defined to estimate the other three motion parameters.

The six motion parameters are then estimated in a two-step estimation algorithm: first, differential flow and depth field are used to recover t_Z, Ω_X and Ω_Y and then these estimated parameters as well as monocular flow and depth field are used to recover the remaining parameters—t_X, t_Y and Ω_Z.

Due to the two-step nature of the algorithm it is expected that the estimation of the first three parameters is more accurate then the other three parameters since these are estimated with data provided by the first estimation step.

This algorithm can be used to compute the motion parameters of the left, right or cyclopean system which are related by Eq. (4).

4. Discrete approach

In this section the discrete versions of both methods (to compute V_Z) are presented and also a minimization method to compute the motion parameters in the discrete formulation using stereo data.

A discrete number of points is considered, and the transformation between consecutive images should be recovered using the relationship given in Eq. (5).

In the continuous formulations of the DCCE and DV methods it was assumed that the depth information was available and so the disparity in time t and $t' - d$

and d'. Feature correspondences are also available as input of the methods.

4.1. Estimation of V_Z—discrete depth constraint

Instead of expanding the DCCE equation (6) around the point (x, y, t), it is used the point (x, y, t') in the Taylor series expansion, yielding

$$Z(x, y, t) + V_Z = Z(x, y, t') + Z_x(x, y, t')(x' - x) + Z_y(x, y, t')(y' - y). \tag{16}$$

The DCCE equation in the discrete formulation is then given by

$$V_Z = Z_t + Z_x\Delta x + Z_y\Delta y, \tag{17}$$

where $Z_t = Z(x, y, t') - Z(x, y, t)$.

In the discrete formulation of the DCCE equation, the image velocities were replaced by the finite differences of the point image coordinates.

4.2. Estimation of V_Z—discrete binocular flow constraint

For the formulation of the binocular flow constraint equation for the discrete approach consider both the disparities in time t and t':

$$x_r' - x_l' = f\frac{X_r' - X_l'}{Z} = -f\frac{b}{Z'} = d' \tag{18}$$

and

$$x_r - x_l = f\frac{X_r - X_l}{Z} = -f\frac{b}{Z} = d. \tag{19}$$

Calculating the difference between the two disparities and re-grouping the terms, it yields

$$\Delta Z = -\frac{ZZ'}{bf}(d' - d) \tag{20}$$

and then the discrete binocular flow equation is given by

$$V_Z \approx -\frac{ZZ'}{bf}(d' - d). \tag{21}$$

4.3. Motion parameters ϕ—discrete formulation

Considering Eq. (5) and expanding it, a relation between discrete motion parameters ($\mathbf{R}$ and $\mathbf{T}$) and

the 3D points $\mathbf{P}$ and $\mathbf{P}'$ is possible to establish, given by

$$X' - X = (r_{11} - 1)X + r_{12}Y + r_{13}Z + t_1,$$
$$Y' - Y = r_{21}X + (r_{22} - 1)Y + r_{23}Z + t_2, \qquad (22)$$
$$Z' - Z = r_{31}X + r_{32}Y + (r_{33} - 1)Z + t_3.$$

Transforming that equation into an over-determined system, matrix $\mathbf{R}$ and vector $\mathbf{T}$ can be recovered by any linear minimization method. Further constraints can be used in order to enhance the accuracy of the estimation (unit determinant and row–column orthogonality).

The translational and rotational velocities of the motion are expected to be recovered using the rotation matrix and translation vector between the two reference frames.

Let $[\Omega]_x$ be the anti-symmetric matrix of the rotational velocity vector such that it can be written as

$$V(t) = \frac{\mathrm{d}}{\mathrm{d}t}P(t) = \mathbf{t} + [\Omega]_x\mathbf{P}, \qquad (23)$$

which is a set of first-order differential equations in $\mathbf{P}$.

The general solution of Eq. (23) is not straightforward. Often some restrictions are used to facilitate the recovery of the solution. Zhang and Faugeras [43] used the Rodrigues formula to prove that, assuming that translational and rotational velocities are constant, the trajectory of point $\mathbf{P}$ is given by

$$\mathbf{P(t)} = \mathbf{RP_0} + \mathbf{Ut}, \qquad (24)$$

where

$$\mathbf{R} = \mathbf{I_3} + \frac{\sin(\theta\Delta t)}{\theta}[\boldsymbol{\Omega}]_\mathbf{x} + \frac{1 - \cos(\theta\Delta t)}{\theta^2}[\boldsymbol{\Omega}]_\mathbf{x}^2, \qquad (25)$$

$$\mathbf{U} = \mathbf{I_3}\Delta t + \frac{1 - \cos(\theta\Delta t)}{\theta^2}[\boldsymbol{\Omega}]_\mathbf{x}$$
$$+ \frac{\theta\Delta t - \sin(\theta\Delta t)}{\theta^3}[\boldsymbol{\Omega}]_\mathbf{x}^2 \qquad (26)$$

with $\theta = \|\boldsymbol{\Omega}\|$, $\Delta t = t - t_0$, $\mathbf{I_3}$ is the 3×3 identity matrix and $\mathbf{P_0} = \mathbf{P}(t_0)$.

The instantaneous approximation (see [11] for more details) is used to obtain the rotational velocities, given by

$$\mathbf{R} \approx \begin{bmatrix} 1 & -\Omega_Z\Delta t & \Omega_Y\Delta t \\ \Omega_Z\Delta t & 1 & -\Omega_X\Delta t \\ -\Omega_Y\Delta t & \Omega_X\Delta t & 1 \end{bmatrix}. \qquad (27)$$

Adiv [1] stated that this approximation is valid only if two conditions are met. First, the translational velocity in the Z-direction has to be small in relation to the distance of the scene to the reference frame ($t_z\Delta t \ll Z$) and, second, the x- and y-components of the rotation must be small relative to the imaging geometry, that is, the field of view has to be narrow ($\forall x\,|x\Omega_Y\Delta t| \ll f$ and $\forall y\,|y\Omega_X\Delta t| \ll f$). These conditions are similar to those of the weak perspective but they are not the same, i.e., if the weak perspective conditions are verified then these conditions are also verified. The reverse may not be true.

Considering Eq. (24) one concludes that the translational motion parameters are given by

$$\mathbf{t} = \mathbf{U}^{-1} \cdot \mathbf{T}, \qquad (28)$$

where $\mathbf{U}$ is given by Eq. (26) and $\mathbf{T}$ is the result of the linear regression algorithm of Eq. (22).

5. Uncertainty propagation

Given the two models for V_Z, both in the differential and discrete approaches, the uncertainty propagation in the equations due to uncertainty in the data inputs is important to analyze. From this analysis, it is possible to determine the critical independent variables that in presence of uncertainties affect the recovery of motion.

The first step is to define the independent variables for each expression:

$$V_{Z_1}(Z_x, Z_y, Z_t, v_x, v_y) = Z_t + Z_x v_x + Z_y v_y,$$
$$V_{Z_2}(Z, v_x^\mathrm{l}, v_x^\mathrm{r}) = -\frac{Z^2}{bf}\Delta v_x,$$
$$V_{Z_3}(Z_x, Z_y, Z_t, x, y, x', y') = Z_t + Z_x\Delta x + Z_y\Delta y, \qquad (29)$$
$$V_{Z_4}(Z, Z', d, d') = -\frac{ZZ'}{bf}(d' - d),$$

where the geometric parameters were assumed to be known, that is, the baseline and the focal length.

Any noise in the values of the disparity maps, depth data, their temporal and spatial derivatives and in the binocular image flows will affect the computation of V_Z.

To study the uncertainty propagation the covariance matrix of an expression that depends on an input variable vector is computed. Let $\mathbf{F}$ be the function vector to be estimated and $\mathbf{S}$ the vector with the independent variables. Consider $\mathbf{S}$ an n-vector random variable

and $\mathbf{F}$ an m-vector random variable function of the n-vector $\mathbf{S}$. Notice that the relation between $\mathbf{F}$ and $\mathbf{S}$ is in general nonlinear. Considering the mean point of the random variables, and using a first-order approximation, the covariance matrix $\mathbf{\Gamma}$ of the function vector $\mathbf{F}$ [14] can be written as

$$\mathbf{\Gamma} = \frac{\partial \mathbf{F}^{\mathrm{T}}}{\partial \mathbf{S}} \cdot \mathbf{\Lambda} \cdot \frac{\partial \mathbf{F}}{\partial \mathbf{S}}, \tag{30}$$

where $\mathbf{\Lambda}$ is the covariance matrix of the input variables $\mathbf{S}$. $\partial \mathbf{F} / \partial \mathbf{S}$ is the Jacobian matrix that maps vector $\mathbf{S}$ to $\mathbf{F}$.

It is assumed that all variables are affected by Gaussian random white noise with zero mean and standard deviation σ_i, where i denotes the variable. Independent noise in the variables is also assumed so that the covariance matrix for this input signal $\mathbf{S}$ is given by

$$\Lambda_{jk} = \begin{cases} \sigma_{ii}^2 & \text{for } j = k, \\ 0 & \text{for } j \neq k. \end{cases} \tag{31}$$

In this study depth is computed from the disparity with $Z = bf/d$ and so the uncertainty analysis will be within the scope of the disparity and optical flow (differential and/or discrete). The V_Z expressions depend on depth and on the temporal and spatial gradients of depth. So, before analyzing each equation, the uncertainty propagation in depth will be derived:

$$\sigma_{ZZ}^2 = \frac{\partial Z}{\partial d} \cdot \sigma_{dd}^2 \cdot \frac{\partial Z}{\partial d} = \frac{Z^4}{(bf)^2} \cdot \sigma_{dd}^2. \tag{32}$$

For the gradients of depth relative to the variable i ($i \in \{x, y, t\}$), one obtains

$$Z_i = \frac{\partial}{\partial i}\left(\frac{bf}{d}\right) = -\frac{bf}{d^2} \cdot d_i, \tag{33}$$

so that

$$\sigma_{Z_i Z_i}^2 = \left[\frac{\partial Z_i}{\partial d}, \frac{\partial Z_i}{\partial d_i}\right] \cdot \begin{bmatrix} \sigma_{dd}^2 & 0 \\ 0 & \sigma_{d_i d_i}^2 \end{bmatrix} \cdot \begin{bmatrix} \dfrac{\partial Z_i}{\partial d} \\ \dfrac{\partial Z_i}{\partial d_i} \end{bmatrix}. \tag{34}$$

The depth gradients covariance expressions become

$$\sigma_{Z_i Z_i}^2 = \left(\frac{2bf}{d^3} d_i\right)^2 \cdot \sigma_{dd}^2 + \left(\frac{bf}{d^2}\right)^2 \cdot \sigma_{d_i d_i}^2. \tag{35}$$

It is possible now to focus the attention on the expressions of V_Z for both the DCCE and DV methods.

5.1. Depth constraint—differential

For the first expression $\mathbf{F_1} = [V_Z]$ and $\mathbf{S_1} = [Z_x, Z_y, Z_t, v_x, v_y]^{\mathrm{T}}$.

