Proceedings of the 1998 IEEE/RSJ
Intl. Conference on Intelligent Robots and Systems
Victoria, B.C., Canada • October 1998

Ground Plane Detection using Visual and Inertial Data Fusion

Jorge Lobo

Institute of Systems and Robotics,
Electrical Engineering Department,
University of Coimbra,
3030 Coimbra, Portugal,
jlobo@isr.uc.pt
Fax: +351-39-406672

Jorge Dias

Institute of Systems and Robotics,
Electrical Engineering Department,
University of Coimbra,
3030 Coimbra, Portugal,
jorge@isr.uc.pt
Fax: +351-39-406672

Abstract

Active vision systems can be used in robotic systems for navigation. The active vision system provides data on the robot's environment. In mobile systems the position and attitude of the cameras relative to the world can be hard to determine. Inertial sensors coupled to the active vision system can provide valuable information to aid the image processing task. In human and other animals the vestibular system plays a similar role.

In this article, we explain our recent steps in the integration of inertial data with an active vision system. The active vision system has a set of stereo cameras capable of vergence, with a common baseline, pan and tilt. A process of visual fixation has already been implemented, enabling symmetric vergence on any selected point. An inertial system prototype, based on low-cost sensors was built. It is used to keep track of the gravity vector, allowing the identification of the vertical in the images. By performing visual fixation of a ground plane point, and knowing the 3D vector normal to level ground, we can determine the ground plane. The image can therefore be segmented, and the ground plane along which the robot can move identified.

For on-the-fly visualisation of the segmented images and the detected points a VRML viewer is used.

1 Introduction

In human and other animals the vestibular system gives inertial information essential for navigation, orientation or equilibrium of the body. In humans this sensorial system is located in the inner ear and it is crucial for several visual tasks and head stabilisation. The human vestibular system appears as a different sensorial modality, that co-operates with other sensorial systems and gives essential information for everyday tasks.

One example of co-operation is between the vestibular sensorial system and the visual system. It is well known that the inertial information plays an important role in some eye and head movements and head-stabilisation behaviours, as well as control of body posture and equilibrium [1].

The inertial information can also be useful on applications with autonomous systems and artificial vision. In the case of active vision systems, the inertial information gives a second modality of sensing that gives useful information for image stabilisation, control of pursuit movements, or ego-motion determination when the active vision system is used with a mobile platform. This kind of sensorial information is also crucial for the development of tasks with artificial autonomous systems where the notion of horizontal or vertical is important, see Viéville for one example [2].

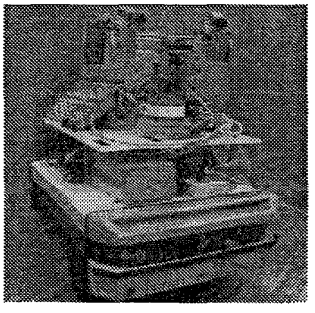

Figure 1: The mobile system with the active vision system.

Our inertial system prototype was coupled to an active vision system used in a mobile robot - see figure 1. The following sections describe the mobile system used and a first approach of inertial and vision data fusion, namely in identifying the ground plane.

2 A Prototype of An Inertial System

To study the integration of the inertial information in artificial autonomous systems that include active vision systems we decided to develop an inertial system prototype composed of low-cost inertial sensors. Their mechanical mounting and the necessary electronics for processing were also designed. The sensors used in the prototype system include a three-axial accelerometer, three gyroscopes and a dual-axis inclinometer. They were mounted on an acrylic cube as seen in figure 2.

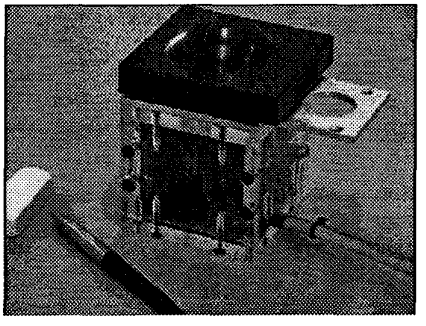

Figure 2: The inertial system prototype.

