Estimation of 3D Motion from Stereo Images - Differential and Discrete Formulations

Nuno Gonçalves and Helder Araújo

Institute of Systems and Robotics
Dept. of Electrical and Computers Engineering
University of Coimbra
Pinhal de Marrocos - POLO II - 3030 Coimbra
PORTUGAL

Abstract

In this paper we analyze the problem of motion estimation from a sequence of stereo images. We formulate both the differential and discrete approaches of two methods. The differential approach uses differential optical ¤ow whereas the discrete approaches uses feature correspondences. We use both methods to compute, £rst, the 3D velocity in the Z direction and, second, the complete motion parameters. The methods were extensively tested using synthetic images as well as real images and several conclusions are drawn from the results. We point out the critical factors for the methods. The real images are used without any illumination control of the scene in order to analyze the behavior of the methods in strongly noisy environments and low resolution depth maps.

1. Introduction

All the methods that solve the motion problem can be classi£ed into discrete methods and differential or continuous methods. Both classes of methods use temporal sequences of images. The former class is called discrete because it uses a set of features and the features correspondences across time are assumed to be known. On the other hand, differential methods use differential optical ¤ow.

In this paper we will present a performance analysis of two methods [4, 6]. One of our goals is the comparison of those methods in terms of their accuracy in the estimation of the 3D velocity in the Z direction. Another goal is their comparison in terms of the estimation of the 3D motion parameters. The estimation of 3D velocity in the Z direction is a relevant problem for the computation of time-to-collision, which is very useful for robot navigation.

We also developed the discrete formulation of both methods to compute both the 3D velocity in the Z direction and the estimation of all the motion parameters.

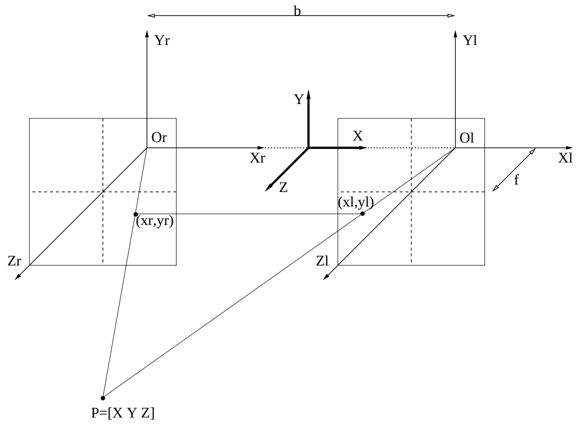


Figure 1. World and stereo coordinate system

The robustness of the methods was also analyzed as a function of the resolution in the depth estimates, in unstructured indoor scenarios.

A performance analysis of both methods in both formulations is performed and the critical input data variables as a function of noise are identi£ed.

2. Motion Estimation Problem

A 3D point in space will be represented by its coordinate vector $\mathbf{P} = [X\ Y\ Z]^T$ and the world coordinate system will be coincident with the cyclopean coordinate system. The cameras (with focal f) are parallel to each other separated by the baseline b. The ¤ow induced in the image planes is represented 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. Figure 1 shows the geometry of the stereo vision system and the world coordinate system.

We consider rigid motion. Let $\mathbf{V}$ be the total 3D velocity of point $\mathbf{P}$. As 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 $\mathbf{\Omega} = [\Omega_X\ \Omega_Y\ \Omega_Z]^T$ we have the 3D velocity given by $\mathbf{V} = \mathbf{\Omega} \times \mathbf{P} + \mathbf{t}$. Expanding this equation we obtain the third component of the 3D velocity of a point along the Z direction - V_Z which we propose to estimate. We have $V_Z = t_Z + \Omega_X Y - \Omega_Y X$.

In the discrete approach two points in space are related by a linear transformation composed by a rotation matrix and a translation vector such that $\mathbf{P}' = R \cdot \mathbf{P} + \mathbf{T}$ where $\mathbf{P}$ is a 3D point at time t and **P'** the same point at time t'. In the discrete formulation we will approximate the velocity of a 3D point by the £nite differences between the point coordinates in time t' and t, that is, $V_Z = Z' - Z$.

3. Differential approach

In this section we will consider 3D motion estimation from a differential standpoint. The differential optical ¤ow is available. First we will present two methods to estimate the 3D velocity in the Z direction - V_Z. The details and proofs of those methods are available in [1, 6, 4].