Independent noise in the variables is also assumed, so the covariance matrix for this input signal $\mathbf{S_1}$ is given by

$$\mathbf{\Lambda_1} = \begin{bmatrix} \sigma_{Z_x Z_x}^2 & & \cdots & & 0 \\ & \sigma_{Z_y Z_y}^2 & & & \\ \vdots & & \sigma_{Z_t Z_t}^2 & & \vdots \\ & & & \sigma_{v_x v_x}^2 & \\ 0 & & \cdots & & \sigma_{v_y v_y}^2 \end{bmatrix}. \tag{36}$$

Eq. (30) is used to compute the covariance matrix of the function vector. It yields

$$\mathbf{\Gamma_1} = [v_x, v_y, 1, Z_x, Z_y]\mathbf{\Lambda_1} \begin{bmatrix} v_x \\ v_y \\ 1 \\ Z_x \\ Z_y \end{bmatrix}. \tag{37}$$

The resulting covariance matrix is a scalar matrix given by the expression

$$\mathbf{\Gamma_1} = \sigma_{Z_x Z_x}^2 v_x^2 + \sigma_{Z_y Z_y}^2 v_y^2 + \sigma_{Z_t Z_t}^2 + \sigma_{v_x v_x}^2 Z_x^2 + \sigma_{v_y v_y}^2 Z_y^2 \tag{38}$$

showing the dependencies on the variances of Z_i ($i \in \{x, y, t\}$). Substituting Eqs. (33) and (35) into Eq. (38) it yields

$$\begin{aligned} \mathbf{\Gamma_1} &= \sigma_{V_Z V_Z, 1}^2 \\ &= \left(\frac{2bf}{d^3}\right)^2 (d_x^2 v_x^2 + d_y^2 v_y^2 + d_t^2)\sigma_{dd}^2 \\ &\quad + \left(\frac{bf}{d^2}\right)^2 (v_x^2 \sigma_{d_x d_x}^2 + v_y^2 \sigma_{d_y d_y}^2 + \sigma_{d_t d_t}^2) \\ &\quad + \left(\frac{bf}{d^2}\right)^2 (d_x^2 \sigma_{v_x v_x}^2 + d_y^2 \sigma_{v_y v_y}^2). \end{aligned} \tag{39}$$

5.2. Binocular flow constraint—differential

Using a similar reasoning for the second method the following is obtained:

$$\mathbf{S_2} = \begin{bmatrix} Z \\ v_x^{\mathrm{l}} \\ v_x^{\mathrm{r}} \end{bmatrix}, \tag{40}$$

$$\mathbf{\Lambda_2} = \begin{bmatrix} \sigma_{ZZ}^2 & 0 & 0 \\ 0 & \sigma_{v_x^{\mathrm{l}} v_x^{\mathrm{l}}}^2 & 0 \\ 0 & 0 & \sigma_{v_x^{\mathrm{r}} v_x^{\mathrm{r}}}^2 \end{bmatrix} \tag{41}$$

and the Jacobian matrix is then

$$\frac{\partial \mathbf{F_2}}{\partial \mathbf{S_2}} = \left[-\frac{2Z}{bf} \Delta v_x, \, -\frac{Z^2}{bf}, \, \frac{Z^2}{bf} \right]. \tag{42}$$

The covariance matrix of the function vector, after arranging the terms, is then

$$\mathbf{\Gamma_2} = 4 V_Z \sigma_{ZZ}^2 + \frac{Z^4}{(bf)^2} (\sigma_{v_x^{\mathrm{l}} v_x^{\mathrm{l}}}^2 + \sigma_{v_x^{\mathrm{r}} v_x^{\mathrm{r}}}^2) \tag{43}$$

and substituting Eq. (35) into Eq. (43) it is obtained

$$\begin{aligned} \mathbf{\Gamma_2} &= \sigma_{V_Z V_Z, 2}^2 \\ &= \frac{(2bf)^2}{d^6} (\Delta v_x)^2 \sigma_{dd}^2 + \frac{(bf)^2}{d^4} (\sigma_{v_x^{\mathrm{l}} v_x^{\mathrm{l}}}^2 + \sigma_{v_x^{\mathrm{r}} v_x^{\mathrm{r}}}^2). \end{aligned} \tag{44}$$

5.3. Depth constraint—discrete

In this case the independent variables vector $\mathbf{S_3}$ is given by

$$\mathbf{S_3} = [Z_x, Z_y, Z_t, x, y, x', y']^{\mathrm{T}}. \tag{45}$$

The Jacobian matrix is straightforward in this function. The covariance matrix dependent on the depth is then given by

$$\begin{aligned} \mathbf{\Gamma_3} &= \left(\frac{\partial \mathbf{F_3}}{\partial \mathbf{S_3}} \right)^{\mathrm{T}} \mathbf{\Lambda_3} \left(\frac{\partial \mathbf{F_3}}{\partial \mathbf{S_3}} \right) \\ &= \sigma_{Z_x Z_x}^2 (x' - x)^2 + \sigma_{Z_y Z_y}^2 (y' - y)^2 + \sigma_{Z_t Z_t}^2 \\ &\quad + (\sigma_{xx}^2 + \sigma_{x'x'}^2) Z_x^2 + (\sigma_{yy}^2 + \sigma_{y'y'}^2) Z_y^2 \end{aligned} \tag{46}$$

and substituting Eqs. (33) and (35) into Eq. (46) it is obtained

$$\begin{aligned} \mathbf{\Gamma_3} &= \sigma_{V_Z V_Z, 3}^2 \\ &= \left(\frac{2bf}{d^3} \right)^2 (d_x^2 (\Delta x)^2 + d_y^2 (\Delta y)^2 + d_t^2) \sigma_{dd}^2 \\ &\quad + \left(\frac{bf}{d^2} \right)^2 \cdot [(\Delta x)^2 \cdot \sigma_{d_x d_x}^2 + (\Delta y)^2 \cdot \sigma_{d_y d_y}^2 \\ &\quad + \sigma_{d_t d_t}^2 + d_x^2 (\sigma_{xx}^2 + \sigma_{x'x'}^2) \\ &\quad + d_y^2 (\sigma_{yy}^2 + \sigma_{y'y'}^2)]. \end{aligned} \tag{47}$$

5.4. Binocular flow constraint—discrete

Using the same reasoning, the independent variables vector for the discrete binocular flow method is

$$\mathbf{S_4} = [Z, Z', d, d']^{\mathrm{T}}. \tag{48}$$

Calculating the Jacobian matrix and substituting in the first-order approximation of the covariance matrix of $\mathbf{F_4}$, it yields

$$\begin{aligned} \mathbf{\Gamma_4} &= \left(1 - \frac{d}{d'} \right)^2 \sigma_{ZZ}^2 + \left(\frac{d'}{d} - 1 \right)^2 \sigma_{Z'Z'}^2 \\ &\quad + \frac{Z^2 Z'^2}{(bf)^2} (\sigma_{dd}^2 + \sigma_{d'd'}^2) \end{aligned} \tag{49}$$

and putting together Eqs. (35) and (49) it yields

$$\mathbf{\Gamma_4} = \sigma_{V_Z V_Z, 4}^2 = \left(\frac{bf}{d^2} \right)^2 \cdot \sigma_{dd}^2 + \left(\frac{bf}{d'^2} \right)^2 \cdot \sigma_{d'd'}^2. \tag{50}$$

5.5. Resolution of depth data

The uncertainty due to random noise in the input variables strongly affects the accuracy of the estimation of V_Z. Besides that, the finite resolution of the disparity maps can be one important source of error and affects even more the V_Z estimation accuracy. Fig. 3 shows how the resolution of the disparity can generate uncertainty in the position of a 3D point, mainly in the depth coordinate (since the uncertainty regions are elongate in the depth direction).

The software used to obtain the disparity fields has a finite resolution of 1/16 of pixel. So, some changes in the real depth of a point do not produce any change in the disparity and since depth is inversely proportional
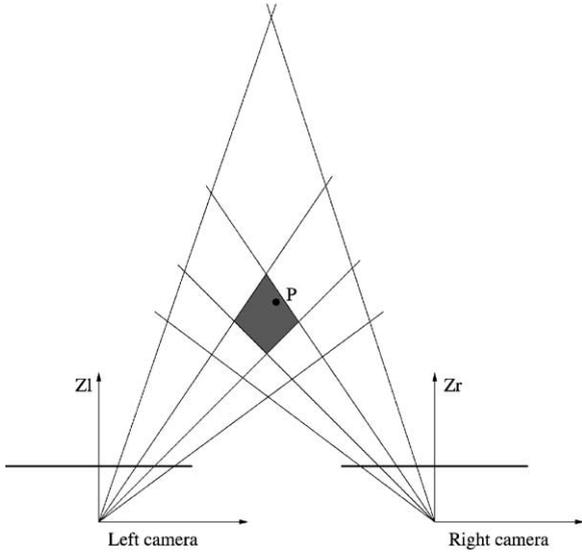
Fig. 3. Effect of finite resolution of disparity maps in depth.

to the disparity its value is calculated with decreasing resolution as the value of the depth itself increases.

Let Δd be the minimum change in disparity. Then for the minimum change in depth to produce change in disparity we have

$$Z = \frac{bf}{d}$$

$$\rightarrow \Delta Z = -\frac{1}{1 + (d/\Delta d)} \cdot \frac{bf}{d}$$

$$= -\frac{1}{1 + (d/\Delta d)} \cdot Z, \tag{51}$$

Eq. (51) indicates that, for close objects, small changes in depth cause significant changes in the disparity and that for distant objects only big changes in depth produce changes in the disparity.

So, consider a realistic situation: $b = 130$, $f = 5$ mm, $\Delta d = 1/16$ px and the pixel width pw $= 0.012$ mm. In that particular case, for example (the disparity is in pixels and the depth is in mm), it is obtained:

- $d = 1 \rightarrow Z = 54\,167 \rightarrow \Delta Z = -3186$ mm;
- $d = 5 \rightarrow Z = 10\,833 \rightarrow \Delta Z = -133.7$ mm;
- $d = 10 \rightarrow Z = 5417 \rightarrow \Delta Z = -33.7$ mm;
- $d = 20 \rightarrow Z = 2708 \rightarrow \Delta Z = -8.4$ mm;
- $d = 50 \rightarrow Z = 1083 \rightarrow \Delta Z = -1.4$ mm.

However, if the resolution lowers to $1/4$ px, for the same case one obtains:

- $d = 1 \rightarrow Z = 54\,167 \rightarrow \Delta Z = -10\,833$ mm;
- $d = 5 \rightarrow Z = 10\,833 \rightarrow \Delta Z = -515.9$ mm;
- $d = 10 \rightarrow Z = 5417 \rightarrow \Delta Z = -132.1$ mm;
- $d = 20 \rightarrow Z = 2708 \rightarrow \Delta Z = -33.4$ mm;
- $d = 50 \rightarrow Z = 1083 \rightarrow \Delta Z = -5.4$ mm.

It can be seen that the low resolution in disparity/depth data can produce large errors with increasing distance to the optical center of the camera. This fact will produce significant errors in the computation of depth field gradients mainly for small motion between two consecutive frames. This also means that it will be difficult to recover motion for distant points.

