Active/Dynamic Stereo Vision

Enrico Grosso and Massimo Tistarelli

Abstract—**Visual navigation is a challenging issue in automated robot control. In many robot applications, like object manipulation in hazardous environments or autonomous locomotion, it is necessary to automatically detect and avoid obstacles while planning a safe trajectory. In this context the detection of corridors of free space along the robot trajectory is a very important capability which requires nontrivial visual processing. In most cases it is possible to take advantage of the active control of the cameras.**

In this paper we propose a cooperative schema in which motion and stereo vision are used to infer scene structure and determine free space areas. Binocular disparity, computed on several stereo images over time, is combined with optical flow from the same sequence to obtain a relative-depth map of the scene. Both the time-to-impact and depth scaled by the distance of the camera from the fixation point in space are considered as good, relative measurements which are based on the viewer, but centered on the environment.

The need for calibrated parameters is considerably reduced by using an active control strategy. The cameras track a point in space independently of the robot motion and the full rotation of the head, which includes the unknown robot motion, is derived from binocular image data.

The feasibility of the approach in real robotic applications is demonstrated by several experiments performed on real image data acquired from an autonomous vehicle and a prototype camera head.

Index Terms—**Active vision, dynamic vision, time-to-impact, stereo vision, motion analysis, navigation.**

I. Introduction

A robot with "intelligent" behavior must be capable of coping with unprecise situations. However, this capability not always imply very sophisticated high-level reasoning capabilities. For example, collecting soda cans in an unpredictable indoor environment and putting them in a predefined place is a kind of intelligent task which has been demonstrated to be solvable by means of basic sensory capabilities coordinated by reflex-type (or insect-like) behaviors [1], [2]. Even though more sophisticated visual and reasoning processes can be envisaged, still many other, even complex operations can be performed relying on reflexes to visual stimuli [3], [4].

One of the most interesting and useful aspects of the active vision paradigm is the use of motion (and in general the ability of the observer to act and interact with the environment) to guide *continuous* image data acquisition to enrich the measurements of the scene. Overall, this implies an increase in the amount of incoming data which, in turn, may require an appropriate data-reduction strategy (for example limiting the frequency content of the image sequence with appropriate transformations like the log-polar mapping [5]). On the other hand, this strategy can enormously simplify the computational schema underlying a given visual task or measurement. In the last years several examples have been presented, where active movements can induce more information and allow simpler computational schema to be adopted [6], [7], [8], [9], [10]. This is very important wishing to design working systems, where simpler processes allow the implementation of real-time systems, which can perform several measurements over time.

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 schema combines optical flow and binocular disparity, computed on several image pairs over time.

In the past the problem of fusing motion and stereo in a mutually useful way has been faced by different researchers. Some of them integrated the measurements obtained from stereo techniques repeated over time according to the relative uncertainty, building a consistent representation of the environment [11], [12], [13], [14]. Ahuja and Abbott [15] used exploratory fixations to apply several modalities like stereo, focus and vergence to recover the object structure and disambiguate occlusions. With the same purpose, Grosso et al. [16] integrated depth measurements derived from both stereo disparity and optical flow. In [17] a simple stereo technique, applied to a single image scan-line, is used to determine the focus of expansion relative to the horizon and control the heading of a robot vehicle. It is interesting to note that the low complexity of this technique allows to continuously repeat the measurements over time, improving the motion control. In general, it is possible to make a distinction between the approaches where the results of stereo and motion analysis are considered separately and the rather different approach based upon more integrated relations. Within this group falls the work reported in [18], [19] where the temporal derivative of disparity is exploited, and the dynamic stereo approach [20], [21] considered in this paper.

Developing a sensory system for autonomous navigation (considered as a possible application of the approach presented in this paper), the integration of multiple visual modalities can be the key to overcome some common problems:

- **Calibration**. Many vision algorithms rely on the knowledge of some well calibrated parameters relative to the camera-robot system [22]. In this paper stereo disparity and optical flow are used to avoid the explicit calibration of external parameters. In particular, this method prevents the drawbacks of earlier techniques like the use of the baseline length and the need for the knowledge of the

Manuscript received Apr. 13. 1993; revised May 22, 1995.

The authors are with the University of Genoa, Department of Communication, Computer, and Systems Science, Integrated Laboratory for Advanced Robotics (LIRA-Lab), Via Opera Pia 13-16145 Genoa, Italy; e-mail: {grenri,tista}@dist.unige.it.

IEEECS Log Number P95123.

robot's motion. Moreover, by exploiting "active behaviors" it is possible to rely on self-calibration techniques which reduce the number of required parameters [23], [24]. This methodology is applied to determine the focal length of the cameras, which is required by the stereo algorithm.

- **Accuracy**. Even though accuracy can be improved by a statistical integration of independent measurements [25], [12], it has been shown [26] that accurate estimates are not necessary for navigation purposes. For this reason this topic is not explicitly addressed in this paper, but the analysis is focussed on the robustness of the measurements which is crucial for safe vehicle control.
- **Robustness**. Mainly for safety reasons the vision system must be very robust with respect to "noise," either electronic, due to the sensors, or dynamic, due to inaccurate motion of the vehicle. Moreover, the result of the visual processing must be error-free, even at the cost of "false alarms." Algorithmic robustness can be certainly improved by providing independent estimates of the same quantity, like those obtained from different visual modalities, and a method to combine them. Moreover, robustness and numerical stability can be also achieved by adopting a cooperative schema in which different visual cues contribute to the estimation of scene structure. In this paper we show that, computing the time-to-impact from the temporal evolution of disparity, the error in the final estimate does not depend on differential measurements (which are notoriously not very robust), which is the case computing the time-to-impact directly from optical flow.
- **Metrics**. It is still not well understood which is the best metric to use representing the environment. Standard metrics, like inches or centimeters, are not well suited because they require an additional calibration to relate them to the image-derived data and to the motion control. On the other hand, if an active interaction with the environment is engaged [7], [6], metrics which are intrinsic to the observer, are certainly best suited to guide the robot's behavior [10], [9], [27]. In this paper the time-to-impact is computed, a viewer-based metric among the most useful for navigation. Other viewer-based metrics are also addressed like the depth scaled by the inter-ocular baseline, or by the distance from a reference point in space [10], [9].

In this paper, binocular disparity, computed for each image pair over time, and monocular optical flow are combined, via simple relations[1], to obtain a 2½D representation of the scene, suitable for visual navigation, which is either in terms of *time-to-impact* or *relative-depth*. The knowledge of the baseline length and the robot motion are not needed. The time-to-impact is computed avoiding the estimation of the Focus of Expansion from optical flow, while stereo reconstruction is used to compute the full rotation of the head-eye system in space. Image-derived quantities are used except for the focal length and vergence angles of the cameras which are actively controlled during the robot motion, and are measured directly on the motor axes by optical encoders.[2]

II. Stereo Vision and Dynamic Visual Processing

The dynamic stereo approach is based on two fundamental *building blocks*: stereo and motion analysis. The algorithms and notations relative to these modalities are briefly presented in the following sections.

A. Stereo Vision

The stereo vision algorithm uses a coarse to fine approach to perform a regional correlation. The resulting disparity map states the correspondence of the points in the left and right image. We skip over a detailed description of the stereo algorithm (see, for instance, [16]); instead, starting from the knowledge of the related disparity, we will concentrate on the computation of relative depth.

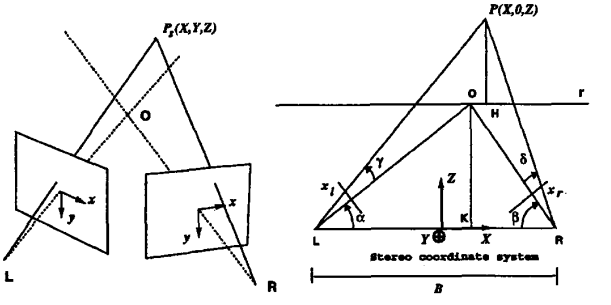


Fig. 1. Schematic representation of the stereo coordinate system. On the left the configuration in space is shown—the y axes of the cameras and of the stereo coordinate system are parallel and orthogonal to the plane defined by the two optical axes. On the right, the projection of the point $P_s = (X, Y, Z)$ on the plane $Y = 0$ (the point P) is shown.