$\mathbf{V_Z}$ - Depth Constraint

The change in the depth of a point or rigid body over time is directly related to its velocity in 3D space. We can use this principle to relate the velocity in the Z direction with depth.

The depth at instant t' of a point should be the depth at the instant t plus the displacement in the Z direction - V_Z. This relationship is given by the following expression, the linear Depth Change Constraint Equation - DCCE:

$$V_Z = Z_t + Z_x \cdot v_x + Z_y \cdot v_y \tag{1}$$

where $Z(x, y, t)$ is the depth at a given time t, Z_x, Z_y and Z_t its spatial-temporal derivatives. v_x and v_y are the components of the optical ¤ow.

$\mathbf{V_Z}$ - Binocular Flow Constraint

1. The $x-$ coordinate of $\mathbf{P}$ is x_l in the left image and x_r in the right image.

Point $\mathbf{P}$ in £gure 1, its projection in each image plane ((x_l, y_l, f) and (x_r, y_r, f)) and the optical centres (O_l and O_r) de£ne two similar triangles, so that we can write the relationship:

$$\frac{Z}{b} = \frac{Z - f}{b - (x^r - x^l)} \tag{2}$$

Now, if we compute the temporal derivative of the equation 2, we obtain:

$$V_Z = -\frac{bf}{(x^r - x^l)^2} \cdot (v_x^r - v_x^l) = -\frac{Z^2}{bf} \Delta v_x \tag{3}$$

is given by equation 3 which yields a way to compute V_Z locally, i.e. one equation for each image point.

Motion parameters - Depth Constraint

The six motion parameters ($\overrightarrow{\Omega}$ and $\overrightarrow{t}$) can also be estimated. Replacing in the DCCE equation 1 the image velocities by their well-known relationships with the motion parameters, we obtain the following equation:

$$-Z_t = \begin{bmatrix} f\frac{Z_x}{Z} \\ f\frac{Z_y}{Z} \\ -\frac{Z + xZ_x + yZ_y}{Z} \\ -fZ_y - \frac{y}{f}(Z + xZ_x + yZ_y) \\ fZ_x + \frac{x}{f}(Z + xZ_x + yZ_y) \\ xZ_y - yZ_x] \end{bmatrix}^T \cdot \overrightarrow{\phi} \tag{4}$$

where $\overrightarrow{\phi}$ is the vector with the six motion parameters that we want to estimate.

Taking several points (more than six) we obtain an overdetermined linear system in $\overrightarrow{\phi}$.

Motion parameters - Binocular Flow Constraint

The 3D velocity in the Z direction (V_Z) can be expressed as a linear equation on three of the six parameters which can be substituted in the equation of binocular ¤ow. We obtain:

$$t_Z + \Omega_X Y - \Omega_Y X = -\frac{Z^2}{bf} \Delta v_x \Leftrightarrow$$

$$\Leftrightarrow \begin{bmatrix} \frac{\Delta v_x}{bf} \end{bmatrix} = \begin{bmatrix} \frac{-1}{Z^2} & \frac{-y}{fZ} & \frac{x}{Z} \end{bmatrix} \begin{bmatrix} t_Z \\ \Omega_X \\ \Omega_Y \end{bmatrix} \tag{5}$$

where we replaced the 3D point coordinates X and Y by its inverse perspective projection equations.

To recover the remaining parameters we propose the use of the optical ¤ow. For each image point, $\hat{t_Z}$, $\hat{\Omega_X}$, $\hat{\Omega_Y}$ and the image ¤ow (v_x,v_y) are known. So we can de£ne another linear system to estimate the other three motion parameters.

4. Discrete Formulation

In this section we present the discrete versions of both methods (to compute V_Z) and we also present a method to compute the motion parameters in the discrete formulation.

A discrete number of images is considered and we wish to recover the transformation between the world projected in consecutive images using the relationship $\mathbf{P}' = R{\cdot}\mathbf{P}{+}\mathbf{T}$.

The disparity information is available and also feature correspondences.

$\mathbf{V_Z}$ - Discrete DCCE

The DCCE equation in the discrete formulation is given by:

$$V_Z = Z_t + Z_x \Delta x + Z_y \Delta y \tag{6}$$

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 £nite differences of the image point coordinates.