The three-axial accelerometer chosen for the system, while minimising eventual alignment problems, did not add much to the equivalent cost of three separate single-axis sensors. The device used was Summit Instruments' 34103A three-axial capacitive accelerometer. In order to keep track of rotation on the x-, y- and z-axis three gyroscopes were used. The piezoelectric vibrating prism gyroscope Gyrostar ENV-011D built by Murata was chosen. To measure tilt about the x and y-axis a dual axis AccuStar electronic inclinometer, built by Lucas Sensing Systems, was used. See [3] for details.

To handle the inertial data acquisition, and also enable some processing, a micro-controller based card was built. This card has analogue filters, an A/D converter as is based on Intel's 80C196KC microcontroller. The robot's master processing unit has an EISA bus interface, where the card is connected along with another for image acquisition and processing. This card is a framegrabber module with two Texas Instruments TMS320C40 DSPs that handles the video processing.

Figure 3 shows the architecture of the system and the computer that supervises the active vision, moving platform and inertial system. The inertial sensors were mounted inside an acrylic cube, enabling the correct alignment of the gyros, inclinometer (mounted on the outside) and accelerometer, as can be seen in figure 2. This cube is connected to, and continuously monitored by, the micro-controller card in the host computer.

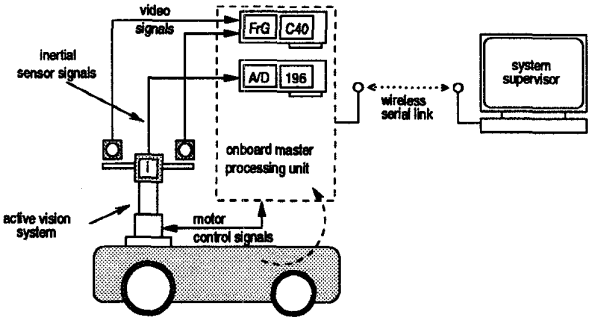

Figure 3: System Architecture. The inertial system processing board uses the Master processing unit as host computer.

3 Identifying the ground plane

3.1 System Geometry

The inertial unit is placed at the middle of the stereo camera baseline $\{C_y\}$, as seen in figure 4. The head co-ordinate frame referential, or Cyclop $\{C_y\}$is defined as having the origin at the centre of the baseline of the stereo cameras.

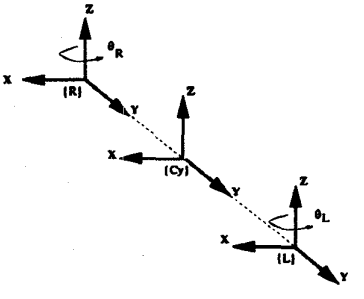

Figure 4: System Geometry.

Each camera position has its own referential, $\{R\}$ and $\{L\}$ being for the right and left positions. The

cameras have vergence capabilities, thus having $\{R\}$ and $\{L\}$ not only translated along the baseline ($\{C_y\}$ y-axis) but also rotated with respect to their z-axis. Notice that in our case we have a common baseline and symmetric vengeance, i.e. $\theta_R = -\theta_L$. To comply with the typical referential convention used for cameras, two additional referentials, $\{C_R\}$ and $\{C_L\}$, are considered - see figure 5.

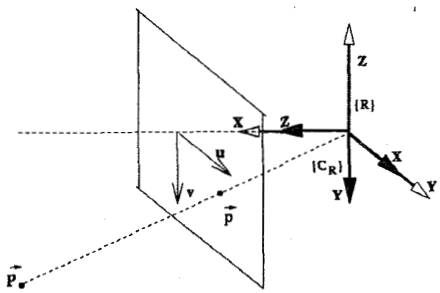

Figure 5: The camera referential and picture co-ordinates.