Fig. 1 shows the considered configuration: let P be the projection on the stereo plane (the plane defined by the two optical axes) of a point P_s in space. We define the ***K function*** as:

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

where α and β are the vergence angles, $\gamma = \arctan\left(\frac{x_l}{F_l}\right)$ and $\delta = \arctan\left(\frac{x_r}{F_r}\right)$ define the position of two corresponding points on the image planes, $x_r = x_l + D$ (D is the known disparity), F_l and F_r are the focal lengths of the left and right camera measured in pixels. It is easy to prove that the depth Z referred to the stereo coordinate system is:

$$Z = \boldsymbol{B}\cdot\boldsymbol{K}(\alpha, \beta, \gamma, \delta), \qquad (2)$$

where $\boldsymbol{B}$ is the baseline length. From the knowledge of the vergence angles α, β, and the angular disparities γ, δ, (2) pro-

1. Defining simple computational schemes is rather important to develop a real-time vision system. This is also true if you wish to repeat the same measurements over time to exploit temporal consistency.

2. Measuring the vergence angles by optical encoders requires calibrating the angle offsets. This operation is performed off-line.

vides a depth measure with respect to the interocular baseline. As expected, the distance from the fixation point $\overline{OK}$ depends only on the vergence angles α and β; in fact:

$$Z_{fp} = \mathcal{B} \cdot \mathcal{K}(\alpha, \beta, 0, 0) = \mathcal{B} \cdot \frac{\tan(\alpha) \cdot \tan(\beta)}{\tan(\alpha) + \tan(\beta)} \tag{3}$$

A.1. Calibration

The computation of the $\mathcal{K}(\alpha, \beta, \gamma, \delta)$ function poses two relevant issues: The former is related to the calibration of the intrinsic parameters of the cameras (in our case limited to the focal lengths), the latter concerns the estimation of the angular offsets for the vergence angles α and β measured on the encoders of the motors.

By exploiting "active behaviors" it is possible to rely on self-calibration techniques which reduce the number of required parameters [23], [24]. Moreover, qualitative estimates of relative depth, like the time-to-impact, do not require the precise calibration of intrinsic parameters of the cameras. In fact, for navigation purposes it is not necessary to define a metric representation of the environment, but an approximate measurement of the *dangerous* and *safe* areas within the visual field always suffices.

In order to estimate the focal length of the cameras, we use a simple, active process: The cameras fixate on a first point in space, then they are moved to fixate a second point. The rotation of the camera, measured on the optical encoder, required to change the fixation point, is related to the displacement of the projections of the points on the image plane. Referring to Fig. 2, we obtain:

$$\tan \omega = \frac{x}{F}, \tag{4}$$

where ω is the rotation of one of the two cameras measured on the motor encoder and x is the displacement of the considered point on the image plane.

Concerning the calibration of the angle offsets, it can be performed off-line by applying standard calibration techniques [28], [29]. In this paper we have not addressed this topic explicitly, but we have compensated the angular offsets of the encoders by measuring manually the vergence angles off-line and then resetting the origin of the encoders counts.

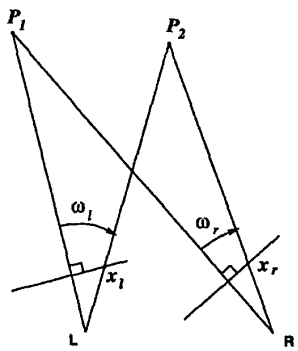


Fig. 2. Moving the fixation from point P_1 to P_2 results in two rotations ω_l and ω_r, measured by the motor encoders.

B. Motion, Optical Flow, and Time-to-Impact

The ability to quickly detect obstacles and evaluate the time to be elapsed before a collision (*time-to-impact*) is of vital importance for animates. This fact has been demonstrated by several studies on the behavior of animals [30] or humans performing specific tasks [31]. The time-to-impact can be computed from the optical flow which is extracted from monocular image sequences acquired during ego-motion. The optical flow is computed by solving an over-determined system of linear equations in the unknown terms $(u, v) = \vec{V}$ [32], [33], [34], [35], [36], [37]:

$$\frac{d}{dt} I = 0 \quad ; \quad \frac{d}{dt} \nabla I = \vec{0}, \tag{5}$$

where I represents the image intensity of the point (x, y) at time t. The instantaneous velocity for each image point can be computed by solving the linear system (5) [37].

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. The rotational part of the flow field $\vec{V}_r$ can be computed from proprioceptive data (e.g. the camera rotation) and the focal length. Once the global optic flow $\vec{V}$ is computed, $\vec{V}_t$ is determined by subtracting $\vec{V}_r$ from $\vec{V}$. From the translational optical flow, the time-to-impact can be computed:

$$T_i = \frac{\Delta_i}{|\vec{V}_t|}, \tag{6}$$

where Δ_i is the distance of the considered point (x_i, y_i), on the image plane, from the *Focus of Expansion* (FOE). The position of the FOE on the image plane can be determined by computing the pseudo intersection of the set of straight lines, obtained by elongating the optical flow vectors.

In general, the estimation of the FOE, is critical. In fact, the spread error in the least squares solution of the FOE increases as the distance of the intersection from the image center increases, because small angular errors in the velocity vectors shift the position of the computed intersection. This is the case when the camera moves along a direction considerably different from the direction of the optical axis.

We will show how the estimation of the FOE can be avoided by using stereo disparity.

III. Active Stereo and Motion Control

The measurement of the time-to-impact from stereo sequences can be faced by analyzing the temporal evolution of the image stream.

We consider the stereo system as sketched in Fig. 3. In this case, even though 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 motion depth is measured along a direction parallel to the optical axis of one (left or right) camera. The relation between the two reference frames is defined by:

$$Z(x, y) = {}^{m}Z(x, y)h(x) \;\; ; \;\; h(x) = \sin\alpha + \frac{x}{F_l}\cos\alpha, \quad (7)$$

where α is the vergence angle of the left camera, F_l is the focal length of the left camera measured in pixels, and x is the horizontal coordinate of the considered point on the image plane (see Fig. 3). We choose to adopt the stereo reference frame, because it is symmetric with respect to the cameras. In the remainder of the paper all symbols referring to the motion reference frame will be denoted by the superscript m (like ${}^{m}Z$ and ${}^{m}T$) while those referred to the symmetric stereo reference frame will be left without any further lettering (like Z and T).

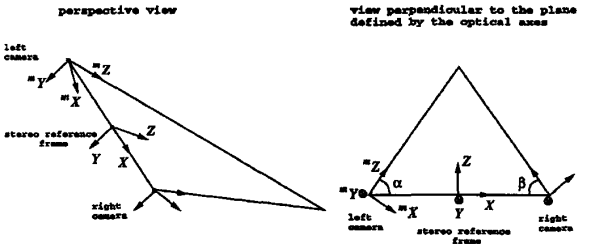


Fig. 3. Schematic representation of the stereo reference frames with their orientation.

From the results presented in the previous sections we can write for a generic point $P_i = (x_i, y_i)$ on the image plane:

$$\bigstar_i = \frac{Z_i}{\div} \;\; ; \;\; T_i = \frac{h(x_i)}{h(0)}{}^{m}T_i = \frac{Z_i}{W_z} \quad (8)$$

where $\mathcal{B}$ is the inter-ocular baseline, W_z is the velocity of the camera along the Z axis in the stereo reference frame, ${}^{m}T_i$ represents the time-to-impact measured in the motion reference frame, and T_i is the time-to-impact referred to the symmetric, stereo reference frame. In a general case, to compute the absolute distance Z_i, either from the time-to-impact or disparity, it is necessary to determine W_z or $\div$. The first parameter requires to measure the translational velocity of the cameras, while the inter-ocular baseline should be calibrated. In order to obtain a common relative-depth estimate, we first consider two different expressions derived from (8).

$${}^{m}T_i = \frac{\div}{W_z}\cdot\bigstar_i\cdot\frac{h(0)}{h(x_i)} \;\; ; \;\; \frac{Z_i}{Z_l} = \frac{{}^{m}T_i}{\bigstar_l}\cdot\frac{h(x_i)}{h(0)}\cdot\frac{W_z}{\div}. \quad (9)$$

The first equation represents the *time-to-impact* with respect to the motion reference frame, while the second equation represents a generic relative measure of the depth of a point (x_i, y_i) with respect to a second point (x_l, y_l). The first expression in (9) can be applied to a generic image point (x_j, y_j) relative to the left camera to compute the ratio $\frac{W_z}{\div}$:

$$\frac{W_z}{\div} = \frac{\bigstar_j}{{}^{m}T_j}\frac{h(0)}{h(x_j)}. \quad (10)$$

Because of ${}^{m}T_j$ at the denominator, this expression applies if the time-to-impact is not null. It is possible to obtain a better estimate of $\frac{W_z}{\div}$ by averaging the expression (10) over all the image points.

