Active/Dynamic Stereo for Navigation

Enrico Grosso, Massimo Tistarelli and Giulio Sandini

University of Genoa
Department of Communication, Computer and Systems Science
Integrated Laboratory for Advanced Robotics (LIRA - Lab)
Via Opera Pia 11A - 16145 Genoa, Italy

Abstract. Stereo vision and motion analysis have been frequently used to infer scene structure and to control the movement of a mobile vehicle or a robot arm. Unfortunately, when considered separately, these methods present intrinsic difficulties and a simple fusion of the respective results has been proved to be insufficient in practice.

The paper presents a cooperative schema in which the binocular disparity is computed for corresponding points in several stereo frames and it is used, together with optical flow, to compute the time-to-impact. The formulation of the problem takes into account translation of the stereo set-up and rotation of the cameras while tracking an environmental point and performing one degree of freedom active vergence control. Experiments on a stereo sequence from a real scene are presented and discussed.

1 Introduction

Visual coordination of actions is essentially a real-time problem. It is more and more clear that a lot of complex operations can rely on reflexes to visual stimuli [Bro86]. For example closed loop visual control has been implemented at about video rate for obstacle detection and avoidance [FGMS90], target tracking [CGS91] and gross shape understanding [TK91].

In this paper we face the problem of "visual navigation". The main goal is to perform task-driven measurements of the scene, detecting corridors of free space along which the robot can safely navigate.

The proposed cooperative schema uses binocular disparity, computed on several image pairs and over time. In the past the problem of fusing motion and stereo in mutually useful way has been faced by different researchers. Nevertheless, there is a great difference between the approaches where the results of the two modalities are considered separately (for instance using depth from stereo to compute motion parameters [Mut86]) and the rather different approach based upon more integrated relations (for instance the temporal derivative of disparity [WD86, LD88]).

In the following we will explain how stereo disparity and image velocity are combined to obtain a $2\frac{1}{2}$D representation of the scene, suitable for visual navigation, which is either in terms of *time-to-impact* or *relative-depth* referred to the distance of the cameras from the fixation point. Only image-derived quantities are used except for the vergence angles of the cameras which could be actively controlled during the robot motion [OC90], and can be measured directly on the motors (with optical encoders).

As a generalization of a previous work [TGS91] we consider also a rotational motion of the cameras around the vertical axes and we derive, from temporal correspondence of image points, the relative rotation of the stereo base-line. This rotation is then used to correct optical flow or relative depth.

* This work has been partially funded by the Esprit projects P2502 VOILA and P3274 FIRST.

2 The experimental set-up

The experimental set-up is based on a computer-controlled mobile platform TRC *Labmate* with two cameras connected to a VDS *7001 Eidobrain* image processing workstation. The cameras are arranged as to verge toward a point in space. Sequences of stereo images are captured at a variable frequency, up to video rate, during the motion of the vehicle (left and right images are captured simultaneously, thanks to the Eidobrain image processor). At present the cameras are not motorized, therefore we moved the vehicle step by step, adjusting manually the orientation of the two cameras as to always verge on the same point in the scene. In this way we simulated a tracking motion of both cameras on a moving vehicle.