V_Z Binocular Flow

The discrete binocular ¤ow method equation is given by:

$$V_Z \approx \frac{ZZ'}{bf}(d' - d) \qquad (7)$$

Motion parameters - Discrete formulation

Regarding the expression $\mathbf{P}' = R \cdot \mathbf{P} + \mathbf{T}$ and expanding it, it is possible to establish a relation between discrete motion parameters (R and $\mathbf{T}$) and the 3D points $\mathbf{P}$ and $\mathbf{P}'$. Transforming that equation into an overdetermined system, matrix R and vector $\mathbf{T}$ can be recovered. To recover the actual rotational motion parameters, considering that the motion parameters are constant over time, it is used the instantaneous approximation yielded by:

$$R \approx \begin{bmatrix} 1 & -\Omega_Z \Delta t & \Omega_Y \Delta t \\ \Omega_Z \Delta t & 1 & -\Omega_X \Delta t \\ -\Omega_Y \Delta t & \Omega_X \Delta t & 1 \end{bmatrix} \qquad (8)$$

and the translational motion parameters are given by:

$$\overrightarrow{t} = V^{-1} \cdot \mathbf{T} \qquad (9)$$

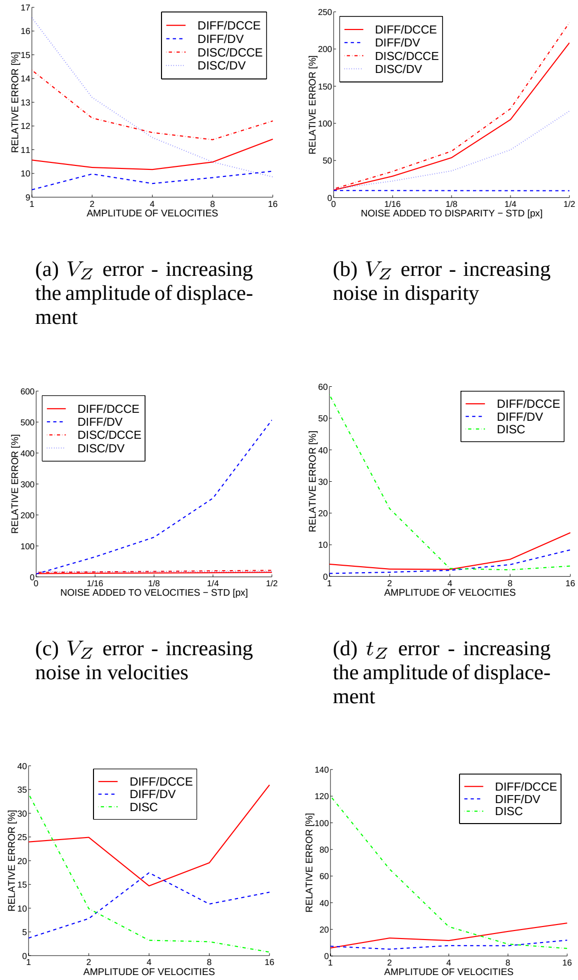
where V is the matrix given by the closed form of a point trajectory in rigid body motion (see [8, 5, 7]).

5. Experiments and Conclusions

We performed two groups of experimental tests. First we used synthetic images to analyze the performance of both methods changing the resolution of disparity/depth £elds, adding noise to disparity and velocities and changing the displacements from image to image, in order to identify the critical variables for both methods. Synthetic images are made up of front and lateral walls, ground and two obstacles. This world was projected into a virtual stereo head mounted on a navigating robot with known motion parameters (the basis velocity equals one focal length per frame).

The second group of experiments was done with real images obtained with known motion parameters. For reasons of lack of space we will report only the main results and conclusions. Extensive testing, results and further details are described in the Technical Report [1].

In £gure 2 we report the relative mean error of estimation of V_Z and of the most numerically stable motion parameters: t_Z, Ω_X and Ω_Y. The sequence used is a complex sequence with translation in all axis and rotation over the X and Y axes. Three sensitivity tests are presented: increasing the amplitude of velocities (multiply the velocity by a factor), increasing the standard deviation of the noise added to disparity and of the noise added to velocities.



(a) V_Z error - increasing the amplitude of displacement

(b) V_Z error - increasing noise in disparity

(c) V_Z error - increasing noise in velocities

(d) t_Z error - increasing the amplitude of displacement

(e) Ω_X error - increasing the amplitude of displacement

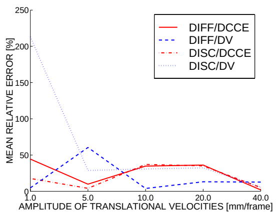
(f) Ω_Y error - increasing the amplitude of displacement

Figure 2. Error analysis in synthetic images.