Substituting now (10) in (9):

$${}^{m}T_i = {}^{m}T_j\cdot\frac{\bigstar_i}{\bigstar_j}\cdot\frac{h(x_j)}{h(x_i)} \;\; ; \;\; \frac{Z_i}{Z_l} = \frac{\bigstar_i}{\bigstar_l}\cdot\frac{{}^{m}T_i}{{}^{m}T_j}\cdot\frac{h(x_i)}{h(x_j)}. \quad (11)$$

These 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 $\bigstar$ function). In general, pointwise ratios like (11) are not robust. Therefore, to reduce the effects of measurement errors, it is necessary to integrate measurements over a neighborhood of the considered point.

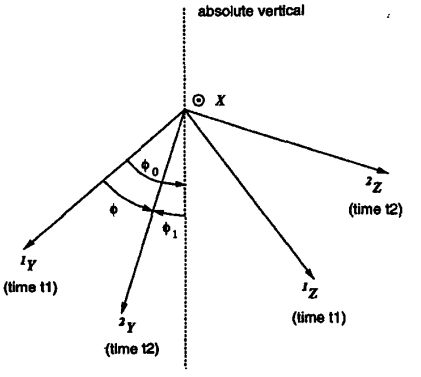


Fig. 4. Diagram showing the rotation of the stereo system during motion.

As already noted [20], [21], the critical factor in (10) and (11) is the computation of the time-to-impact which usually requires the estimation of the FOE position. To avoid this measurement it is necessary to exploit also the temporal evolution of disparity. The problem is not trivial because, considering different instants of time, the stereo reference frame changes its position and orientation while the rotational component of motion deeply affects the depth measurements performed by means of (11). Theoretically, it is possible to recover the rotational motion of the stereo cameras from visual information only, but the formulation is far too complex to allow a closed form solution [38], [28], [39]. In other words, even knowing the pan and tilt angles of the cameras, it is generally impossible to compute the *effective rotation* of the stereo system in space.

A first attempt to solve this problem has been presented in [21], where a solution was given based on two main assumptions:

1) The cameras were actively controlled to track a point in space during the robot motion,
2) The stereo rig was kept parallel to the ground plane during the motion of the vehicle.

In the remainder of the paper a more general solution is presented, where the cameras are allowed to rotate about the x axis, parallel to the ground plane, and about the y axis, perpendicular to the ground plane. This added capability is very important because it allows the binocular head to direct the gaze everywhere in space.

A. Estimation of Camera Rotation

The motion of the stereo system between two different time instants t_1 and t_2 can be expressed as a generic roto-translation:

$$^2\vec{P} = \mathbf{R} \cdot {}^1\vec{P} + {}^2\vec{T} \tag{12}$$

where $^1\vec{P} = (^1X, {}^1Y, {}^1Z)$ and $^2\vec{P} = (^2X, {}^2Y, {}^2Z)$ are the coordinates of the same point in space measured at time t_1 and t_2, $\mathbf{R}$ is the rotation matrix, and $^2\vec{T} = (^2T_X, {}^2T_Y, {}^2T_Z)$ is the translation vector. Therefore, to compute the 3D motion in the general case, it is necessary to solve a linear system with 12 unknowns (which can be reduced to six, including three rotational angles and three translational displacements), and at least three different points are needed [28], [40].

As we are interested in recovering the rotational motion only, it is possible to avoid the computation of the translational vector $\vec{T}$ by subtracting the expressions for four different points in space. Considering four points P_a, P_b, P_c, and P_d at two time instants t_1 and t_2, we obtain:

$$^2\vec{P}_a - {}^2\vec{P}_i = \mathbf{R} \cdot \left({}^1\vec{P}_a - {}^1\vec{P}_i\right) \;;\; i = b, c, d, \tag{13}$$

which is a linear system of nine equations in nine unknowns.

Let us consider the general case, defined by (12), and suppose that the stereo reference frame rotates during motion around two different axes: the absolute vertical axis, perpendicular to the ground plane, and the x axis, parallel to the ground plane.[3] The rotation matrix $\mathbf{R}$ can be written as the composition of three elementary rotations:

$$^2\vec{P} = \mathbf{R}_x(\phi_1) \cdot \mathbf{R}_y(\theta) \cdot \mathbf{R}_x(\phi_0) \cdot {}^1\vec{P} + {}^2\vec{T}, \tag{14}$$

where ϕ_0 is the rotation angle between the y axis of the stereo system and the absolute vertical, and $\phi = \phi_0 + \phi_1$ is the effective rotation around the x axis. The angle θ is the only *real* unknown in our case: ϕ_0 and ϕ_1 are measured from the motor encoders, but θ can not be measured because it is related to the rotation of the robot carrying the stereo system and not of the cameras. Computing explicitly the matrix $\mathbf{R}$ we obtain:

$$\mathbf{R} = \begin{bmatrix} \cos\theta & \sin\phi_0 \sin\theta & -\cos\phi_0 \sin\theta \\ \sin\phi_1 \sin\theta & \cos\phi_0 \cos\phi_1 - \sin\phi_0 \sin\phi_1 \cos\theta & \sin\phi_0 \cos\phi_1 + \cos\phi_0 \sin\phi_1 \cos\theta \\ \cos\phi_1 \sin\theta & -\cos\phi_0 \sin\phi_1 - \sin\phi_0 \cos\phi_1 \cos\theta & -\sin\phi_0 \sin\phi_1 + \cos\phi_0 \cos\phi_1 \cos\theta \end{bmatrix}. \tag{15}$$

Let us consider two points $\vec{P}_h$ and $\vec{P}_k$:

$$\begin{cases} {}^2\vec{P}_h = \mathbf{R}\, {}^1\vec{P}_h + {}^2\vec{T} \\ {}^2\vec{P}_k = \mathbf{R}\, {}^1\vec{P}_k + {}^2\vec{T} \end{cases}, \tag{16}$$

by applying (13) and denoting by $\Delta^1\vec{P} = (\Delta^1X, \Delta^1Y, \Delta^1Z) = {}^1\vec{P}_h - {}^1\vec{P}_k$ and $\Delta^2\vec{P} = (\Delta^2X, \Delta^2Y, \Delta^2Z) = {}^2\vec{P}_h - {}^2\vec{P}_k$ we obtain:

$$\begin{cases} \Delta^2X = \Delta^1X\cos\theta + \Delta^1Y\sin\phi_0\sin\theta + \\ \qquad -\Delta^1Z\cos\phi_0\sin\theta \\ \Delta^2Z = \Delta^1X\cos\phi_1\sin\theta + \\ \qquad -\Delta^1Y(\cos\phi_0\sin\phi_1 + \sin\phi_0\cos\phi_1\cos\theta) + \\ \qquad -\Delta^1Z(\sin\phi_0\sin\phi_1 - \cos\phi_0\cos\phi_1\cos\theta) \end{cases}, \tag{17}$$

posing $a = \sin\phi_0$, $b = \cos\phi_0$, $c = \sin\phi_1$, $d = \cos\phi_1$, dividing the two expressions and solving for $\tan\theta$:

$$\tan\theta = \frac{\left(\Delta^2Z + bc\Delta^1Y + ac\Delta^1Z\right)\Delta^1X - \left(bd\Delta^1Z - ad\Delta^1Y\right)\Delta^2X}{\left(b\Delta^1Z - a\Delta^1Y\right)\left(\Delta^2Z + bc\Delta^1Y + ac\Delta^1Z\right) + d\Delta^1X\Delta^2X}. \tag{18}$$

It is important to note that the factors 'Δ^i' depend on the length of the baseline $\div$. In fact, according to the diagram in Fig. 3:

$$\begin{cases} X = \div \cdot \left(\dfrac{\star}{\tan(\alpha - \gamma)}\right) - \dfrac{\div}{2} \\ Y = \div \cdot \dfrac{y}{F_i} \cdot \star \cdot \dfrac{\cos\gamma}{\sin(\alpha - \gamma)} \\ Z = \div \cdot \star \end{cases}. \tag{19}$$

But, as $\div$ is a proportionality factor for the coordinates (X, Y, Z), then (18) does not depend on the stereo baseline $\div$.