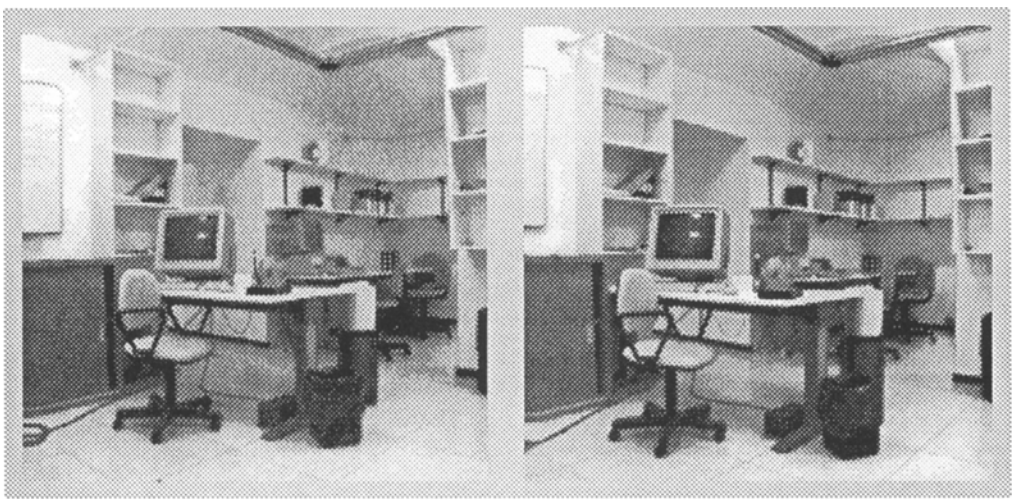

Fig. 1. First stereo pair of the acquired sequence.

In figure 1 the first stereo pair, from a sequence of 15, is shown. The vehicle was moving forward at about 100 mm per frame. The sequence has been taken inside the LIRA lab. Many objects were in the scene at different depths. The vehicle was undergoing an almost straight trajectory with a very small steering toward left, while the cameras were fixating a stick on the desk in the foreground.

3 Stereo and motion analysis

3.1 Stereo analysis

The stereo vision algorithm is based on a regional multiresolution approach [GST89] and produces, at each instant of time, a disparity map between the points in the left and in the right image (see figure 3).

With reference to figure 2 we define the *K function* [TGS91] as:

$$K(\alpha,\beta,\gamma,\delta) = \frac{\tan(\alpha-\gamma)\cdot\tan(\beta+\delta)}{\tan(\alpha-\gamma)+\tan(\beta+\delta)} \quad (1)$$

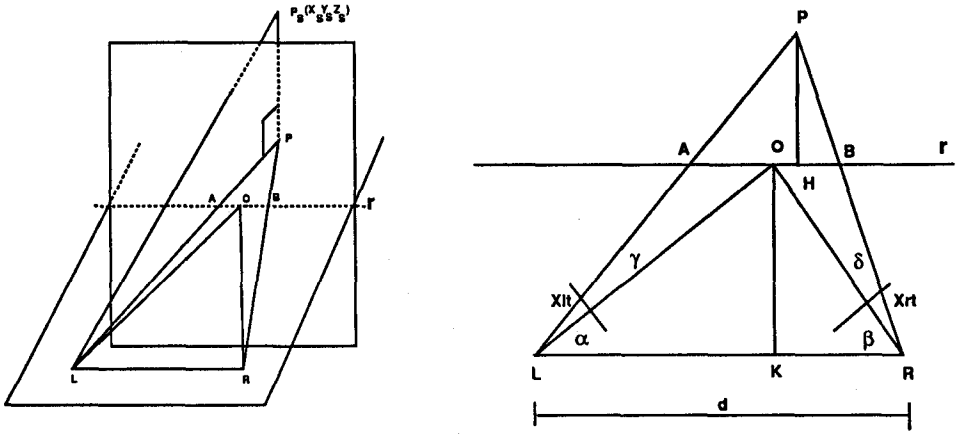

Fig. 2. Schematic representation of the stereo coordinate system.

where α and β are the vergence angles, $\gamma = \arctan\left(\frac{x_{lt}}{F}\right)$ and $\delta = \arctan\left(\frac{x_{rt}}{F}\right)$ define the position of two corresponding points on the image planes and $x_{rt} = x_{lt} + D$ where D is the known disparity. The depth is computed as:

$$Z_s = d \cdot K(\alpha, \beta, \gamma, \delta) \tag{2}$$

The knowledge of the focal length is required to compute the angular quantities.