The perturbation caused by rounding/quantization error (limited resolution) is given by the following equation [31]:

$$\sigma_{dd}^2 = \frac{step^2}{12} \tag{52}$$

where *step* is the minimum increment due to finite resolution.

5.6. Discussion

To analyze quantitatively the uncertainty equations (39), (44), (47) and (50), we constructed a synthetic world (for details, see Section 7.1 and Fig. 4 that shows the left and right images of a synthetic stereo pair and the corresponding disparity map).

This world was projected into two equal cameras mounted in a virtual navigation robot with baseline 130 mm, focal length 5.0 mm, square pixels with side of 0.012 mm. The virtual robot performed several paths (translational, rotational and mixed paths) and the data stored includes: left and right images, disparity in high resolution (map of floats) and continuous and discrete image velocities (in high resolution).

Given the disparity maps, their spatial and temporal gradients and the continuous and discrete velocities, the uncertainty for each point can be computed using the uncertainty equations as a function of the variance of the input variables.

For that purpose the following assumptions are made: the variances of the differential and discrete image velocities are equal in both x- and y-coordinates ($\sigma_{vv}^2 = \sigma_{v_x^l v_x^l}^2 = \sigma_{v_y^l v_y^l}^2 = \sigma_{v_x^r v_x^r}^2 = \sigma_{v_y^r v_y^r}^2$) and the same holds for the discrete velocities and for the gradients of the disparity $\sigma_{d_i d_i}^2 = 0.5\sigma_{dd}^2$ because the deriva-

(a) Left image   (b) Disparity   (c) Right image
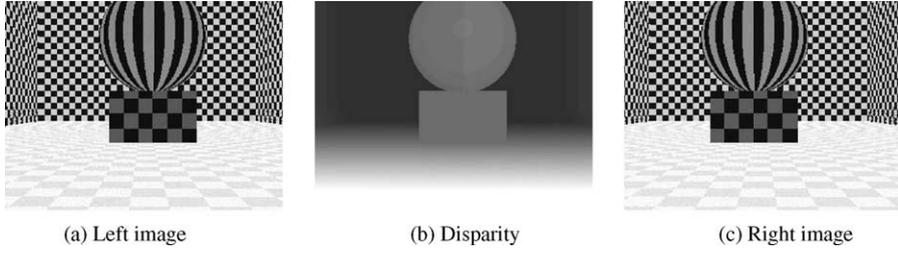
Fig. 4. Intensity images and disparity field for synthetic world.

tives of the disparity maps are approximated by a finite differences equation (for example, $d_t(x, y, t) \approx 0.5d(x, y, t+1) - 0.5d(x, y, t-1)$).

The uncertainty propagation equations are then given by

$$\mathbf{\Gamma_j} = \text{Coef}_{j\,\text{disp}} \cdot \sigma_{dd}^2 + \text{Coef}_{j\,\text{vel}} \cdot \sigma_{vv}^2, \qquad (53)$$

where $j \in \{1, 2, 3, 4\}$ represents one of the methods (DCCE/DIF, DV/DIF, DCCE/DISC and DV/DISC, respectively). $\text{Coef}_{j\,\text{disp}}$ and $\text{Coef}_{j\,\text{vel}}$ are the weights of the disparity and velocities due to random noises, respectively.

From the variance equations of all expressions it is clear that depth is one of the most important factors. Fig. 5 plots the variance value for Eq. (39) where darker points represent values with low variance and lighter points have high variance (saturation for values equal to and above 3000 mm^2). It can be seen that farther objects have higher variances. The corresponding
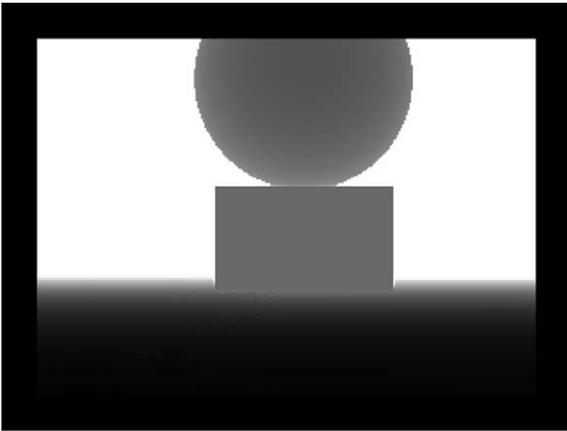

Fig. 5. Variance distribution for $\mathbf{\Gamma_1}$. The quantization error for a resolution of $1/4$ px ($\sigma_{dd}^2 = \sigma_{vv}^2 = 0.0052$) was used.

representations of uncertainty equations $\mathbf{\Gamma_2}$ to $\mathbf{\Gamma_4}$ are not presented since they are similar to the one shown.

To see more explicitly the relation between the uncertainty coefficients and the depth of the points used to compute V_Z, Fig. 6 plots these uncertainty coefficients when a sphere is moved from 2.5 to 5 m with the same motion conditions.

Fig. 6 plots the coefficients of Eq. (53) in two groups: (a) all coefficients that increase highly with the depth and (b) all coefficients that grow much more slowly with the increase of the depth. All coefficients in Fig. 6(a) have very close values in such a way that its distinction is difficult and the same happens with coefficients $\text{Coef}_{1\,\text{vel}}$ and $\text{Coef}_{3\,\text{vel}}$ in Fig. 6(b). Additionally, the scales of Fig. 6(a) and (b) differ three orders of magnitude. Thus it is possible to point out which is the critical factor for each method.

We conclude that, for the DCCE method, both in the differential and in the discrete formulations, the critical factor is the disparity ($\text{Coef}_{1\,\text{disp}} \gg \text{Coef}_{1\,\text{vel}}$ and $\text{Coef}_{3\,\text{disp}} \gg \text{Coef}_{3\,\text{vel}}$). The uncertainty coefficients increase as the uncertainty in the disparity itself increase.

For the DV method, however, the two formulations have distinct behaviors. For the differential one, the critical factor is the uncertainty on velocities ($\text{Coef}_{2\,\text{vel}} \gg \text{Coef}_{2\,\text{disp}}$) and for the discrete one the critical factor is the uncertainty on the disparity ($\text{Coef}_{4\,\text{vel}} = 0$). The former approach presents an increase in the uncertainty coefficients as velocities increase. On the other hand, in the case of the discrete approach, the corresponding coefficients decrease their values.

In the DCCE method as well as in the DV method (in both approaches), the critical factor coefficients were always higher than the other ones (between one and five orders of magnitude).

(a) $Coef_{1disp}$, $Coef_{2vel}$, $Coef_{3disp}$ and $Coef_{4disp}$

(b) $Coef_{1vel}$, $Coef_{2disp}$ and $Coef_{3vel}$
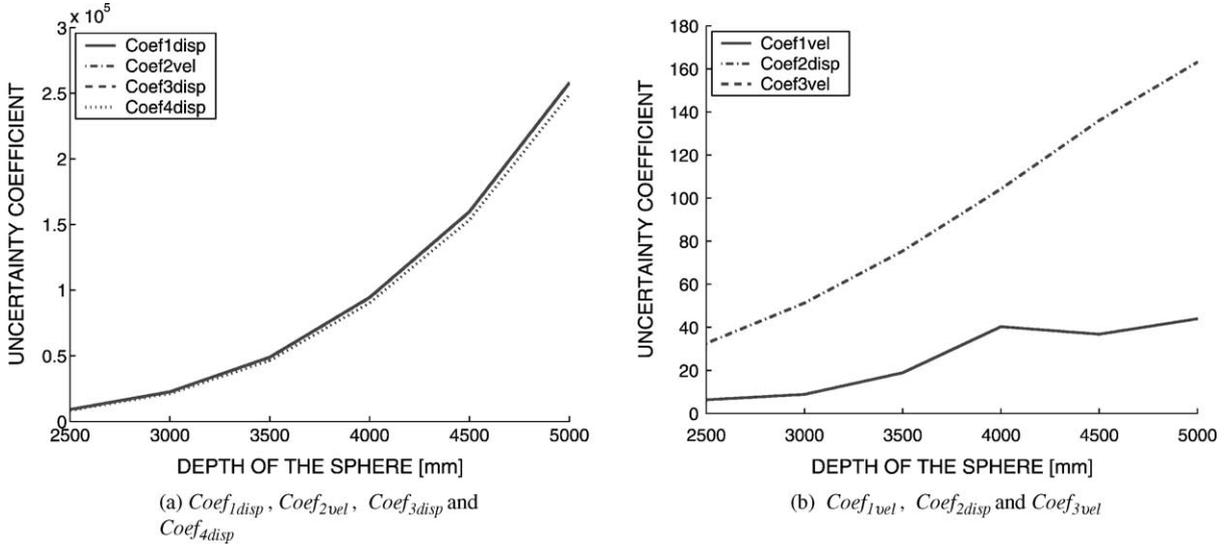
Fig. 6. Uncertainty coefficients of Eq. (53)—depth effect. All coefficients in (a) are almost coincident and they are separated from (b) since they increase much more quickly than those in (b). The indices 1, 2, 3 and 4 refer, respectively, to the methods DCCE/DIFF, DV/DIFF, DCCE/DISC and DV/DISC.

It was also observed that the 3D point depth relative to the cameras is very important to the uncertainty coefficients. Those coefficients increase with a high power (between 2 and 4) of the depth coordinate, so farther objects have much higher uncertainty which suggests that the V_Z is more accurate when closer points are used.

6. Estimation methods

In this section two methods used to estimate motion parameters are presented. Consider the following equation:

$$y_i = x_{i1}\phi_1 + \cdots + x_{ip}\phi_p + e_i \quad \text{for } i = 0, \ldots, n, \tag{54}$$

where n is the total number of observations, the variables $x_{i1}, \ldots, x_{ip}$ are called explanatory variables and y_i is called the response variable. e_i represents the error of the observation and is assumed to have a Gaussian distribution of zero mean, and standard deviation σ. The explanatory and response variables are measured values in several observations. One wants to estimate the unknown parameters vector ϕ.

Let $\hat{\phi}$ be the best estimated parameters vector and $\hat{y}_i$ the estimated response variable calculated using the estimated parameters. It is possible to construct the equation

$$\hat{y}_i = x_{i1}\hat{\phi}_1 + \cdots + x_{ip}\hat{\phi}_p. \tag{55}$$

The main task is to best estimate the parameters vector ϕ using some criterion.

6.1. Least squares

The most used criterion is to minimize the sum of least squares, given by the equation

$$\underset{\hat{\phi}}{\text{Minimize}} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \tag{56}$$

If the problem is in the form of Eq. (12) or (15) ($\mathbf{b} = \mathbf{H}\phi$) and $\mathbf{H}$ is over-determined, it can be shown that mathematically the solution for least squares criterion is given by

$$\hat{\phi} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{b}. \tag{57}$$

This estimator is the most used since its solution is very simple and fast to compute using Eq. (57). However, its main problem is the behavior in the presence

of strong noisy measurements. The estimated solution is the one that minimizes the residuals (square of the difference between the response variable and the estimated response). If there are some observations with very strong deviation (called outliers) the estimated parameters will be truly affected. For a complete study on this estimator and many others, see [30].