If the optical axes of the cameras are maintained parallel to the ground plane during the robot motion (as in [21]), we simply obtain $a = c = 0$ and $b = d = 1$. In this case:

$$\tan\theta = \frac{\Delta^1X\Delta^2Z - \Delta^2X\Delta^1Z}{\Delta^1X\Delta^2X + \Delta^1Z\Delta^2Z}. \tag{20}$$

As a consistency verification we can note that if the stereo system does not rotate during vehicle motion $\Delta^1X = \Delta^2X$ and $\Delta^1Z = \Delta^2Z$, and $\theta = 0$.

The rotation angle θ can be used to eliminate the rotational effect from the depth estimate. From (14) applied to a generic point P:

$$Z_t = {}^2Z - {}^2T_Z = d^1X\sin\theta - (bc + ad\cos\theta)^1Y - (ac - bd\cos\theta)\,{}^1Z, \tag{21}$$

where Z_t represents the distance ^1Z of the considered point projected along the direction of ^2Z. Dividing both sides of (21) by $\div$ and applying the first expression in (8):

$$\star_t = \frac{^1X}{\div} d\sin\theta - (bc + ad\cos\theta)\frac{^1Y}{\div} - (ac - bd\cos\theta)\frac{^1Z}{\div}, \tag{22}$$

which is an expression of the $\star$ function projected on the translational path. Also, in this case, if the cameras do not tilt, we obtain simpler expressions:

$$Z_t = {}^1X\cos\theta + {}^1Z\cos\theta = \left[\left(\frac{^1\star}{\tan(^1\alpha - {}^1\gamma)} - \frac{1}{2}\right)\sin\theta + {}^1\star\cos\theta\right] \cdot \div \tag{23}$$

3. This is equivalent to consider a moving vehicle equipped with a stereo camera system where the baseline is kept parallel to the ground plane, while the stereo reference frame can freely perform pan and tilt rotations. Only *roll* is not included in the allowed movements.

$$\star_t = \left[\left(\frac{{}^1\star}{\tan({}^1\alpha - {}^1\gamma)} - \frac{1}{2}\right)\sin\theta + {}^1\star\cos\theta\right]. \quad (24)$$

If the baseline of the cameras does not rotate $Z_t = {}^1Z$ and $K_t = {}^1K$. In the remainder of the paper we will generically denote as Z and K the distance Z_t and the function K_t projected along the translational path.

B. Using the Temporal Evolution of Disparity

Considering a general motion of the stereo system with uniform velocity, we obtain:

$$\begin{aligned} Z_i(t-\Delta t) - Z_i(t) &= W_z\,\Delta t \\ \star_i(t-\Delta t) - \star_i(t) &= \frac{W_z}{\dotplus}\Delta t, \end{aligned} \quad (25)$$

where Z_i and $\mathcal{K}_i$ are the corresponding measurements, for a generic point (x_i, y_i), projected along the translational path. This is a very interesting expression showing that K is a linear function of time, but also it can be used to estimate the unknown parameters W_z and B. Moreover, given the factor $\frac{W_z}{\dotplus}$, it is possible to make a prediction of the values of disparity over time to facilitate the matching process. 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 $\frac{W_z}{\dotplus}$ is a constant factor for a given stereo image pair, it is possible to compute a robust estimate by taking the average over a neighborhood [20]:

$$\frac{W_z}{\dotplus} = \Delta_\star = \frac{1}{\Delta t\, N^2}\sum_i\left[\star_i(t-\Delta t) - \star_i(t)\right]. \quad (26)$$

Given the optical flow $\vec{V} = (u, v)$ and the map of the values of the $\mathcal{K}$ function at time $t - \Delta t$, the value of $\mathcal{K}_i(t)$ is obtained by considering the image point $(x_i + u_i, y_i + v_i)$ on the map at time t. This expression reminds the temporal evolution of disparity formulated in [18]. The basic difference between the two approaches is the fact that Waxman and Duncan analyzed a stereo set-up with parallel optical axes, while in this paper the cameras have convergent optical axes. On the other hand, the expression developed in [18] is related to binocular image flows, which are used to establish stereo correspondence, while in our approach a monocular optical flow is used, together with stereo disparity, to apply (26) and then compute the time-to-impact.

Combining (9) and (26) we obtain:

$$T_i = \frac{h(x_i)}{h(0)}\,{}^mT_i = \frac{\star_i}{\Delta_\star} \quad (27)$$

$$\frac{Z_i}{Z_l} = \frac{\Delta_\star}{\star_l}\cdot\frac{h(x_i)}{h(0)}\,{}^mT_i. \quad (28)$$

The estimate (27) of the time-to-impact is very robust and does not require the computation of the FOE. The value of $\frac{h(x_i)}{h(0)}$ in (28) can be easily computed from the pixel position, the focal length of the camera and the vergence angle of the camera on which the optical flow has been computed:

$$\frac{h(x_i)}{h(0)} = 1 + \frac{x_i}{F_l\tan\alpha}. \quad (29)$$

In summary, in order to compute the time-to-impact or relative depth, the following quantities need to be computed or measured:

- the stereo disparity field at two time instants;
- monocular optical flow field from an image sequence acquired from the left camera;
- the vergence angles of the cameras (measured from the optical encoders of the motors);
- the focal length of the cameras, or the conversion factor between linear and angular displacements on the image plane.

C. Sensitivity Analysis in the Computation of the Time-to-Impact

It is beyond the aim of this paper to perform an exhaustive error analysis for all the equations presented, but it is interesting to analyze the sensitivity of (27) with respect to noise. The variation of T in relation to $\star$ can be expressed by differentiating (27):

$$\left|\frac{dT_i}{d\star_i}\right| = \left|\frac{\partial T_i}{\partial\Delta_\star}\cdot\frac{d\Delta_\star}{d\star_i}\right| + \left|\frac{\partial T_i}{\partial\star_i}\right|. \quad (30)$$

From (26) it is possible to observe that Δ_χ is a function of $\mathcal{K}_i(t)$ only. Therefore, computing explicitly the total derivative $\frac{d\Delta_\star}{d\star_i} = -\frac{1}{N^2\Delta t}$ and the partial derivatives of T_i, and substituting in (30):

$$\left|\frac{dT_i}{d\star_i}\right| = \left|\frac{\star_i}{\Delta_\star^2}\cdot\frac{1}{N^2\Delta t}\right| + \left|\frac{1}{\Delta_\star}\right| = \left|\frac{1}{N^2\Delta t}\cdot\frac{T_i}{\Delta_\star}\right| + \left|\frac{1}{\Delta_\star}\right|, \quad (31)$$

dividing both sides by T_i and taking the differential δT_i of T_i, we obtain:

$$\left|\frac{\delta T_i}{T_i}\right| = \left|\frac{1}{N^2\Delta t}\cdot\frac{\delta\star_i}{\Delta_\star}\right| + \left|\frac{\delta\star_i}{T_i\Delta_\star}\right| = \left|\frac{1}{N^2\Delta t}\cdot\frac{\delta\star_i}{\Delta_\star}\right| + \left|\frac{\delta\star_i}{\star_i}\right|. \quad (32)$$

If the value of Δ_K is obtained using a sufficient number N of image points, then the term $\left|\frac{1}{N^2\Delta t}\cdot\frac{\delta\star_i}{\Delta_\star}\right|$ is negligible. Therefore, the relative error on the time-to-impact is equal to the relative error on the disparity function $\mathcal{K}$. This result can be compared with the error in computing the time-to-impact from optical flow. In this case the relative error can be computed from (6):

$$\left|\frac{dT_i}{dV_t}\right| = \left|\frac{\partial T_i}{\partial\Delta_i}\cdot\frac{d\Delta_i}{dV_t}\right| + \left|\frac{\partial T_i}{\partial V_t}\right|, \quad (33)$$

where $V_t = |\vec{V}_t|$ is the amplitude of the translational component of the velocity vector $\vec{V}(x_i, y_i)$ and Δ_i is the distance of the