3.2 Motion analysis

The temporal evolution of image fetures (corresponding to objects in the scene) is described as the instantaneous image velocity (optical flow). The optical flow $\mathbf{V} = (u, v)$ is computed from a monocular image sequence by solving an over-determined system of linear equations in the unknown terms (u, v) [HS81, UGVT88, TS90]:

$$\frac{d}{dt} I = 0 \qquad \frac{d}{dt} \nabla I = 0$$

where I represents the image intensity of the point (x, y) at time t. The least squares solution of these equations can be computed for each point on the image plane [TS90].

In figure 4 the optical flow of the sixth image of the sequence is shown. The image velocity can be described as a function of the camera parameters and split into two terms depending on the rotational and translational components of camera velocity respectively.

If the rotational part of the flow field $\mathbf{V}_r$ can be computed (for instance from proprioceptive data), $\mathbf{V}_t$ is determined by subtracting $\mathbf{V}_r$ from $\mathbf{V}$. From the translational optical flow, the time-to-impact can be computed as:

$$T = \frac{\Delta}{|\mathbf{V}_t|} \tag{3}$$

where Δ is the distance of the considered point, on the image plane, from the FOE.

The estimation of the FOE position is still a critical step; we will show how it can be avoided by using stereo disparity.

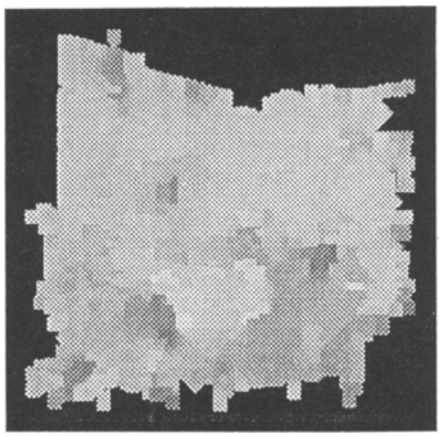

Fig. 3. Disparity computed for the 6th stereo pair of the sequence; negative values are depicted using darker gray levels.

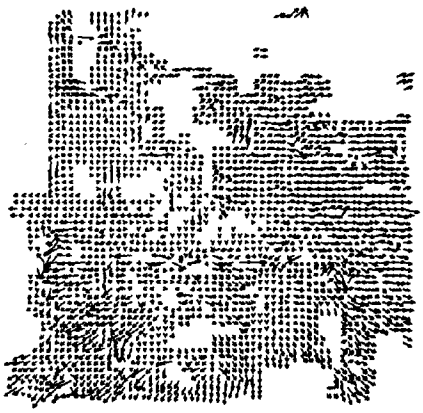

Fig. 4. Optical flow relative to the 6th left image of the sequence.

4 Stereo - motion geometry

Even though the depth estimates from stereo and motion are expressed using the same metric, they are not homogeneous because they are related to different reference frames. In the case of stereo, depth is referred to an axis orthogonal to the baseline (it defines the stereo camera geometry) while for motion it is measured along a direction parallel to the optical axis of the (left or right) camera. We have to consider the two reference frames and a relation between them:

$$Z_s(x,y) \;=\; Z_m(x,y)\; h(x) \qquad\qquad h(x) \;=\; \sin\alpha \;+\; \frac{x}{F}\cos\alpha \tag{4}$$

where α is the vergence angle of the left camera, F is the focal length of the camera in pixels and x is the horizontal coordinate of the considered point on the image plane (see figure 2). We choose to adopt the stereo reference frame, because it is symmetric with respect to the cameras, therefore all the measurements derived from motion are corrected accordingly to the factor $h(x)$.

5 Rotational motion and vergence control

The translational case analyzed in [TGS91] can be generalized by considering a planar motion of the vehicle with a rotational degree of freedom and with the two cameras tracking a point in space.

As the cameras and the vehicle are moving independently, we are interested in computing the global camera rotation resulting from both vehicle global motion and vergence/tracking motion of the cameras.