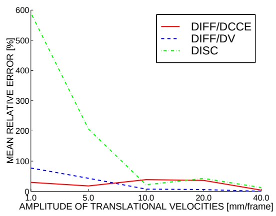
To analyze the performance of both methods with real images a pair of cameras with focal length of 6.5 mm and a baseline of 130 mm was used. The pixel width is 12.0 μm. The stereo head was attached to a manipulator with reasonable precision allowing complex paths to be performed. The real images were acquired without any special care with illumination, shadows and other lighting effects. Figure 3 shows one left frame and the corresponding disparity map. To compute the optical ¤ow it was used the Lucas-Kanade algorithm and for feature correspondences it was used a simple corner detection and matching algorithm.

Figure 4 shows the relative error of V_Z and of the motion parameters t_Z and Ω_Y.

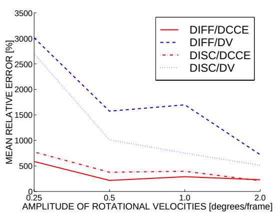
Figure 3. Real image and corresponding disparity map.



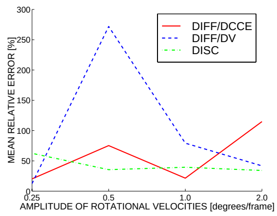
(a) Relative error of V_Z with increasing displacement



(b) Relative error of t_Z with increasing displacement



(c) Relative error of V_Z with increasing rotation



(d) Relative error of Ω_Y with increasing rotation

Figure 4. Error analysis in real images

5.1. Conclusions

Several conclusions can be drawn regarding the results obtained with synthetic and real images.

In synthetic sequences one can observe that the DCCE method is very sensitive to the disparity resolution and to the noise in disparity. The DV method, on the other hand, is more sensitive to the noise in optical ¤ow and feature correspondences. However, both methods get better results when the displacement between consecutive images is bigger. As was expected the discrete approach is better only for big displacements (translation and/or rotation).

Concerning the estimation of V_Z, it can be concluded that the path is very important in the accuracy. 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, however, the estimation results are very poor. The DCCE method presents slightly better results than the DV method in translational and mixed paths and 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 are necessary (in order to allow for the cancellation of the error).

For the computation of the complete motion parameters - $\overrightarrow{\phi}$, 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 optical axis direction is the parameter with best estimation values. The DCCE method is again the best one.

Generally, the results get better for higher velocities/displacements.

Concerning the comparison of the differential and discrete formulations, it can be concluded that in the estimation of V_Z, there are few differences between both approaches but, in the estimation of $\overrightarrow{\phi}$, the discrete one presents smaller errors in the unstable parameters and again few differences in the other three parameters. It is observed that the discrete formulation gets better results for higher displacements.

It can be also concluded that for the DCCE method the critical factor is the resolution of the depth £eld. On the other hand, for the DV method the critical factor is the accuracy of the optical ¤ow (as was expected from the analysis of the uncertainty propagation [2, 3]).

6. Acknowledge

The authors gratefully acknowledge the support of project PRAXIS/P/EEI/10252/1998, funded by the Portuguese Foundation for Science and Technology.

References

[1] N. Gonçalves. Estimação de movimento usando estereovisão. Master's thesis, Department of Electrical and Computers Engineering of the Faculty of Science and Technology of the University of Coimbra, 2002.
[2] N. Gonçalves and H. Araújo. Analysis of two methods for the estimation of partial 3d velocity. In *9th SIRS*, Toulouse, France, 2001.
[3] N. Gonçalves and H. Araújo. Uncertainty propagation in estimation of partial 3d velocity. In *10th MED2002*, Lisbon, Portugal, 2002.
[4] M. Harville, A. Rahimi, T. Darrell, G. Gordon, and J. Wood£ll. 3d pose tracking with linear depth and brightness constraints. In *ICCV*, Corfu, Greece, 1999.
[5] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade. Three-dimensional scene ¤ow. In *ICCV*, 1999.
[6] A. M. Waxman and J. H. Duncan. Binocular image ¤ows: Steps towards stereo - motion fusion. *PAMI*, 8(6):715–729, Nov. 1986.
[7] Y. Zhang and C. Kambhamettu. On 3d scene ¤ow and structure estimation. In *CVPR*, 2001.
[8] Z. Zhang and O. Faugeras. Three-dimensional motion computation and object segmentation in a long sequence of stereo frames. *IJCV*, 7(3):211–241, 1992.