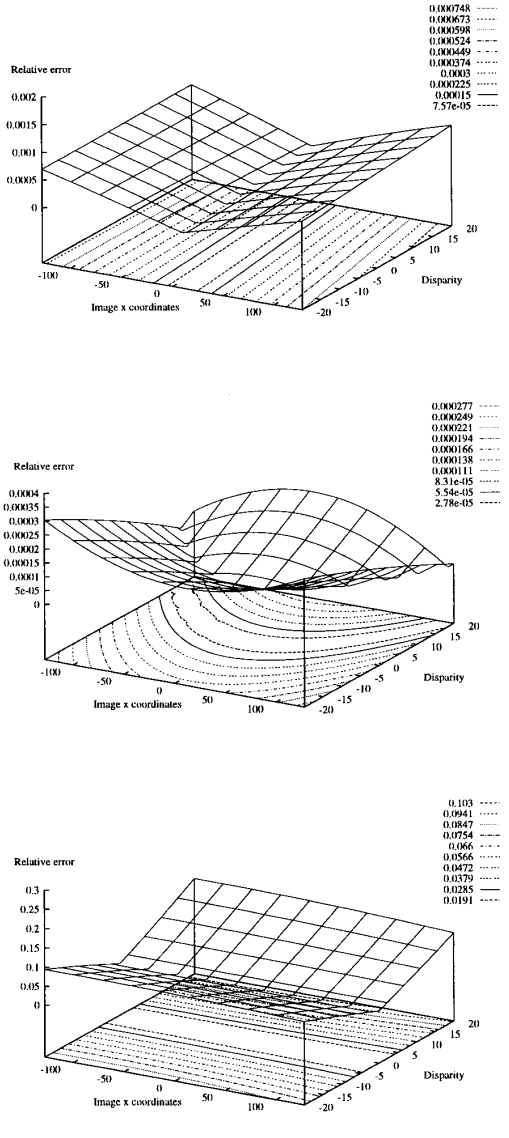


Fig. 5. Graphs showing the relative variation of $\mathcal{K}$ with respect to the pixel coordinates and image disparity. a) Variation of the K function in relation to the error δx of the image x coordinate. b) Variation of the K function in relation to the error δF in the focal length F. c) Variation of the K function in relation to the relative error $\frac{\delta D}{D}$ in the computed image disparity D.

considered point (x_i, y_i) from the FOE. By developing this expression and taking the differential δT_i of T_i, we obtain:

$$\left|\delta T_i\right| = \left|\frac{1}{V_t} \cdot \delta\Delta_i\right| + \left|\frac{\Delta_i}{V_t^2} \cdot \delta V_t\right|, \tag{34}$$

where $\delta\Delta_i$ represents the spread error of the least squares solution for the location of the FOE, which is obtained as the pseudo intersection of the directions of the optical flow vectors. Dividing both sides by T_i and substituting (6):

$$\left|\frac{\delta T_i}{T_i}\right| = \left|\frac{\delta\Delta_i}{\Delta_i}\right| + \left|\frac{\delta V_t}{V_t}\right|. \tag{35}$$

This relation is very similar to (32) but the two expressions have different meanings. The second addendum in (35) constitutes the relative error in the computation of the image displacement field, while the second addendum in (32) is the relative error of the $\star$ function which is not a direct measurement of the displacement on the image, but rather the distance with respect to the stereo reference frame. Consequently, while the error (35) strongly depends on the image coordinates and, in particular, presents a singularity close to the focus of expansion, the error (32) is quite flat as a function of the image coordinates. In fact, by differentiating (1) with respect to the x image coordinate and dividing by $\star$:

$$\left|\frac{\delta\star}{\star}\right| = \left|\frac{-\frac{F}{(x^2+F^2)}\frac{\tan^2(\beta+\delta)}{\cos^2(\alpha-\gamma)} + \frac{F}{(x^2+F^2+D^2+2xD)}\frac{\tan^2(\alpha-\gamma)}{\cos^2(\beta+\delta)}}{[\tan(\alpha-\gamma)+\tan(\beta+\delta)]\tan(\alpha-\gamma)\tan(\beta+\delta)}\right| |\delta x|, \tag{36}$$

where x is the image coordinate, D is the image disparity measured along the x coordinate, F is the focal length of the camera, $\gamma = \arctan(\frac{x}{F})$, and $\delta = \arctan(\frac{x+D}{F})$. Similar expressions are obtained by differentiating $\mathcal{K}$ with respect to the other variables related to the image: the disparity D and the focal length F. These expressions have been plotted by assuming unitary errors for δx, δF, and δD and using values for the other parameters which are standard in our set-up ($F = 600$ *pixels* and $\alpha = \beta = 1.4$ radians). The resulting graphs are shown in Fig. 5. It is worth noting that each expression of the error is quite flat as a function of the pixel coordinates and also, from the level curves of the function, it can be noticed that the amplitude of the error terms is quite low.

By adding each expression the global relative error results:

$$\left|\frac{\delta\star}{\star}\right| = \left|\frac{-\frac{F}{(x^2+F^2)}\frac{\tan^2(\beta+\delta)}{\cos^2(\alpha-\gamma)} + \frac{F}{(x^2+F^2+D^2+2xD)}\frac{\tan^2(\alpha-\gamma)}{\cos^2(\beta+\delta)}}{[\tan(\alpha-\gamma)+\tan(\beta+\delta)]\tan(\alpha-\gamma)\tan(\beta+\delta)}\right| |\delta x| + \left|\frac{\frac{x}{(x^2+F^2)}\frac{\tan^2(\beta+\delta)}{\cos^2(\alpha-\gamma)} - \frac{x+D}{(x^2+F^2+D^2+2xD)}\frac{\tan^2(\alpha-\gamma)}{\cos^2(\beta+\delta)}}{[\tan(\alpha-\gamma)+\tan(\beta+\delta)]\tan(\alpha-\gamma)\tan(\beta+\delta)}\right| |\delta F| + \left|\frac{\frac{FD}{(x^2+F^2)}\frac{\tan^2(\alpha-\gamma)}{\cos^2(\beta+\delta)}}{[\tan(\alpha-\gamma)+\tan(\beta+\delta)]\tan(\alpha-\gamma)\tan(\beta+\delta)}\right| \frac{|\delta D|}{D}. \tag{37}$$

This error function is clearly bounded, because each of the three addenda is bounded, and also varies smoothly with the image coordinates and disparity. Therefore, the measurement of $\mathcal{K}$, and from (32) also the time-to-impact, degrade gracefully with increasing errors in the image variables. The behavior of the relative error in the measurement of the image displacement $\left|\frac{\delta V_t}{V_t}\right|$ as well as the relative error in the image disparity $\left|\frac{\delta D}{D}\right|$ depend on the computational schema applied. On the other hand, the computation of $\vec{V}_t$ requires to compute and differentiate the rotational component of the optical flow. In this paper a differential technique has been applied for the computation of the optical flow, because allows to estimate a dense and accurate flow field almost everywhere on the image plane, assumed the motion to be small. However, as the image velocity becomes close to zero, the error term $\left|\frac{\delta V_t}{V_t}\right|$ grows quite quickly, consequently, the measurements of the time-to-impact from the optical flow only can be unstable and sensitive to high frequency noise. This is not the case for the error term $\left|\frac{\delta D}{D}\right|$ because rather large disparities are computed by applying a correlation technique.

The first addendum in (35) represents the relative error in the computation of the FOE. This measurement is generally more robust than the local computation of the image velocity itself, because it is obtained by integrating several velocity estimates on the image plane. However, due to the sensitivity to noise of the velocity estimation, the accuracy in the localization of the FOE is poor, in particular close to the FOE, while the term $\delta\Delta_i$ becomes very large, of the order of tens of pixels. Also the error term $\left|\frac{\delta\Delta_i}{\Delta_i}\right|$ becomes very large computing the time-to-impact close to the FOE. In conclusion, the time-to-impact computed from the optical flow only can be unstable and very sensitive to noise wherever the optical flow is small and, in particular, presents a singularity in the FOE. The accuracy in the computation of the time-to-impact degrades quite quickly as the FOE departs from the image plane. On the contrary, the absolute error computing the time-to-impact from the dynamic stereo approach can be larger than using the optical flow only, but it is more robust and stable with respect to noise.

IV. Experimental Results

A. Planar Tracking

In the first experiment a computer-controlled mobile platform TRC *Labmate* with two stereo cameras has been used. The cameras were arranged as to verge toward a point in space. A sequence of stereo images has been captured during a tracking motion of the cameras and keeping the stereo rig parallel to the ground plane. In Fig. 6 the first and last stereo pair, from a sequence of 11, are shown. The disparity map computed from the sixth stereo pair is shown in Fig. 7. The vehicle was moving forward about 100 mm per frame. The sequence has been taken inside the LIRA lab, with many objects 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 which can be seen on the desk, in the foreground. The value of Δ_x for the sixth stereo pair has been computed by applying (26) at each image point. By taking the average of the values of Δ_x over all the image points, a value of $\frac{W_x}{f}$ equal to 0.23 has been obtained. This value must be compared to the ground truth equal to 0.29, computed from the velocity of the vehicle, which was about 100 millimeters per frame along the Z axis, and the inter-ocular 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, also in this case, the given values of the velocity and baseline are slightly wrong.