Figure 5 helps to clarify the problem. Previous work [KP86] shows that, from a theoretical point of view, it is possible to compute the vergence of the two cameras from

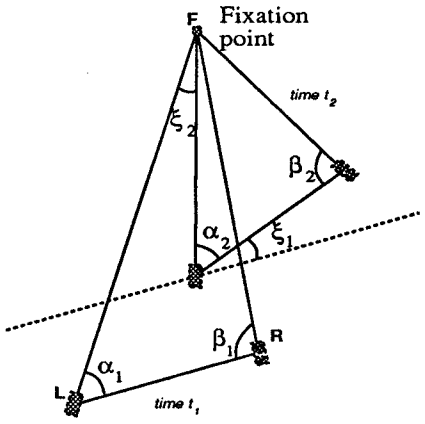

Fig. 5. Rotation of the stereo system during the motion.

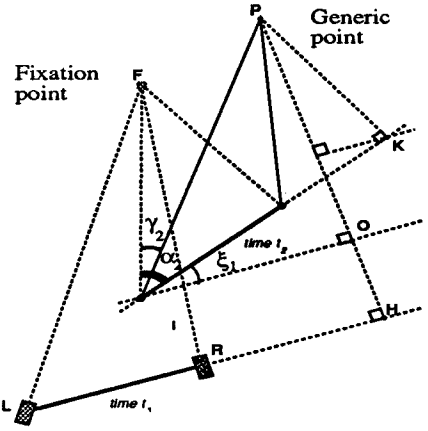

Fig. 6. Correction of the relative depth using rotation.

corresponding left and right image points. However, camera resolution and numerical instability problems make difficult a practical application of the theory. Moreover, as it appears in figure 5 the computation of the vergence angle is insufficient to completely determine the position of the stereo pair in space. For this reason the angle ξ_1 or, alternatively, ξ_2 must be computed. We first assume the vergence angles of the cameras to be known at a given time instant; for example they can be measured by optical encoders mounted directly on the motors.

The basic idea for computing the rotational angle is to locate two invariant points on the scene space and use their projection, along with the disparity and optical flow measurements, to describe the temporal evolution of the stereo pair. The first point considered is the fixation point, as it is "physically" tracked over time and kept in the image center. Other points are obtained by computing the image velocity and tracking them over successive frames.

In figure 7 the position of the two invariant points (F and P), projected on the ZX plane, with the projection rays is shown.

Considering now the stereo system at time t_1 we can compute, by applying basic trigonometric relations, the oriented angle between the 2D vectors **FP** and **LR**:

$$\tan(\theta_1) = \frac{\tan(\alpha_1-\gamma_1)\cdot\tan(\beta_1+\delta_1)\cdot[\tan(\alpha_1)+\tan(\beta_1)]}{\tan(\beta_1+\delta_1)\cdot[\tan(\alpha_1)+\tan(\beta_1)]-\tan(\beta_1)\cdot[\tan(\alpha_1-\gamma_1)+\tan(\beta_1+\delta_1)]} -$$

$$\frac{\tan(\alpha_1)\cdot\tan(\beta_1)\cdot[\tan(\alpha_1-\gamma_1)+\tan(\beta_1+\delta_1)]}{\tan(\beta_1+\delta_1)\cdot[\tan(\alpha_1)+\tan(\beta_1)]-\tan(\beta_1)\cdot[\tan(\alpha_1-\gamma_1)+\tan(\beta_1+\delta_1)]} \quad (5)$$

It is worth noting that the angle θ_1 must be bounded within the range $[0-2\pi)$. In a similar way the angle θ_2 at time t_2 can be computed. ξ_1 is derived as:

$$\xi_1 = \theta_1 - \theta_2 \quad (6)$$

In this formulation ξ_1 represents the rotation of the base-line at time t_2 with respect to the position at time t_1; the measurements of ξ_1 can be performed using a subset or