6.2. Least median of squares

The breakdown point of an estimator is the smallest fraction of contamination data that can cause the estimator to take on values arbitrarily far from the true value. In other words, the breakdown point of an estimator is the fraction of data that can contain strong error without strongly affecting the estimated parameters. Obviously, the estimator is better if its breakdown point is higher.

The least squares estimator has a 0 breakdown point since one sole bad observation can cause the estimated parameters to be very bad. Thus one looks for an estimator with the highest possible breakdown point (0.5). Rousseeuw and Leroy [30] proposes the least median of squares estimator with a breakdown point of 0.5, given by

$$\underset{\hat{\phi}}{\text{Minimize}}\{\underset{i}{\text{median}}(y_i - \widehat{y_i})^2\}. \tag{58}$$

This estimator also provides a mean to detect the outliers and to reject them. Although its algorithm is very simple, the computation time is much higher than the least squares estimator.

Suppose there are n observations and p explanatory variables. The minimum number of observations to get a vector estimate is p. So, try all combinations of p observations and get the estimated vector $\hat{\phi}$— possibly using Eq. (57). Compute for all observations the estimated response and calculate the median of the obtained squared error (Eq. (58)). The best vector of estimated parameters is the one that minimizes the median of squares.

The total number of trials to compute is the combinations C_n^p which increases very fast with n and p. This is impracticable to do to some values of n. So, Rousseeuw and Leroy [30] propose that in big systems it can be taken a certain number of random sets of p observations and chose the one that minimizes the median of squares. This number m of sets of ob-

servations to try can be calculated using the following equation:

$$P = 1 - (1 - (1 - \epsilon)^p)^m, \tag{59}$$

where P is the probability of at least one of the random sets be "good" and so the estimated vector is "good", that is, unaffected by noise. ϵ represents the maximum fraction of data affected by noise (at most 0.5).

This parameter m is then a trade off between the quality of the solution and velocity of the algorithm. Its value should depend on these two criteria.

As expected, the least median of squares (LMS) estimator, produced, in our experiments, more accurate results than the first presented—least squares (LS).

7. Experiments and results

Two groups of experimental tests were performed. First synthetic images were used to analyze the performance of both methods. Several sensitivity tests were done measuring the accuracy of the estimation of V_Z and ϕ, by changing the resolution of the disparity/depth fields, by adding noise to the disparities and velocities, and by changing the displacement from image to image, in order to identify the critical variables for both methods (in both the differential and discrete approaches).

The second group of experiments was done with real images obtained with known motion parameters.

For both groups of experiments, several paths, including translational, rotational and mixed paths were considered.

For reasons of lack of space only the main results and conclusions will be reported. Extensive testing, results and further details are described in the technical report [9].

7.1. Synthetic sequences

To generate the synthetic images a virtual world was designed. This virtual world consisted of a ground, a front wall, left and right walls and two objects in the middle of the scenario. Using a stereo pair of virtual cameras every point was projected in the image planes. Their disparities and image flows (differential and discrete) were also computed and saved. Fig. 4

Table 1
Motion parameters for the synthetic sequences

Sequence	t_X	t_Y	t_Z	Ω_X	Ω_Y	Ω_Z	V_Z
A	0.0	0.0	−5.0	0.0	0.0	0.0	−5.0
B	0.0	0.0	0.0	0.0	−0.25	0.0	$0.25X$
C	10.0	−10.0	−15.0	0.1	0.1	0.0	$-15.0 + 0.1(Y - X)$

shows an example of the intensity and disparity images obtained.

The stereo system motion is made up of sequences with only translational velocity along the optical axis (Z-direction) and also combined with translational components along the other two axis and sequences with rotational motion over the vertical and horizontal axis (Y- and X-directions—pan and tilt motion). There are also some sequences with combined translational and rotational motion. The motion parameters of the three sequences here reported are listed in Table 1. The true value of V_Z is also presented.

Concerning the methodology used to study the accuracy of the V_Z and ϕ estimation, six sensitivity tests were performed:

- Displacement between two consecutive images, that is, the amplitude of the velocity vector. The parameter STEP (in the set {1, 2, 4, 8, 16, 32}) corresponds to a factor applied to the velocity vector. It is expected to study the influence of the velocities amplitude in the estimation accuracy.
- Disparity resolution. The parameter used, ROUND, represents the number of resolution steps between two consecutive integer values. ROUND is in the set {∞, 32, 16, 8, 4}, where ∞ represents the infinite resolution (since in programming languages the infinite value is not trivial, the ROUND value was substituted by zero, representing infinite resolution).
- Noise added to the disparity data. The parameter STD_DISP is the standard deviation of the random white noise with zero mean added to the disparity data (before the disparity is rounded off). The values used are in the set {0, 1/16, 1/8, 1/4, 1/2}.
- Noise added to the differential image flows. The parameter STD_VEL is the standard deviation of the random white noise with zero mean added to the differential image flow data. The values used

are in the set {0, 1/16, 1/8, 1/4, 1/2}. This study is performed only in the differential approach.
- Noise added to the discrete image flows. The parameter STD_TRACK is the standard deviation of the random white noise with zero mean added to the discrete image flow data (feature tracking). The values used are in the set {0, 1/16, 1/8, 1/4, 1/2}. This study is performed only in the discrete approach.
- Number of features. The parameter FEATURES represents the number of features tracked through the stereo sequence and used in the V_Z and ϕ computation. The values used are in the set $\{2^7, 2^9, 2^{11}, 2^{13}\}$. This study is performed only in the discrete approach.

The accuracy analysis of V_Z is based on the mean value of the relative error computed in all points where V_Z is estimated through the image sequence (several points in several images). This measure, which is called ERM, allows the comparison of both methods. In the estimation of the motion parameters ϕ, however, only one value is computed for each frame (since each point adds a restriction to the over-determined linear system of equations (12) and (15). The values computed for ϕ are valid to all points in the scene. The ERM, in the complete motion parameters estimation, is thus the mean of that value for all frames of the sequence. When the motion parameter is zero, the mean of the absolute error (EAM) is computed instead of the relative error.

The relative mean error (ERM) values obtained for sequence C are plotted in Figs. 7–9, for the V_Z and ϕ estimation. The following acronyms are used—DCCE: depth constraint method; DV: binocular flow method; DIFF: differential approach; DISC: discrete approach.

Fig. 7 shows that the critical factors for each formulation of the method are in accordance with those
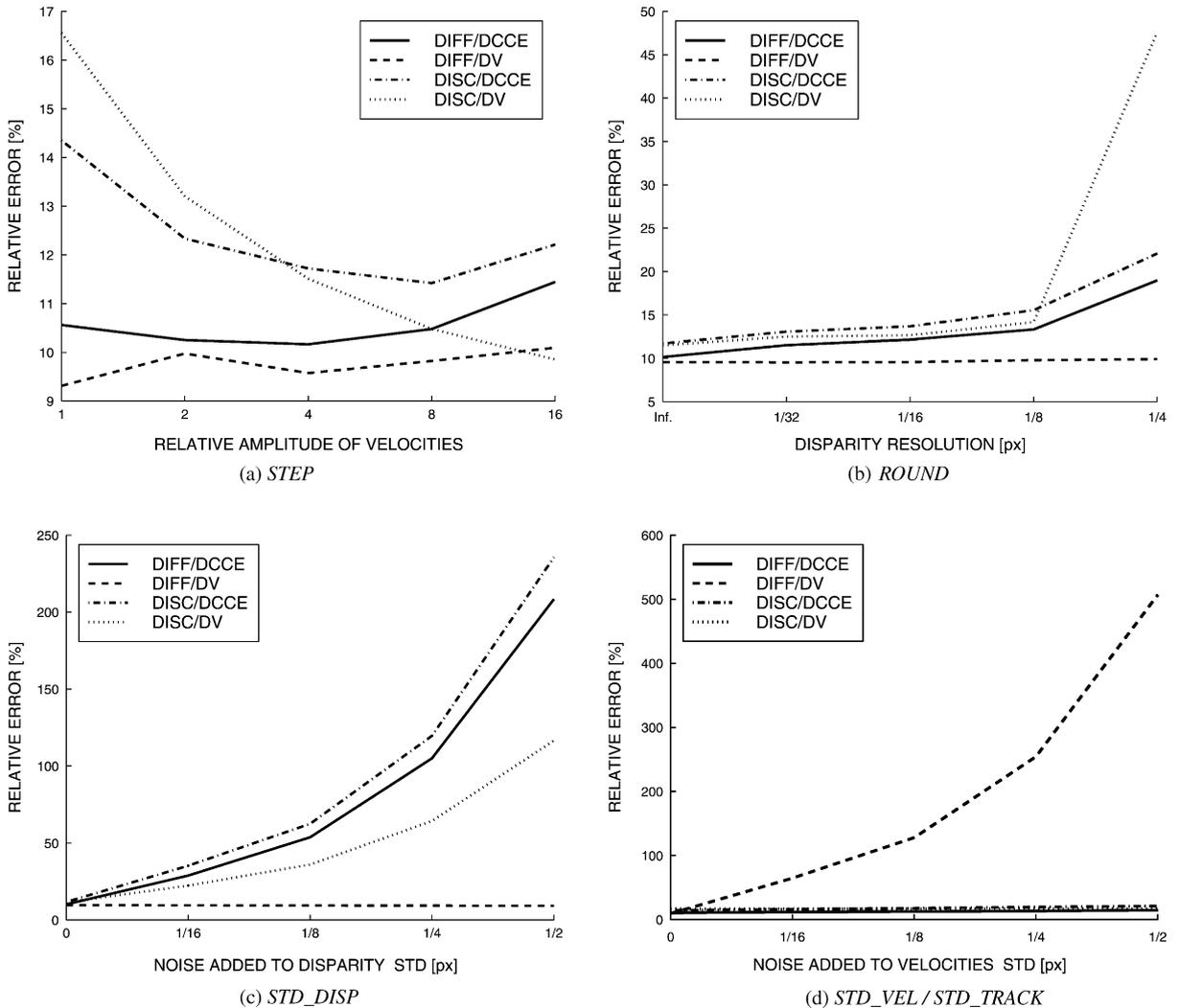
Fig. 7. Relative mean error of V_Z for synthetic sequence C—sensibility analysis. The amplitude of the velocity of the camera is given in number of focal lengths per frame, the disparity resolution is given in pixels (Inf. means floating point resolution) and the standard deviation of the noise added to the disparity and to the velocities is given in pixels. Sequence C motion parameters are given in Table 1.

pointed out by the uncertainty analysis. It can be seen that the estimation error of V_Z tends to increase for slow velocities (one or two focal lengths per frame) and slightly increase for much higher velocities (more than eight focal lengths per frame). For the resolution sensibility test the estimation error increases for all methods as was expected. This error increase occurs mainly for the discrete approach of the DV method. In what concerns to the noise added to the disparity

and to the velocities, the behavior is as was expected: the discrete approach of the DV method is very sensitive to the noise added to the velocities (see Fig. 7(d)) and almost insensitive to the noise added to the disparity although all other methods are sensitive to the noise added to the disparity (see Fig. 7(c)) and almost insensitive to the noise in the velocities.