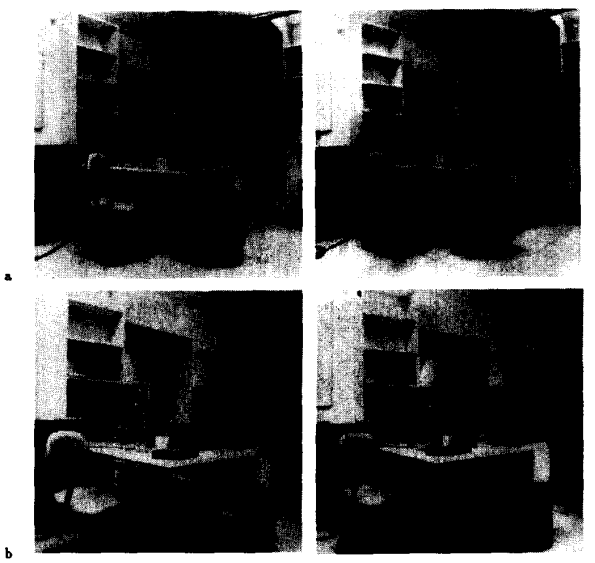


Fig. 6. First (a) and last (b) stereo image pair of the sequence.

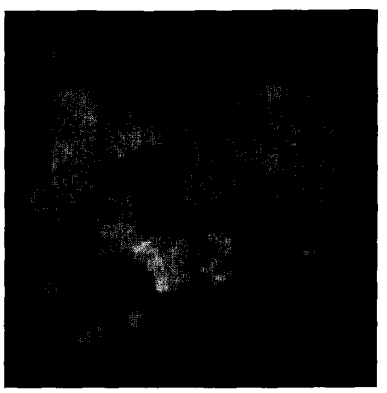


Fig. 7. Map of the values of the K function obtained from the sixth stereo pair of the sequence.

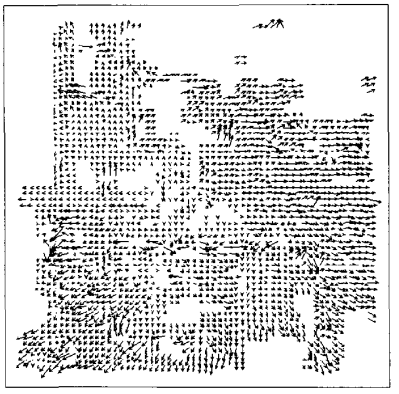


Fig. 8. Optical flow relative to the sixth left image of the sequence.

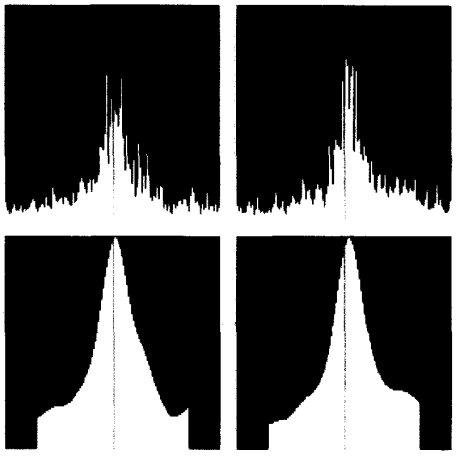


Fig. 9. Rough and smoothed histograms of the angles computed from frames 5-6 and frames 7-8, respectively. The abscissa scale goes from –0.16 radians to 0.16 radians. The maxima computed in the smoothed histograms corresponds to 0.00625 and 0.0075 radians, respectively.

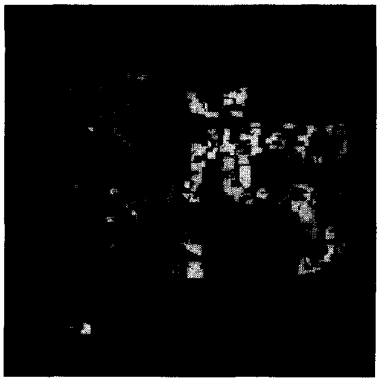


Fig. 10. Time-to-impact computed using (27) for the sixth pair of the sequence; darker regions correspond to closer objects.

In this experiment the cameras were moved keeping the optical axes on a plane parallel to the ground plane, therefore only one rotation angle is involved during tracking. The rotation angle of the stereo baseline between two successive time instants has been computed by applying (20) to all the image points. In the noiseless case all the image points will produce identical estimates. In a real scene, like in this experiment, it is necessary to discard all the wrong angular values coming from errors and loss of accuracy in the computation. This is obtained by ranking the angular values on an histogram and considering as correct the angular value corresponding to the peak in the histogram. In the case of objects moving within the field of view it should be easy to separate the different peaks corresponding to moving and still objects. In Fig. 9 two histograms related to frames 5-6 and 7-8, respectively, are shown. The computed camera rotation is 0.00625 radians, corresponding to about 0.36 degrees.

In Fig. 8 the optical flow of the sixth left image of the sequence is shown. From the optical flow and the values of the K function the map of the time-to-impact has been computed by applying (27). The map is shown in Fig. 10; the values of the time-to-impact are coded as gray levels, darker meaning lower time-to-impact.

B. General Tracking

In the first experiment the cameras were kept fixed, while an object was rotating on a turntable at a speed of two degrees per frame. The cameras were arranged as to verge toward the center of the turntable. In this way a tracking motion was simulated. In Fig. 11 the 10th and 20th stereo pair, from a sequence of 24, are shown. The images are 256 × 256 pixels with 8 bits of resolution in intensity.

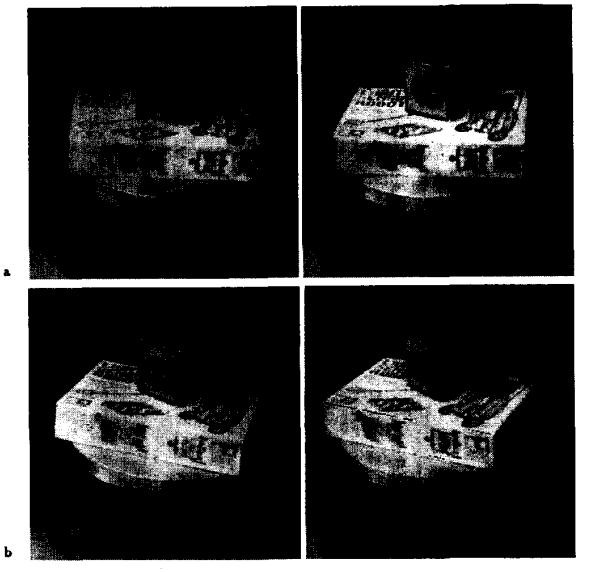


Fig. 11. Tenth (a) and 20th (b) stereo image pair of the sequence.

The captured sequence has been used to perform two experiments: In the former we took the position of 10 relevant

points (like corners), directly from the original images (by hand using a pointing device); in the latter we computed the optical flow and disparity maps from two successive image pairs, and used the output to compute the rotational angle and the time-to-impact.

For the first experiment the 10th and 20th image pairs of the sequence have been used, which correspond to a rotation of the object of 20° circa.[4] The images are shown in Fig. 11. The coordinates of the points, taken by hand, were fed to (18). Using more data than equations, we simply computed the average of the angles resulting by considering all the data points. Of course a least square method would have been more appropriate to find a better solution. The angular value obtained is 18.6° , which must be compared to the measured rotation of 20° . It is worth noting that, by locating the position of the points by hand, we introduced an error of at least one pixel in the localization of the points.

In the second experiment two successive image pairs were used, corresponding to a rotation of two degrees circa. The disparity maps were computed for both image pairs and the optical flow for the left image of the first frame. The computed camera rotation (around the vertical axis) is 0.0427 radians, corresponding to about 2.4° . The rotation recovered from the optical flow and disparity with (18), has been used to correct the K values as from (22) and compute the time-to-impact by applying (27). The map in Fig. 12 codes the values of the time-to-impact as gray levels, darker means smaller time-to-impact (or an object closer to the cameras).