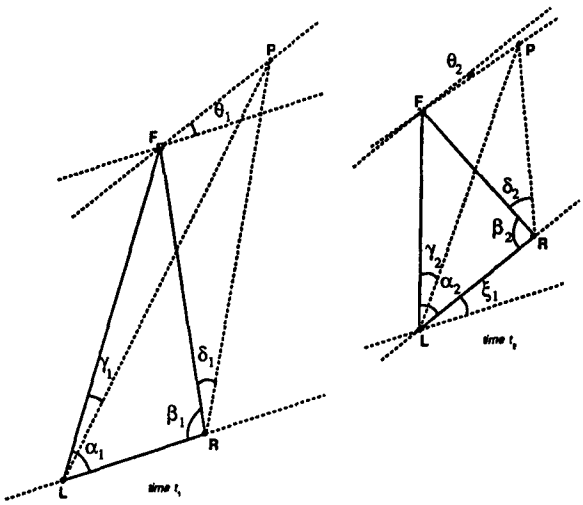

Fig. 7. Geometry of the rotation (projection on the stereo plane).

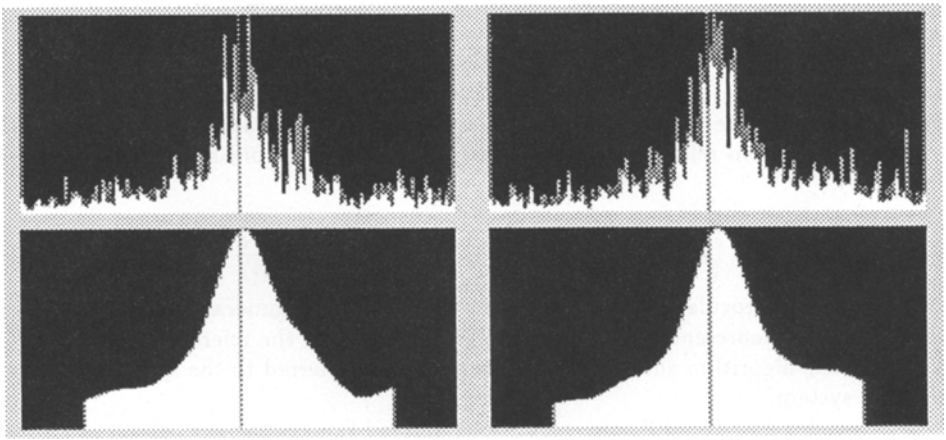

Fig. 8. Rough and smoothed histograms of the angles computed from frames 5-6 and frames 7-8, respectively. The abscissa scale goes from -0.16 radians to 0.16 radians. The maxima computed in the smoothed histograms correspond to 0.00625 and 0.0075 radians respectively.

also all the image points. In the noiseless case, all the image points will produce identical estimates. In the case of objects moving within the field of view it should be easy to separate the different peaks in the histogram of the computed values corresponding to moving and still objects. Figure 8 shows two histograms related, to frames 5-6 and 7-8 respectively.

In order to compute the angle ξ_1 the following parameters must be known or measured:

- α and β, the vergence angles of the left and right camera respectively, referred to the stereo baseline.
- γ and δ, the angular coordinates of the considered environmental point computed on the left and right camera respectively. The computed image disparity is used to

establish the correspondence between image points on the two image planes.
- The optical flow computed at time t_1.

5.1 Using rotation to correct image-derived data

The rotation of the stereo system affects the optical flow adding a rotational component to the translational one. As a consequence, the computed rotation from image-data allows, by difference, the computation of the component V_t and, finally, the time-to-impact.