Figs. 8 and 9 plot the t_Z, Ω_X and Ω_Y estimation error for the amplitude of velocities sensibility
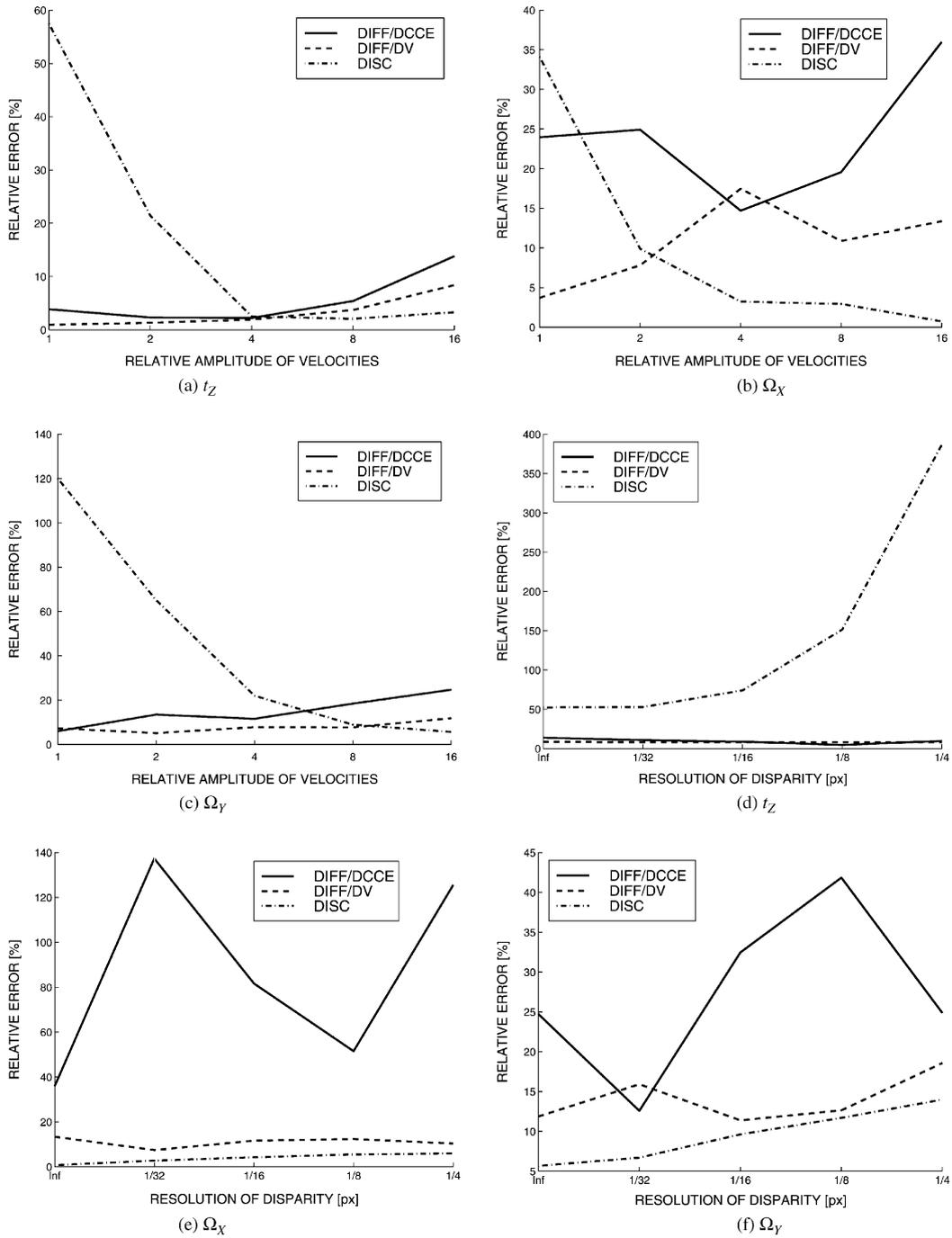
Fig. 8. Relative mean error of ϕ for synthetic sequence C—sensibility analysis in the magnitude of velocities (a)–(c) and disparity resolution (d)–(f). The magnitude of the velocities is given in number of focal lengths per frame and the disparity resolution is given in pixels (Inf. means floating point resolution). Sequence C motion parameters are given in Table 1.
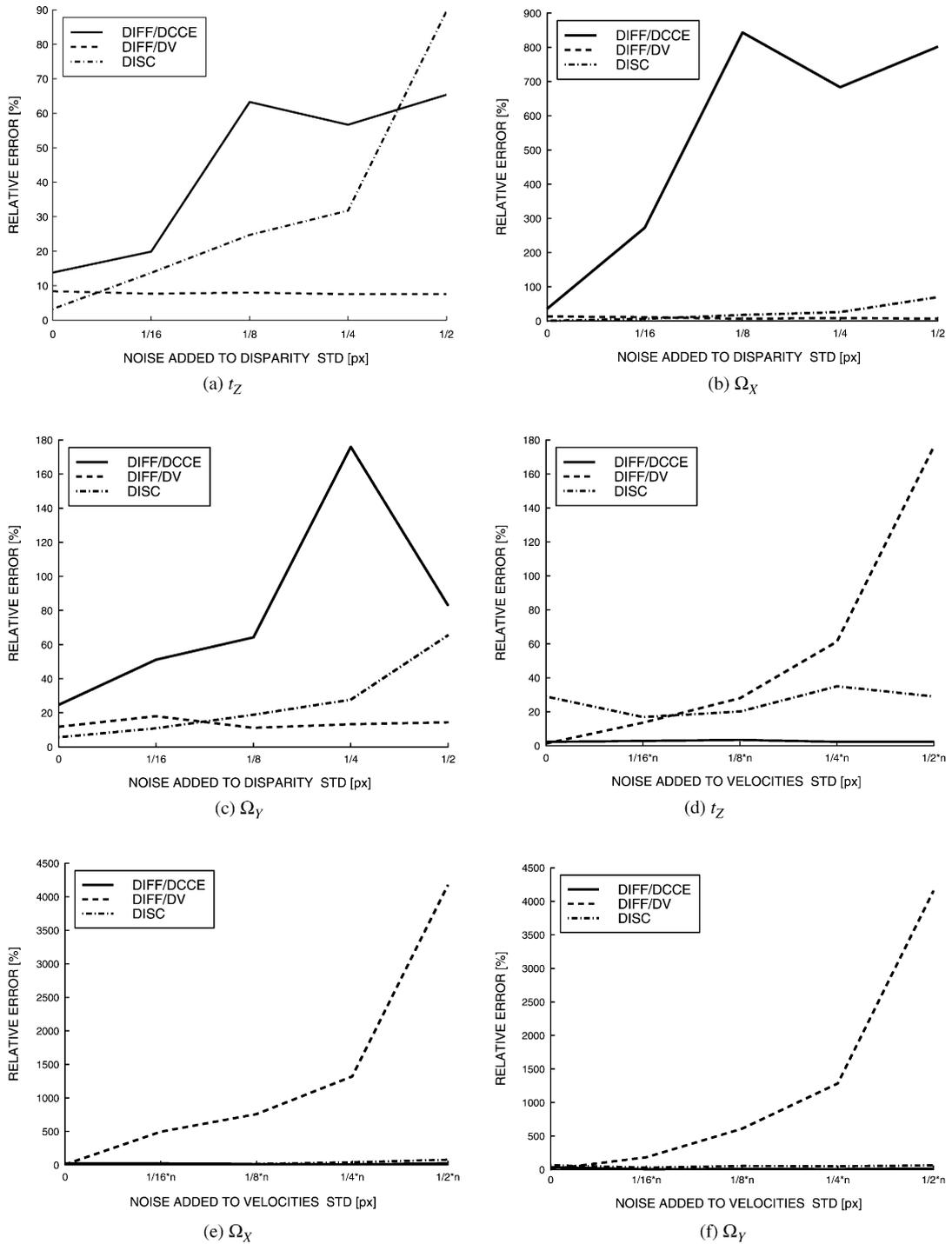
Fig. 9. Relative mean error of ϕ for synthetic sequence C—sensibility analysis in the noise added to the disparity (a)–(c) and velocities (d)–(f). The standard deviation of the noise added to the disparity and to the velocities is given in pixels. Sequence C motion parameters are given in Table 1.
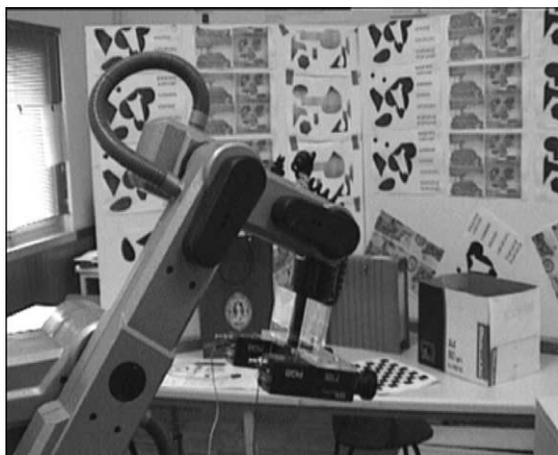
Fig. 10. Stereo head attached to a manipulator.

Table 2
Motion parameters for the real sequences

Sequence	t_X	t_Y	t_Z	Ω_X	Ω_Y	Ω_Z	V_Z
A	0.0	0.0	0.0	0.0	−0.25	0.0	$0.25° \cdot X$
B	0.0	0.0	0.0	0.0	−0.5	0.0	$0.5° \cdot X$
C	0.0	0.0	0.0	0.0	−1.0	0.0	$1° \cdot X$
D	0.0	0.0	0.0	0.0	−2.0	0.0	$2° \cdot X$
E	0.0	0.0	1.0	0.0	0.0	0.0	1.0
F	0.0	0.0	5.0	0.0	0.0	0.0	5.0
G	0.0	0.0	10.0	0.0	0.0	0.0	10.0
H	0.0	0.0	20.0	0.0	0.0	0.0	20.0
I	0.0	0.0	40.0	0.0	0.0	0.0	40.0
J	0.0	5.0	10.0	0.0	0.0	0.0	10.0
L	0.0	10.0	20.0	0.0	0.0	0.0	20.0
M	0.0	5.0	5.0	0.0	−0.25	0.0	$5.0 + 0.25 \cdot X$

analysis (Fig. 8(a)–(c)), resolution sensibility analysis (Fig. 8(d)–(f)), noise in disparity sensibility analysis (Fig. 9(a)–(c)) and noise in the velocities sensibility analysis (Fig. 9(d)–(f)). It can be seen that the error is much smaller in the estimation of t_Z than in the estimation of the rotational velocities Ω_X and Ω_Y and that the same behaviors observed in the estimation of V_Z can be here observed.