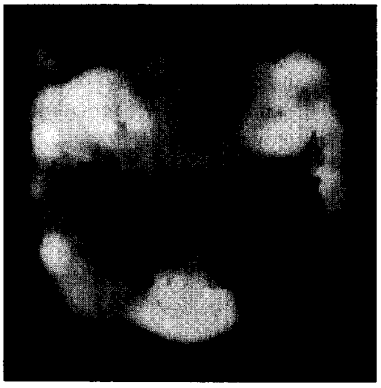


Fig. 12. Time-to-impact computed using (27) for the 15th pair of the sequence; darker regions correspond to closer objects.

A different experiment was performed by processing a sequence of stereo images acquired from a prototype robotic head, built at the LIRA-Lab. A picture and a schematic diagram of the head with the cameras is shown in Fig. 13. The head is composed of four independent degrees of freedom. The "neck" rotation and the common tilt of the two cameras is controlled by two DC torque motors. The independent vergence of each camera is controlled by two stepper motors.

4. This value has been measured approximately during the image acquisition phase.

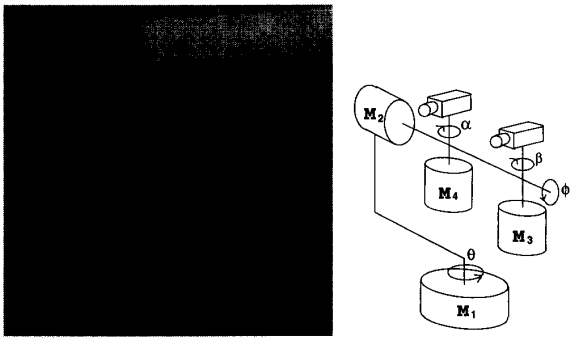


Fig. 13. Picture and diagram of the robotic head used to acquire the images used in one experiment.

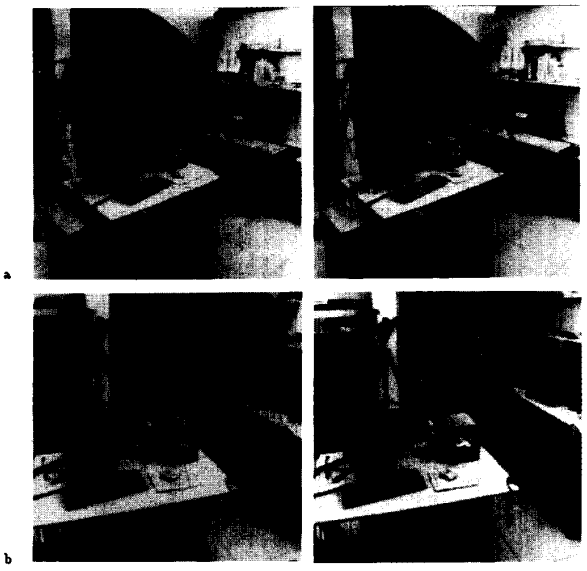


Fig. 14. First (a) and last (b) stereo image pair of the sequence acquired from the head.

The head was placed on a wheeled cabinet. The gaze and vergence of the head were automatically adjusted to compensate for the motion of the cabinet, keeping the fixation on the same point in space. The forward motion was a translation of about 7 cm per each step. At each step a pair of stereo images was captured, at a resolution of 256 × 256 pixels with 8 bits per pixel. In Fig. 14, the first and last image pair from a sequence of 16 is shown. The vergence, pan and tilt angles of the cameras were recorded from the optical encoders of the head motors. It is worth noting that only the degrees of freedom under active control were effectively recorded, while the rotation of the cabinet was completely unknown and without precise control. The recorded rotations were used to compute the effective rotation of the camera system from one frame to the following. The procedure adopted to compute the effective rotation and the time-to-impact, is the same followed in the previous experiment. The disparity maps were computed for

the sixth and seventh image pairs and the optical flow for the left image of the sixth frame. In Figs. 15 and 16 the computed map of the $\star$ function and optical flow relative to the sixth frame, are shown. As in the previous experiment, we computed the average from all the resulting angles. The obtained camera rotation (around the vertical axis) is 0.014 radians, corresponding to about 0.8 degrees.

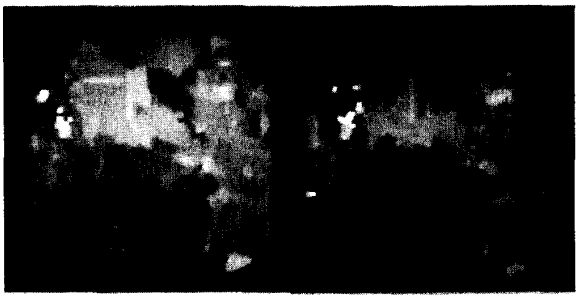


Fig. 15. Map of the values of the K function obtained from the sixth and seventh stereo pair of the sequence.

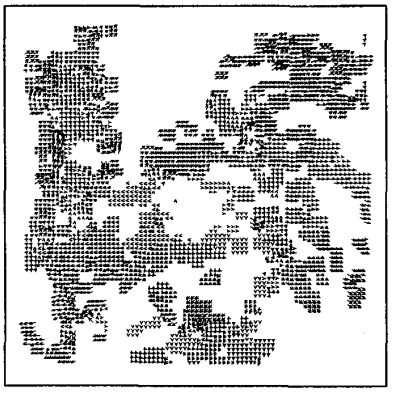


Fig. 16. Optical flow relative to the sixth left image of the sequence.

Fig. 17. Time-to-impact computed using (27) for the sixth pair of the sequence; darker regions correspond to closer objects.

The rotation recovered from the optical flow and disparity has been used to correct the K values and compute the time-to-impact. The map in Fig. 17 codes the values of the time-to-impact as gray levels. As it can be noticed, the time-to-impact map closely reflects the structure of the environment. The lamp on the desk, which can be hardly seen in the original images, appears clearly closer than the background, while the desk in the foreground has the lowest time-to-impact values.

V. Conclusion

In this paper we have addressed the problem of the extraction of relevant visual information for robot operations.

Whenever a spatial goal has to be reached, either by animals or robots, it is important to be able to decide the direction of motion. This decision is crucial in visual navigation because can be taken on the basis of visual information only. It is necessary to identify free space areas in the scene which can be safely crossed by the robot from its current position in space. Stereo vision and motion parallax have been considered as cues to identify corridors of free space.

Redundancy is enforced by defining a cooperative schema in which both stereo and motion provide information to estimate a relative-depth map of the observed scene. In the cooperation process binocular disparity, computed on several image pairs over time, is *merged* with optical flows to cope for the need of critical parameters relative to the cameras and/or the robot. An important aspect is certainly the possibility of computing dynamic quantities, like the time-to-impact, directly from stereo disparity, using the optical flow to determine the temporal evolution of disparity in the image sequence.

One of the advantages of the proposed approach is the possibility to compute the effective rotation of the stereo camera system, even in presence of unknown rotations of the moving vehicle. In fact, the only angular displacements which have to be measured, are the rotations performed by the motors of the head-eye system. Therefore, the system's behavior is closely related to the capability of visual stabilization of a target. This is consistent with the need to stabilize an image in order to compute meaningful data relative to the environment and/or the observer. In this sense the system's activity helps for motion and structure recovery.

The measured parameters are closely coupled together and strongly depend on the behaviour of the system. In fact, not only the recovered rotation is necessary to compute the time-to-impact map, but also it is naturally referred to the imaging system and not to the motor system. As long as the system keeps tracking the same point in space, or it does not performs saccadic movements, successive measurements are perfectly coherent and can be integrated over time to enforce the robustness of the recovered rotation and time-to-impact.

Acknowledgments

This work has been partially funded by the ESPRIT Basic Research Action projects P3274 FIRST and P6769 SECOND and by grants of the Italian National Research Council.

We thank Prof. G. Sandini for his helpful comments and insights during the development of this work.

REFERENCES

[1] R.D. Beer, *Intelligence as Adaptive Behavior*. Academic Press, 1990.

[2] R.A. Brooks, "A robust layered control system for a mobile robot," *IEEE Trans. Robotics and Automation*, vol. 2, pp. 14–23, Apr. 1986.

[3] F. Ferrari, E. Grosso, M. Magrassi, and G. Sandini, "A stereo vision system for real time obstacle avoidance in unknown environment," *Proc. Int'l Workshop Intelligent Robots and Systems*, Tokyo, July 1990.

[4] G. Sandini and M. Tistarelli, "Robust obstacle detection using optical flow," *Proc. IEEE Int'l Workshop Robust Computer Vision*, pp. 396–411, Seattle, Oct. 1-3, 1990.