An alternative way is to correct the relative depth coming from stereo analysis taking into account the rotation of the stereo pair. In this case, with reference to figure 6, $Z_s(t_1) = \overline{PH}$ is the depth at time t_1 and $Z_s(t_2) = \overline{PK}$ is the depth of the same point at time t_2. The correction must eliminate only the rotational component, therefore we can write:

$$\overline{PO} \;=\; Z_{str}(t_2) = Z_s(t_2) \cdot \left[\cos\xi_1 + \frac{\sin\xi_1}{\tan(\alpha_2 - \gamma_2)}\right] \tag{7}$$

In the remainder of the paper we denote with Z the translational component Z_{str}.

6 Using neighborhoods to compute *time-to-impact* and relative depth

From the results presented in the previous sections we can observe that both stereo- and motion-derived relative-depth depend on some external parameters. More explicitly, writing the equations related to a common reference frame (the one adopted by the stereo algorithm):

$$K_i \;=\; \frac{Z_i}{d} \qquad\qquad T_i^s \;=\; \frac{h(x_i)}{h(0)} T_i \;=\; \frac{Z_i}{W_z} \tag{8}$$

where d is the interocular baseline, W_z is the velocity of the camera along the stereo reference frame, T_i represents the time-to-impact measured in the reference frame adopted by the motion algorithm and T_i^s is the time-to-impact referred to the symmetric, stereo reference system.

We consider now two different expressions derived from equations (8).

$$T_i = \frac{d}{W_z} \cdot K_i \cdot \frac{h(0)}{h(x_i)} \qquad\qquad \frac{Z_i}{Z_l} \;=\; \frac{T_i}{K_l} \cdot \frac{h(x_i)}{h(0)} \cdot \frac{W_z}{d} \tag{9}$$

First equation represents the *time-to-impact* with respect to the motion system while the second equation represents a generic relative measure of the point (x_i, y_i) with respect to the point (x_l, y_l). Our goal is now to eliminate the ratio $\frac{W_z}{d}$. A first way to proceed is to rewrite the first equation of (9) for a generic point (x_j, y_j) whose velocity with respect to (x_i, y_i) is zero.

$$T_j = \frac{d}{W_z} \cdot K_j \cdot \frac{h(0)}{h(x_j)} \tag{10}$$

Using the first equation of (9) and equation (10) we can compute a new expression for $\frac{W_z}{d}$:

$$\frac{W_z}{d} \;=\; \frac{K_i \;-\; K_j}{T_i h(x_i) \;-\; T_j h(x_j)} h(0) \tag{11}$$

where (x_i, y_i) and (x_j, y_j) are two points on the image plane of the left (or right) camera. This formulation is possible if we are not measuring the distance of a flat surface, because of the difference of K at the numerator and the difference of T at the denominator. Substituting now in equations (9) we obtain:

$$T_i = T_j \cdot \frac{K_i}{K_j} \cdot \frac{h(x_j)}{h(x_i)} \qquad \frac{Z_i}{Z_l} = \frac{(K_i - K_j)\, T_i}{\left(T_i - T_j \cdot \frac{h(x_j)}{h(x_i)}\right) K_l} \tag{12}$$

The two equations are the first important result. In particular the second equation directly relates the relative-depth to the time-to-impact and stereo disparity (i.e. the K function). The relative-depth or *time-to-impact* can be computed more robustly by integrating several measurements over a small neighborhood of the considered point (x_j, y_j), for example with a simple average. The only critical factor in the second equation of (12) is the time-to-impact which usually requires the estimation of the FOE position. However, with a minimum effort it is possible to exploit further the motion equations to directly relate time-to-impact and also relative-depth to stereo disparity (the K function) and optical flow only.

7 Using optical flow to compute *time-to-impact*

We will exploit now the temporal evolution of disparity. If the optical flow and the disparity map are computed at time t, the disparity relative to the same point in space at the successive time instant, can be obtained by searching for a matching around the predicted disparity, which must be shifted by the velocity vector to take into account the motion.