7.2. Real sequences

The acquisition of the real stereo image sequences was made using a stereo head mounted on an Eshed manipulator (see Fig. 10) with precision of 0.1 mm in the translational movements and $0.1°$ in the rotational movements. The manipulator permitted five degrees of freedom—translation over all three axes and rotation over the *X*- and *Y*-axes. The manipulator does not allow rotation around the *Z*-axis.

The disparity data was calculated by a commercial algorithm—SVS [18]—which produces disparity maps with resolution of 1/16 px. To compute the optical flow, a well-known algorithm (Lucas–Kanade) was used [3,21]. The scene was structured without light and shadows control. It was composed of two front walls and some objects. Fig. 11 shows a stereo pair of real images as well as their disparity map.

The report of the results in this section is similar to that presented in the previous one. It is based on the ERM and EAM. Furthermore, the standard deviation (STD) of the estimated results distribution is used.

The motion parameters as well as the true value of V_Z for each sequence used is summarized in Table 2.

In the differential approximation, the results obtained for V_Z are reported in Table 3 and the results obtained for ϕ are reported in Tables 4 and 5.

Concerning the estimation of V_Z, it can be concluded that the path is very important in the accuracy obtained. For translational paths it is possible to obtain V_Z with relative errors of about 10–30%, using low resolution disparity maps. In the rotational paths,


(a) Left image   (b) Disparity data   (c) Right image
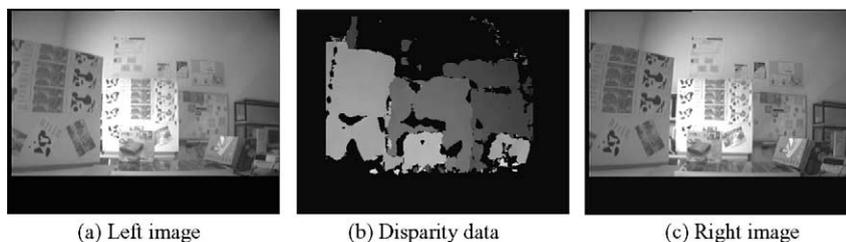
Fig. 11. Intensity images and their disparity map for the real world.

Table 3
Estimation of V_Z—real sequences: differential approach[a]

	Estimated value		True value	ERM (%)		STD (%)	
	DCCE	DV		DCCE	DV	DCCE	DV
A	–	–	–	586.98	3010.7	1898.5	8178.7
B	–	–	–	216.85	1573.8	574.38	5024.4
C	–	–	–	290.18	1698.0	782.49	4990.3
D	–	–	–	230.60	722.29	560.62	2204.5
E	0.556	1.048	1.0	−44.39	4.75	1715.6	714.22
F	5.491	8.022	5.0	9.82	60.45	250.51	211.29
G	6.520	10.391	10.0	−34.80	3.91	182.69	110.53
H	12.763	17.374	20.0	−36.18	−13.13	182.83	85.36
I	40.765	34.918	40.0	1.91	12.71	128.47	102.15
J	10.427	16.043	10.0	4.27	60.43	138.94	153.78
L	21.621	27.251	20.0	8.10	36.26	83.09	116.79
M	3.097	3.549	≈5.0	−38.06	−29.03	4159.1	4413.0

[a] The image sequences A–D are rotational sequences, E–L translational sequences and sequence M both rotational and translational. ERM represent the relative mean error of the estimation of V_Z and STD is the measured standard deviation of the estimation. There is no estimated value and true value in rotational sequences since V_Z varies from point to point. All estimated values and true values are in mm/frame. Error and standard deviation are in percentage. Sequence motion parameters are given in Table 2.

however, the estimation results are very poor. The depth constraint method (DCCE) presents slightly better results than the binocular flow method (DV) in translational and mixed paths and, although poor, much better results in rotational paths.

The V_Z standard deviation is almost always very high. This is a relevant fact since it suggests that when computing V_Z, a high number of measurements may be necessary to allow for a decrease in the standard deviation.

For the computation of the complete motion parameters, ϕ, which is a multi-linear regression problem, there are numerical instability problems for the parameters t_X, t_Y and Ω_Z due to ill-conditioning of the observation matrix. On the other hand, it is possible to estimate with reasonable accuracy the other three parameters (t_Z, Ω_X and Ω_Y). The parameter t_Z, which represents the translational motion in the depth direction is the parameter with best estimation values. The DCCE method is again the best one.

Table 4
Estimation of ϕ—real sequences—rotational sequences: differential approach[a]

		t_X	t_Y	t_Z	Ω_X	Ω_Y	Ω_Z
A	Exact	0.0	0.0	0.0	0.0	−0.25	0.0
	DCCE	249.90	15.01	0.67	0.041	−0.160	0.069
	DV	12.10	2.79	−0.15	0.074	−0.224	−0.038
B	Exact	0.0	0.0	0.0	0.0	−0.50	0.0
	DCCE	258.62	57.05	−1.81	−0.018	−0.124	16.936
	DV	−24.81	−72.52	−5.49	−2.455	0.860	−0.128
C	Exact	0.0	0.0	0.0	0.0	1.00	0.0
	DCCE	−621.67	−99.28	2.86	0.842	0.786	−16.819
	DV	−10.76	−58.70	−4.26	−1.352	0.208	0.531
D	Exact	0.0	0.0	0.0	0.0	2.00	0.0
	DCCE	−222–20	0.0	11.05	0.606	−0.187	−0.781
	DV	−45.91	−36.72	−13.70	−0.889	1.159	0.488

[a] All translational velocities are measured in mm/frame and rotational velocities in degrees/frame. Sequence motion parameters are given in Table 2.

Table 5
Estimation of ϕ—real sequences—translational and mixed sequences: differential approach[a]

		t_X	t_Y	t_Z	Ω_X	Ω_Y	Ω_Z
E	Exact	0.0	0.0	−1.0	0.0	0.0	0.0
	DCCE	−9.66	6.23	−0.71	−0.003	−0.013	2.096
	DV	−4.32	−0.45	−1.77	−0.015	0.083	0.008
F	Exact	0.0	0.0	−5.0	0.0	0.0	0.0
	DCCE	−30.41	21.64	−4.11	−0.018	0.038	0.696
	DV	−4.88	−8.41	−7.12	−0.194	0.126	0.173
G	Exact	0.0	0.0	−10.0	0.0	0.0	0.0
	DCCE	−27.60	1.05	−6.14	−0.006	−0.013	2.275
	DV	−7.41	−1.88	−10.73	−0.085	0.207	0.045
H	Exact	0.0	0.0	−20.0	0.0	0.0	0.0
	DCCE	−99.25	−70.17	−12.78	0.067	−0.071	4.660
	DV	−31.39	10.60	−18.86	0.234	0.874	−0.122
I	Exact	0.0	0.0	−40.0	0.0	0.0	0.0
	DCCE	−118.76	−130.54	−38.49	−0.515	0.020	−0.252
	DV	−110.39	−6.64	−40.00	−0.303	2.344	−0.148
J	Exact	0.0	5.0	−10.0	0.0	0.0	0.0
	DCCE	−42.65	86.22	−10.16	−0.115	0.042	1.574
	DV	−5.78	−28.29	−15.86	−0.475	0.141	0.228
L	Exact	0.0	10.0	−20.0	0.0	0.0	0.0
	DCCE	−243.77	351.56	−19.41	−0.215	−0.029	22.193
	DV	−61.79	−61.04	−28.10	−1.068	1.333	1.030
M	Exact	0.0	5.0	−5.0	0.0	0.25	0.0
	DCCE	−296.22	78.99	−5.94	−0.140	0.130	−8.490
	DV	25.71	110.80	8.35	2.436	0.702	−1.122

[a] All translational velocities are measured in mm/frame and rotational velocities in degrees/frame. Sequence motion parameters are given in Table 2.

Table 6
Estimation of V_Z—real sequences: discrete approach[a]

	Estimated value		Real value	ERM (%)		STD (%)	
	DCCE	DV		DCCE	DV	DCCE	DV
A	–	–	–	775.11	2599.6	2296.3	6407.0
B	–	–	–	376.49	1011.5	1012.6	2550.2
C	–	–	–	396.40	751.63	1012.9	1806.6
D	–	–	–	210.73	514.37	474.67	1207.8
E	0.823	3.131	1.0	−17.68	213.08	1220.5	649.67
F	5.205	6.437	5.0	4.09	28.73	225.14	130.09
G	6.547	6.824	10.0	−34.53	−31.77	156.48	54.29
H	12.991	13.532	20.0	−35.04	−32.34	176.97	64.05
I	43.513	43.034	40.0	8.78	7.58	123.43	37.45
J	11.395	11.121	10.0	13.05	11.21	151.52	83.39
L	24.444	18.578	20.0	22.22	−7.11	80.90	54.99
M	4.910	6.705	≈5.0	−1.81	34.10	348.90	327.04

[a] The image sequences A–D are rotational sequences, E–L translational sequences and sequence M both rotational and translational. ERM represent the relative mean error of the estimation of V_Z and STD the measured standard deviation of the estimation. There is no estimated value and true value in rotational sequences since V_Z varies from point to point. All estimated values and true values are in mm/frame. Error and standard deviation are in percentage. Sequence motion parameters are given in Table 2.

Table 7
Estimation of ϕ—real sequences: discrete approach[a]

		t_X	t_Y	t_Z	Ω_X	Ω_Y	Ω_Z
A	Exact	0.0	0.0	0.0	0.0	−0.25	0.0
	Estimated	1.59	1.52	0.00	0.200	−0.118	0.017
B	Exact	0.0	0.0	0.0	0.0	−0.5	0.0
	Estimated	0.97	−10.75	7.99	−0.138	−0.323	−0.226
C	Exact	0.0	0.0	0.0	0.0	1.0	0.0
	Estimated	−17.66	18.13	233.35	0.607	1.395	0.105
D	Exact	0.0	0.0	0.0	0.0	2.0	0.0
	Estimated	−15.28	35.49	136.00	0.800	2.683	0.461
E	Exact	0.0	0.0	−1.0	0.0	0.0	0.0
	Estimated	−3.39	0.07	4.90	0.185	0.207	0.028
F	Exact	0.0	0.0	−5.0	0.0	0.0	0.0
	Estimated	−3.52	−0.54	5.32	0.169	0.179	0.016
G	Exact	0.0	0.0	−10.0	0.0	0.0	0.0
	Estimated	−2.01	−1.84	−12.18	0.105	0.164	0.022
H	Exact	0.0	0.0	−20.0	0.0	0.0	0.0
	Estimated	3.62	−3.46	−28.53	−0.081	0.107	−0.006
I	Exact	0.0	0.0	−40.0	0.0	0.0	0.0
	Estimated	−5.74	−2.29	−44.91	0.164	0.340	−0.096
J	Exact	0.0	5.0	−10.0	0.0	0.0	0.0
	Estimated	−2.74	−0.19	−12.13	0.093	0.235	−0.006
L	Exact	0.0	10.0	−20.0	0.0	0.0	0.0
	Estimated	−0.15	6.44	−17.83	0.156	0.201	0.067
M	Exact	0.0	5.0	−5.0	0.0	0.25	0.0
	Estimated	−5.42	−3.99	−6.57	0.050	0.319	0.010

[a] All translational velocities are measured in mm/frame and rotational velocities in degrees/frame. Sequence motion parameters are given in Table 2.