[5] M. Tistarelli and G. Sandini, "Dynamic aspects in active vision," *CVGIP*, Special issue on Purposive and Qualitative Active Vision, Y. Aloimonos, ed., vol. 56, no. 1, pp. 108–129, July 1992.

[6] J. Aloimonos, I. Weiss, and A. Bandyopadhyay, "Active vision," *Int'l J. Computer Vision*, vol. 1, no. 4, pp. 333–356, 1988.

[7] G. Sandini and M. Tistarelli, "Active tracking strategy for monocular depth inference over multiple frames," *IEEE Trans. Pattern Aanalysis and Machine Intelligence*, vol. 12, no. 1, pp. 13-27, Jan. 1990.

[8] R.K. Bajcsy, "Active perception vs passive perception," *Proc. Third IEEE CS Workshop Computer Vision: Representation and Control*, pp. 13–16, Bellaire, Mich., 1985.

[9] D.H. Ballard, "Animate vision," *Artificial Intelligence*, vol. 48, pp. 57–86, 1991.

[10] D.H. Ballard and C.M. Brown, "Principles of animate vision," *CVGIP*, Special issue on Purposive and Qualitative Active Vision, Y. Aloimonos, ed., vol. 56, no. 1, pp. 3–21, July 1992.

[11] N.J. Bridwell and T.S. Huang, "A discrete spatial representation for lateral motion stereo," *CVGIP*, vol. 21, pp. 33–57, 1983.

[12] N. Ayache and O.D. Faugeras, "Maintaining representations of the environment of a mobile robot," *IEEE Trans. Robotics and Automation*, vol. 5, no. 6, pp. 804–819, Dec. 1989.

[13] D.J. Kriegman, E. Triendl, and T.O. Binford, "Stereo vision and navigation in buildings for mobile robots," *IEEE Trans. Robotics and Automation*, vol. 5, no. 6, pp. 792–803, Dec. 1989.

[14] L. Matthies, T. Kanade, and R. Szeliski, "Kalman filter-based algorithms for estimating depth from image sequences," *Int'l J. Computer Vision*, vol. 3, no. 3, pp. 209–238, 1989.

[15] N. Ahuja and L. Abbott, "Surfaces from dynamic stereo: Integrating camera vergence, focus, and calibration with stereo surface reconstruction," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1,007-1,029, Oct. 1993.

[16] E. Grosso, G. Sandini, and M. Tistarelli, "3D object reconstruction using stereo and motion," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 19, no. 6, Nov./Dec. 1989.

[17] R.A. Brooks, A.M. Flynn, and T. Marill, "Self calibration of motion and stereo vision for mobile robot navigation," *Proc. DARPA Workshop Image Understanding*, pp. 398–410, Morgan and Kaufman, eds., 1988.

[18] A.M. Waxman and J.H. Duncan, "Binocular image flows: Steps toward stereo-motion fusion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 715–729, 1986.

[19] L. Li and J.H. Duncan, "3D translational motion and structure from binocular image flows," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 7, pp. 657–667, July 1993.

[20] M. Tistarelli, E. Grosso, and G. Sandini, "Dynamic stereo in visual navigation," *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, pp. 186-193, Lahaina, Hawaii, June 1991.

[21] E. Grosso, M. Tistarelli, and G. Sandini, "Active-dynamic stereo for navigation," *Proc. Second European Conf. Computer Vision*, pp. 516–525, S. Margherita Ligure, Italy, May 1992.

[22] A. Izaguirre, P. Pu, and J. Summers, "A new development in camera calibration: Calibrating a pair of mobile cameras," *The Int'l J. Robotics Research*, pp. 104–116, 1988.

[23] Y.L. Chang, P. Liang, and S. Hackwood, "Adaptive self-calibration of vision-based robot systems," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 19, no. 4, Jul./Aug. 1989.

[24] A.P. Tirumalai, B.G. Schunck, and R.C. Jain, "Dynamic stereo with self-calibration," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 12, pp. 1,184–1,189, Dec. 1992.

[25] L. Matthies and T. Kanade, "Using uncertainty models in visual motion and depth estimation," *Proc. Fourth Int'l Symp. Robotics Research*, pp. 120–138, Santa Cruz, Calif., Aug. 9-14, 1987.

[26] R.C. Nelson and J. Aloimonos, "Using flow field divergence for obstacle avoidance in visual navigation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 10, pp. 1,102-1,106, Oct. 1989.

[27] D.H. Ballard, R.C. Nelson, and B. Yamauchi, "Animate vision," *Optics News*, vol. 15, no. 5, pp. 17–25, 1989.

[28] B.K.P. Horn, *Robot Vision*. Cambridge, Mass.: MIT Press, 1986.

[29] P. Puget and T. Skordas, "Calibrating a mobile camera," *Image and Vision Computing*, vol. 8, no. 4, pp. 341–348, 1990.

[30] D.N. Lee and P E. Reddish, "Plummeting gannets: A paradigm of ecological optics," *Nature*, vol. 293, pp. 293–294, 1981.

[31] J. Cutting, *Perception with an Eye for Motion*. Cambridge, Mass.: MIT Press, 1988.

[32] B.K.P. Horn and B.G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–204, 1981.

[33] H.H. Nagel, "Direct estimation of optical flow and of its derivatives," *Artificial and Biological Vision Systems*, G.A. Orban and H.H. Nagel, eds., pp. 193–224, Springer Verlag, 1992.

[34] H.H. Nagel, "On the estimation of optical flow: Relations between differenet approaches and some new results," *Artificial Intelligence*, vol. 33, pp. 299–324, 1987.

[35] S. Uras, F. Girosi, A. Verri, and V. Torre, "A computational approach to motion perception," *Biological Cybernetics*, vol. 60, pp. 79–87, 1988.

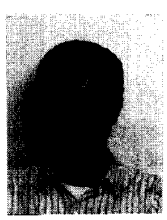
[36] M. Tistarelli and G. Sandini, "Estimation of depth from motion using an anthropomorphic visual sensor," *Image and Vision Computing*, vol. 8, no. 4, pp. 271–278, 1990.

[37] M. Tistarelli, "Multiple constraints for optical flow," *Proc. Third European Conf. Computer Vision*, pp. 61–70, Stockholm, Sweden, May 1994.

[38] B.K.P. Horn, "Relative orientation," *Int'l J. Computer Vision*,, no. 4, pp. 59–78, 1990.

[39] B. Kamgar-Parsi, "Practical computation of pan and tilt angles in stereo," Tech. Rep. CS-TR-1640, Univ. of Maryland, College Park, Md., Mar. 1986.

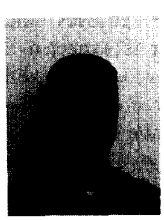
[40] B. Sabata and J.K. Aggarwal, "Estimation of motion from a pair of range images: A review," *CVGIP*, special issue on Image Understanding, vol. 54, no. 3, pp. 309–324, Nov. 1991.



Enrico Grosso received a degree in electronic engineering in 1988 and the PhD in computer science and electronic engineering in 1993 from the University of Genoa.

Since 1988, Dr. Grosso has been involved as project investigator and task manager in various ESPRIT projects funded by the European Community. During 1990 he visited the National Institute of Scientific Research LIFIA of Grenoble, France. In 1992, he was a visiting scientist at the Department of Computer Science of the University of Rochester, New York.

Dr. Grosso is currently an assistant professor in the Department of Communication, Computer, and Systems Science at the University of Genoa. His main research interests cover biological and artificial vision, visuo-motor coordination, and robotics.



Massimo Tistarelli received a degree in electronic engineering and the PhD in computer science and electronic engineering (in 1992) from the University of Genoa.

Since 1984, Dr. Tistarelli has been working on image processing and computer vision at the Integrated Laboratory for Advanced Robotics of the Department of Communication, Computer, and Systems Science of the University of Genoa where he is currently an assistant professor. In 1986, he was a research assistant at the Department of Computer Science of Trinity College, Dublin, Ireland, where he was developing a system for the analysis of image data primarily aimed at the investigation of low-level visual proceses. In 1989, he was a visiting scientist at Thinking Machines Co. developing parallel algorithms for dynamic image processing on the Connection Machine.

Dr. Tistarelli's research interests include robotics, computer vision (particularly in the area of three-dimensional and dynamic scene analysis), image processing, and artificial intelligence.