As $\frac{W_x}{d}$ is a constant factor for a given stereo image, it is possible to compute a robust estimate by taking the average over a neighborhood [TGS91]:

$$\frac{W_x}{d} = \Delta_K = \frac{1}{\Delta t\, N^2} \sum_i [K_i(t) - K_i(t + \Delta t)] \tag{13}$$

Given the optical flow $\mathbf{V} = (u, v)$ and the map of the values of the K function at time t, the value of $K_i(t + \Delta t)$ is obtained by considering the image point $(x_i + u_i, y_i + v_i)$ on the map at time $t + \Delta t$.

The value of Δ_K for the 6th stereo has been computed by applying equation (13) at each image point. By taking the average of the values of Δ_K over the all image, a value of $\frac{W_x}{d}$ equal to 0.23 has been obtained. This value must be compared to the velocity of the vehicle, which was about 100 millimeters per frame along the Z axis and the interocular baseline which was about 335 millimeters. Due to the motion drift of the vehicle and the fact that the baseline has been measured by hand, it is most likely that the given values of the velocity and baseline are slightly wrong.

By using equation (13) to substitute $\frac{W_x}{d}$ in the first equation of (9), it is possible to obtain a simple relation for the time-to-impact, which involves the optical flow to estimate Δ_K:

$$T_i^s = \frac{h(x_i)}{h(0)}\, T_i = \frac{K_i}{\Delta_K} \tag{14}$$

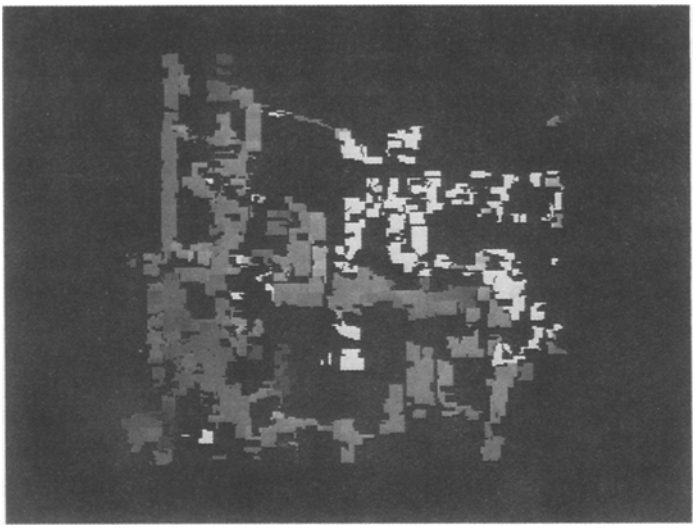

Fig. 9. Time-to-impact computed using eq. (14) for the 6th pair of the sequence; darker regions correspond to closer objects.

This estimate of the time-to-impact is very robust and does not require the computation of the FOE (see figure 9). From equation (13) and the second equation of (9), it is possible to obtain a new expression for the relative-depth:

$$\frac{Z_i}{Z_l} = \frac{\Delta_K}{K_l} \cdot \frac{h(x_i)}{h(0)} T_i \tag{15}$$

There is, in principle, a different way to exploit the optical flow. For completeness we will briefly outline this aspect.

From the knowledge of the translational component $\mathbf{V}_t$ of $\mathbf{V}$ it is possible to write for a generic point (x_h, y_h):

$$\begin{cases} V_{tx}(x_h, y_h) = \frac{x_h}{T_h} - \frac{FW_x}{Z} \\ V_{ty}(x_h, y_h) = \frac{y_h}{T_h} - \frac{FW_y}{Z} \end{cases}$$

Considering two points (x_i, y_i) and (x_j, y_j) and eliminating F/Z and W_x/W_y we express a relation between T_i and T_j:

$$\frac{x_i - T_i V_{tx}(x_i, y_i)}{y_i - T_i V_{ty}(x_i, y_i)} = \frac{x_j - T_j V_{tx}(x_j, y_j)}{y_j - T_j V_{ty}(x_j, y_j)} \tag{16}$$

Now, combining equation (16) with the first of (12) we can obtain a new equation of order two for T_i. T_i is expressed in this case as a function of the coordinates and the translational flow of the two considered points. As in the case of equations (12) the *time-to-impact* can be computed more robustly by averaging the measurements over a small neighborhood of the considered point (x_i, y_i).