(a) Translational path
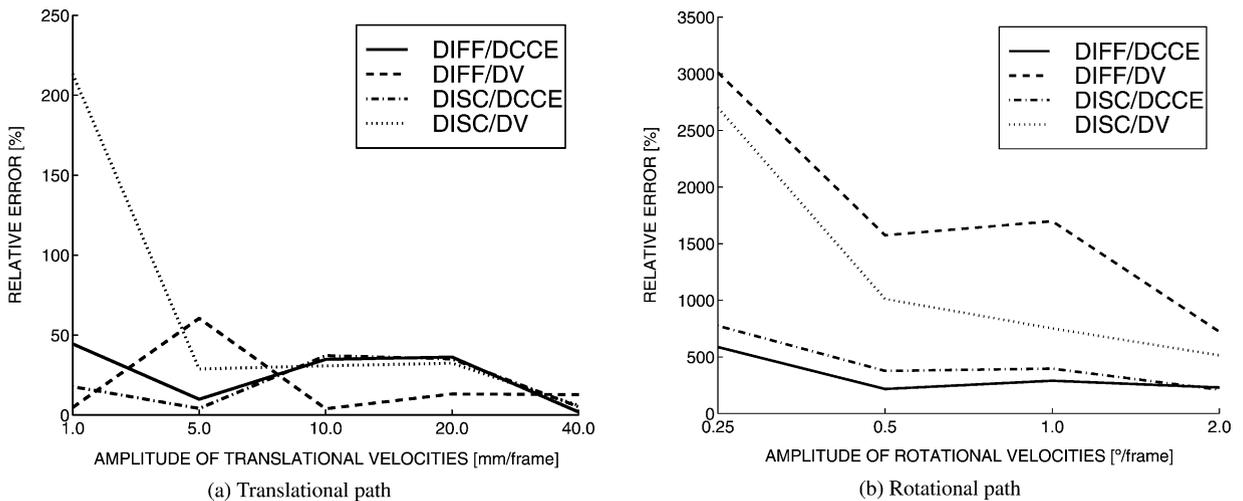
(b) Rotational path

Fig. 12. Relative mean error of V_Z for translational and rotational paths—effect of increasing the amplitude of velocities. Translational velocities are given in mm/frame and rotational velocities in degrees/frame.
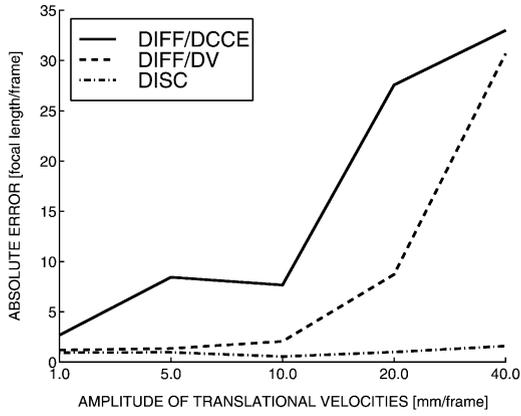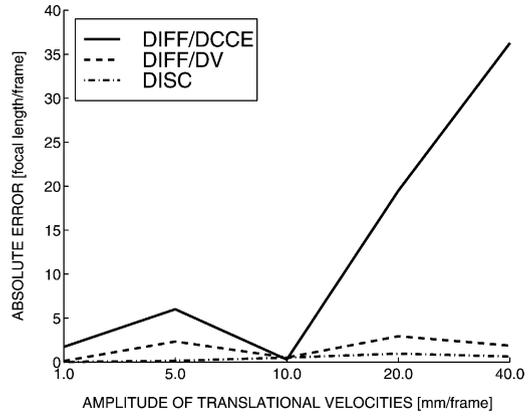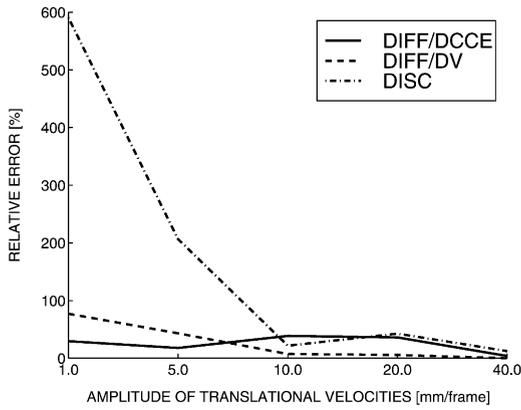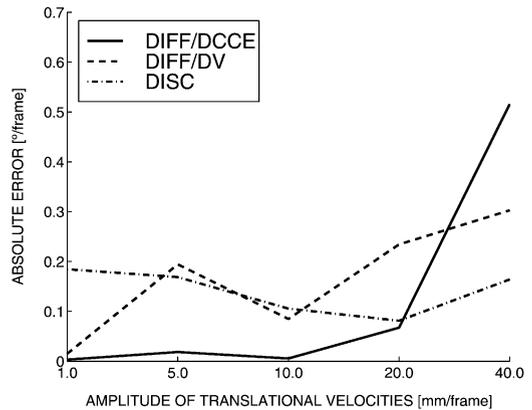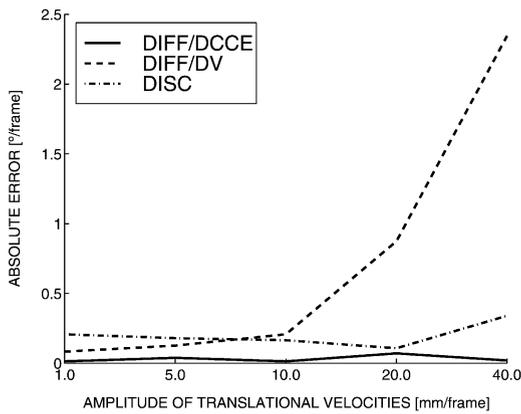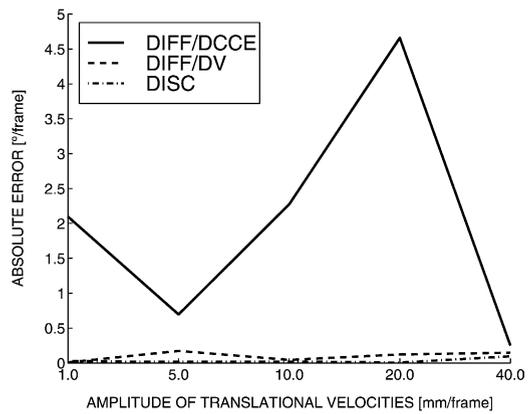
Fig. 13. Relative mean error of ϕ for real images—translational path. Effect of increasing the amplitude of velocities. Velocities in mm/frame.
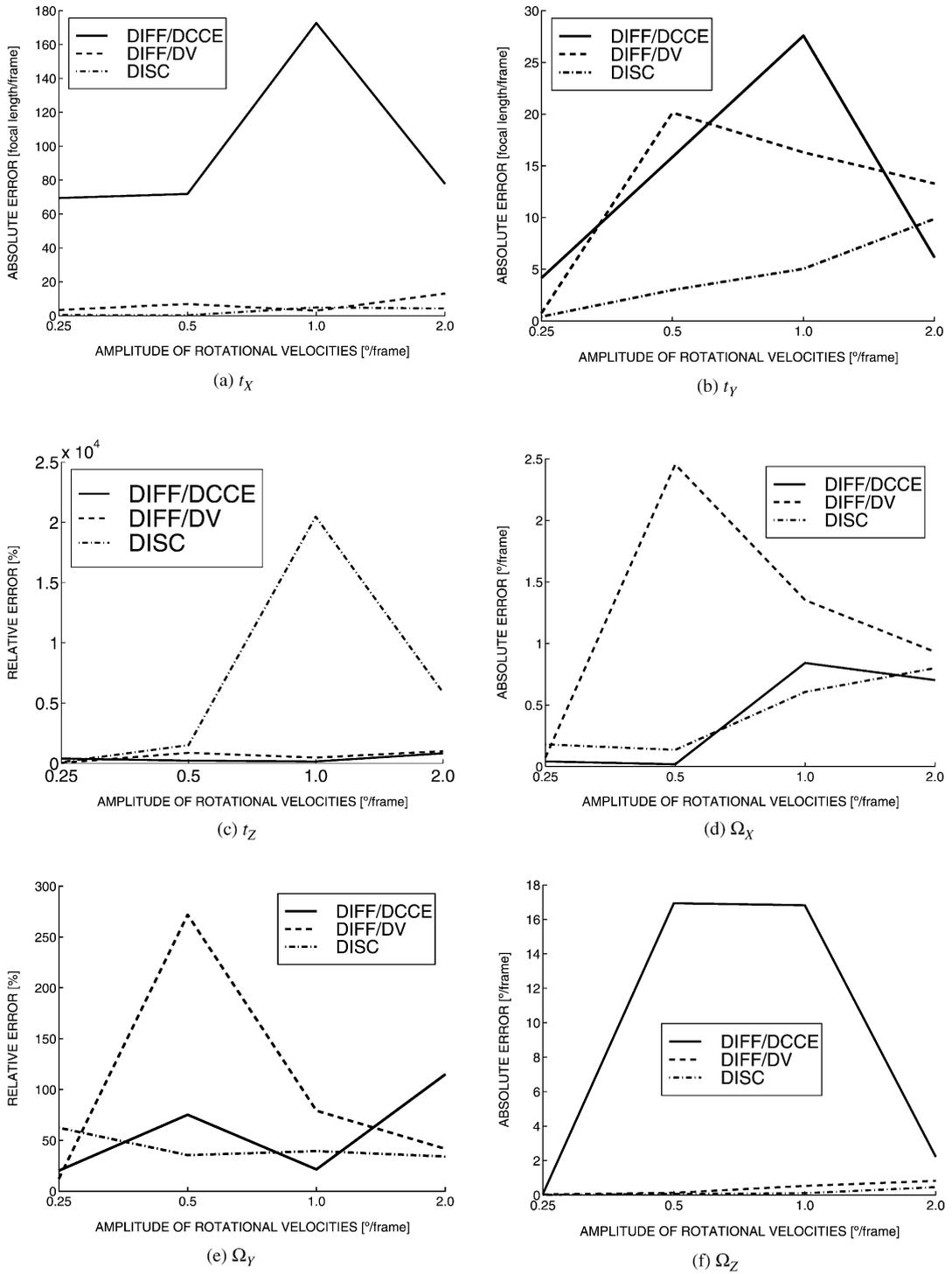
Fig. 14. Relative mean error of ϕ for real images—rotational path. Effect of increasing the amplitude of velocities. Velocities in degrees/frame.

For the discrete approximation, the results obtained to V_Z are reported in Table 6 and the results obtained to ϕ are reported in Table 7.

Regarding the values in these tables one can observe in the estimation of V_Z the same behavior observed in the differential approach. That is, good accuracy (once low resolution disparity is used) for translational sequences and poor accuracy for rotational sequences. In what concerns to the estimation of ϕ, the same instability problem in t_X, t_Y and Ω_Z exists and t_Z, Ω_X and Ω_Y are recovered with reasonable accuracy.

Comparing the differential and discrete methods in the estimation of V_Z, few differences between both methods are noticed. However, in the estimation of ϕ, the discrete approach presents smaller errors in the unstable parameters (and again few differences in the other three).

Additionally, the translational and rotational sequences were grouped to study the effect of increasing the amplitude of velocities. The relative mean error (ERM) values obtained for those sequences are plotted in Figs. 12–14, for the V_Z and ϕ estimation. The following acronyms are used—DCCE: depth constraint method; DV: binocular flow constraint method; DIFF: differential approach; DISC: discrete approach.

These figures plots the values of Tables 3–7. The same observations and conclusions can be drawn. The effect of increasing the amplitude of velocities (translational or rotational) is possible, however, to disclose now. Generally, the V_Z estimation error decreases for greater velocities as well as the t_Z estimation error for translational velocities. For the other parameters, some instability is noticed due to numerical problems.

8. Conclusions

This paper studies the performance of the estimation of 3D motion in the Z-direction and also the estimation of the rigid motion parameters from stereo images. Two formulations were presented: differential and discrete. Two methods to compute both V_Z and ϕ were derived and compared in terms of their accuracy in the estimation.

The uncertainty propagation expressions of the third component of 3D velocity estimation were derived.

Those expressions were written as a function of the uncertainty on the disparity map and the uncertainty on the velocities (differential and discrete).

For the DCCE method, both in the differential and in the discrete formulations, the critical factor is the disparity. There is an increasing tendency of the uncertainty coefficients when the velocities themselves increase.

For the DV method, however, the two formulations have distinct behaviors. For the differential formulation, the critical factor is the uncertainty on velocities and for the discrete one the critical factor is the uncertainty on the disparity.

Furthermore, it was concluded that the depth is a key parameter in the estimation of both V_Z and ϕ. The accuracy is better for closer points. Additionally, it was also concluded that rotational sequences increase the difficulty of the estimation process. The results with real and synthetic images are in agreement with the discussion made in the uncertainty study.

Regarding the experiments and results obtained with synthetic and real images some conclusions can be drawn.

The estimation of V_Z is obtained with good accuracy (once low resolution disparity maps are used) for translational sequences and poor accuracy for rotational ones. The standard deviation of the values recovered is very high suggesting the use of as many points as possible.

In what concerns the estimation of the full motion parameters (ϕ), ill-conditioning of the observation matrix cause very poor accuracy in parameters t_X, t_Y and Ω_Z although good values are recovered for the remaining parameters, mainly for t_Z, the best estimated.

Concerning the comparison of the differential and discrete methods, it can be concluded that in the estimation of V_Z and ϕ, there are few differences between both methods. However, in the estimation of ϕ, the discrete approach presents slightly better results than the differential one. The discrete formulation gets better results for higher displacements.

When using both methods in real environments one concludes that the critical factors are the accuracies in the disparities and in the velocities. In our tests, the best method was, in general, the DCCE. Depending on the type of the most accurate values (disparity or velocities) the best one should be the DCCE or the DV, respectively.

Acknowledgements

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 PAMI-7 (4) (1985) 384–401.

[2] J.L. Barron, R. Eagleson, Motion and structure from long binocular sequences with observer rotation, in: Proceedings of the International Conference on Image Processing, October 1995, pp. 193–196.

[3] J.L. Barron, D.J. Fleet, S.S. Beauchemin, Performance of optical flow techniques, IEEE International Journal of Computer Vision 12 (1) (1994) 43–77.

[4] A. Berthoz, The Brain's Sense of Movement, Harvard University Press, Cambridge, MA, 1997.

[5] C. Colombo, A. Del Bimbo, Generalized bounds for time to collision from first-order image motion, in: Proceedings of the Seventh IEEE International Conference on Computer Vision, Corfu, Greece, September 1999, pp. 220–226.

[6] D. Demirdjian, R. Horaud, Motion–egomotion discrimination and motion segmentation from image-pair streams, Computer Vision and Image Understanding 78 (2000) 53–68.

[7] N. Gonçalves, H. Araújo, Estimation of 3D motion from stereo images—differential and discrete formulations, in: Proceedings of the ICPR'2002, Quebec City, Quebec, IEEE Press, New York, August 2000.

[8] N. Gonçalves, H. Araújo, Estimation of 3D motion from stereo images—uncertainty analysis and experimental results, in: Proceedings of IROS'02, Lausanne, Switzerland, IEEE, New York, October 2002.

[9] N. Gonçalves, Estimação de movimento usando estereovisão, Master's Thesis, Department of Electrical and Computer Engineering Faculty of Science and Technology, University of Coimbra, June 2002.

[10] M. Harville, A. Rahimi, T. Darrell, G. Gordon, J. Woodfill, 3D pose tracking with linear depth and brightness constraints, in: Proceedings of the IEEE International Conference on Computer Vision, Corfu, Greece, 1999.

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

[12] T.S. Huang, S.D. Blostein, Robust algorithm for motion estimation based on two sequential stereo image pairs, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, 1985, pp. 518–523.

[13] A.D. Jepson, J.L. Barron, J.K. Tsotsos, Determining egomotion and environmental layout from noisy time-varying image velocity in binocular image sequences, in: Proceedings of IJCAI-1987, Milan, Italy, vol. 2, 1987, pp. 822–825.

[14] K. Kanatani, Statistical optimization for geometric computation: theory and practice, in: Machine Intelligence and Pattern Recognition, vol. 18, North-Holland, Elsevier, 1996.

[15] K. Kanatani, S. Takeda, 3D motion analysis of a planar surface by renormalization, IEICE Transactions on Information and Systems E78-D8 (1995) 1074–1079.

[16] P.J. Kellman, M.K. Kaiser, Extracting object motion during observer motion: Combining constraints from optic flow and binocular disparity, Journal of the Optical Society of America 12 (3) (1995) 623–625.

[17] Y.C. Kim, J.K. Aggarwal, Determining object motion in a sequence of stereo images, IEEE Journal on Robotics and Automation 3 (1987) 599–614.

[18] K. Konolige, Small Vision System—Development System. http://www.ai.sri.com/~konolige/svs/svs.htm.

[19] S. Lee, Y. Kay, A Kalman filter approach for accurate 3D motion estimation from a sequence of stereo images, CVGIP: Image Understanding 54 (1991) 244–258.

[20] L. Li, J.H. Duncan, 3D translational motion and structure from binocular image flows, IEEE Transactions on Pattern Analysis and Machine Intelligence 15 (7) (1993) 657–667.

[21] B. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: Proceedings of the DARPA Image Understanding Workshop, 1981, pp. 121–130.

[22] E. Martínez, C. Torras, Qualitative vision for the guidance of legged robots in unstructured environments, Pattern Recognition 34 (8) (2001) 1585–1599.

[23] S.A. Matthies, L. Shafer, Error modeling in stereo navigation, IEEE Journal on Robotics and Automation 3 (3) (1987) 239–248.

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

[25] A. Mitiche, On kineopsis and computation of structure and motion, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (1) (1986) 109–112.

[26] N. Molton, M. Brady, Practical structure and motion from stereo when motion is unconstrained, International Journal of Computer Vision 39 (1) (2000) 5–23.

[27] S. Palmer, Vision Science: Photons to Phenomenology, MIT Press, Cambridge, MA, 1999.

[28] W. Richards, Structure from stereo and motion, Journal of the Optical Society of America 2 (1985) 343–349.

[29] J.W. Roach, J.K. Aggarwal, Determining the movement of objects from a sequence of images, IEEE Transactions on Pattern Analysis and Machine Intelligence 2 (6) (1980) 544–562.

[30] P.J. Rousseeuw, A.M. Leroy, Robust Regression and Outlier Detection, Wiley, New York, 1987.

[31] K. Shanmugan, Digital and Analog Communication Systems, Wiley, New York, 1979.

[32] J.-Y. Shieh, H. Huang, R. Sudhakar, Motion estimation from a sequence of stereo images: A direct method, IEEE Transactions on Systems, Man and Cybernetics 24 (7) (1994) 1044–1053.

[33] G. Stein, A. Shashua, Direct estimation of motion and extended scene structure from a moving stereo rig, in: Proceedings of the Conferene on Computer Vision and Pattern Recognition, CVPR'98, Santa Barbara, CA, June 1998, pp. 211–218.

[34] R. Sudhakar, H. Zhuang, P. Haliyur, J. Shieh, Estimation of motion parameters using binocular camera configurations, in: Proceedings of the SPIE Conference on Intelligent Robots, November 1991, pp. 24–32.

[35] W. Wang, J.H. Duncan, Recovering the three-dimensional motion and structure of multiple moving objects from binocular image flows, Computer Vision and Image Understanding 63 (1996) 430–446.

[36] A.M. Waxman, J.H. Duncan, Binocular image flows: steps towards stereo-motion fusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (6) (1986) 715–729.

[37] A.M. Waxman, S. Sinha, Dynamic stereo: passive ranging to moving objects from relative image flows, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (1986) 406–412.

[38] J. Weng, P. Cohen, N. Rebibo, Motion and structure estimation from stereo image sequences, IEEE Transactions on Robotics and Automation 8 (3) (1992) 362–382.

[39] J. Weng, T.S. Huang, N. Ahuja, 3D motion estimation, understanding, and prediction from noisy image sequences, IEEE Transactions on Pattern Analysis and Machine Intelligence 9 (1987) 370–389.

[40] G. Xu, S. Tsuji, M. Asada, A motion stereo method based on coarse to fine control strategy, IEEE Transactions on Pattern Analysis and Machine Intelligence 9 (2) (1987) 332–336.

[41] G.J. Young, R. Chellappa, 3D motion estimation of using a sequence of noisy stereo images: Models, estimation and uniqueness results, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1990) 735–759.

[42] Z. Zhang, O. Faugeras, Estimation of displacements from two 3D frames obtained from stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (12) (1992) 1141–1155.

[43] Z. Zhang, O. Faugeras, Three-dimensional motion computation and object segmentation in a long sequence of stereo frames, International Journal of Computer Vision 7 (3) (1992) 211–241.

Nuno Gonçalves is a Teaching Assistant in the Department of Electrical and Computer Engineering, University of Coimbra, Portugal, and Researcher in the Portuguese Institute of Systems and Robotics. He received his B.Sc. degree in Electrical Engineering and his MS degree in Electrical and Computer Engineering, Automation and Control, Computer Vision, from the University of Coimbra, in 1999 and 2002, respectively. He is currently working toward the Ph.D. degree in Computer Vision. His research interests are computer vision, robotics, omnidirectional vision and tracking.

Helder Araújo is an Associate Professor in the Department of Electrical and Computer Engineering, University of Coimbra, Coimbra, Portugal. He is co-founder of the Portuguese Institute for Systems and Robotics (ISR), where he is now a Researcher. His primary research interests are in computer vision and mobile robotics.