8 Conclusions

Robot navigation requires simple computational schemes able to select and to exploit relevant visual information. The paper addressed the problem of stereo motion cooperation proposing a new approach to merge binocular disparity and optical flow. The formulation of the problem takes into account active vergence control but limits the rotation of the cameras around the vertical axes. The rotation of the stereo base-line is first extracted using stereo disparity and optical flow; after that it is used to correct stereo relative depth and to compute dynamic quantities like *time-to-impact* directly from stereo information, using optical flow to determine the temporal evolution of disparity in the sequence. Future work will be addressed to include in the formulation the tilt angles of the two cameras.

References

[Bro86] R.A. Brooks. A robust layered control system for a mobile robot. *IEEE Trans. on Robotics and Automat.*, RA-2:14–23, April 1986.

[CGS91] G. Casalino, G. Germano, and G. Sandini. Tracking with a robot head. In *Proc. of ESA Workshop on Computer Vision and Image Proc essing for Spaceborn Applications*, Noordwijk, June 10-12, 1991.

[FGMS90] F. Ferrari, E. Grosso, M. Magrassi, and G. Sandini. A stereo vision system for real time obstacle avoidance in unknown environment. In *Proc. of Intl. Workshop on Intelligent Robots and Systems*, Tokyo, Japan, July 1990. IEEE Computer Society.

[GST89] E. Grosso, G. Sandini, and M. Tistarelli. 3d object reconstruction using stereo and motion. *IEEE Trans. on Syst. Man and Cybern.*, SMC-19, No. 6, November/December 1989.

[HS81] B. K. P. Horn and B. G. Schunck. Determining optical flow. *Artificial Intelligence*, 17 No.1-3:185–204, 1981.

[KP86] B. Kamgar-Parsi. Practical computation of pan and tilt angles in stereo. Technical Report CS-TR-1640, University of Maryland, College Park, MD, March 1986.

[LD88] L. Li and J.H. Duncan. Recovering three-dimensional translational velocity and establishing stereo correspondence from binocular image flows. Technical Report CS-TR-2041, University of Maryland, College Park, MD, May 1988.

[Mut86] K.M. Mutch. Determining object translation information using stereoscopic motion. *IEEE Trans. on PAMI - 8*, No. 6, 1986.

[OC90] T.J. Olson and D.J. Coombs. Real-time vergence control for binocular robots. Technical Report 348, University of Rochester - Dept. of Computer Science, 1990.

[TGS91] M. Tistarelli, E. Grosso, and G. Sandini. Dynamic stereo in visual navigation. In *Proc. of Int. Conf. on Computer Vision and Pattern Recognition*, Lahaina, Maui, Hawaii, June 1991.

[TK91] C. Tomasi and T. Kanade. Shape and motion from image streams: a factorization method. Technical Report CS-91-105, Carnegie Mellon University, Pittsburgh, PA, January 1991.

[TS90] M. Tistarelli and G. Sandini. Estimation of depth from motion using an anthropomorphic visual sensor. *Image and Vision Computing*, 8, No. 4:271–278, 1990.

[UGVT88] S. Uras, F. Girosi, A. Verri, and V. Torre. Computational approach to motion perception. *Biological Cybernetics*, 1988.

[WD86] A.M. Waxman and J.H. Duncan. Binocular image flows: Steps toward stereo-motion fusion. *IEEE Trans. on PAMI - 8*, No. 6, 1986.

This article was processed using the LaTeX macro package with ECCV92 style