To express a world point $\vec{P}$, given in the camera referential, on the Cyclop referential $\{C_y\}$ we have

$$^{C_y}\vec{P} = {}^{C_y}T_{C_R}.{}^{C_R}\vec{P} = {}^{C_y}T_R.{}^{R}T_{C_R}.{}^{C_R}\vec{P} \quad (1)$$

and

$$^{C_y}\vec{P} = {}^{C_y}T_{C_L}.{}^{C_L}\vec{P} = {}^{C_y}T_L.{}^{L}T_{C_L}.{}^{C_L}\vec{P} \quad (2)$$

where

$$^{R}T_{C_R} = {}^{L}T_{C_L} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

$$^{C_y}T_R = \begin{bmatrix} \cos\theta_R & -\sin\theta_R & 0 & 0 \\ \sin\theta_R & \cos\theta_R & 0 & -\frac{b}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

$$^{C_y}T_L = \begin{bmatrix} \cos\theta_L & -\sin\theta_L & 0 & 0 \\ \sin\theta_L & \cos\theta_L & 0 & \frac{b}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5)$$

and b is the baseline distance.

A projection point $\vec{p} = (u, v)$ in each camera image is related with a 3D point $\vec{P} = (X, Y, Z)$ by the perspective relations

$$u = S_u f \frac{X}{Z} \qquad v = S_v f \frac{Y}{Z} \quad (6)$$

where u and v are the pixel co-ordinates with origin at the image centre, f is the camera focal distance, S_u and S_v are the scale factors and $\vec{P}$ is expressed in the camera referential.

If we know $\vec{P} = (X, Y, Z)$, finding the projection (u, v) is trivial. The reverse problem involves matching points between the left and right images. Establishing this correspondence will give us enough equations to determine the 3D co-ordinates, if a few vision system parameters are known. However if visual fixation is used, the geometry is simplified and the reconstruction of the 3D fixated point is simplified, as can be seen in figure 6. Notice that the visual fixation can be achieved by controlling the active vision system and the geometry generated by the process allows a fast and robust 3D reconstruction of the fixation point - see [4] and [5] for details.

3.2 Inclinometer gives the ground plane orientation

The inclinometer data can be used to determine the orientation of the ground plane. In order to locate this plane in space at least one point belonging to the ground plane must be known. When the vehicle is stationary or subject to constant speed the inclinometer gives the direction of $\vec{g}$ relative to the Cyclop referential $\{C_y\}$. Assuming the ground is levelled, and with α_x and α_y being the sensed angles along the x and y-axis, the normal to the ground plane will be

$$\begin{aligned} \hat{n} &= -\frac{\vec{g}}{\|\vec{g}\|} \\ &= \frac{1}{\sqrt{1-\sin^2\alpha_x-\sin^2\alpha_y}} \begin{bmatrix} -\cos\alpha_x \sin\alpha_y \\ \cos\alpha_y \sin\alpha_x \\ \cos\alpha_y \cos\alpha_x \end{bmatrix} \end{aligned} \quad (7)$$

given in the Cyclop frame of reference. Using this inertial information the equation for the ground plane will be given by

$$\hat{n}.\vec{P} + h = 0 \quad (8)$$

where $\vec{P}$ is a point in the plane and h is the distance from the origin of $\{C_y\}$ down to the ground plane.

3.3 Segmentation of ground plane from visual fixation and inertial sensing

To obtain a point belonging to the ground plane it is necessary to establish a mechanism to achieve visual fixation. This mechanism was developed in our

laboratory and is described in [4] and [5]. If the active vision system fixates in a point that belongs to the ground plane, the ground plane could be determined in the Cyclop referential $\{C_y\}$ using the inclinometer data. Hence, any other correspondent point in the image can be identified as belonging or not to the ground plane.

If the fixation point is identified as belonging to the ground plane, the value of h in (8), can be determined. As seen in figure 6 (where only α_y is non null to keep the diagram simple) h will be given by

$$h = -\hat{n}.\vec{P}_f \tag{9}$$

An algorithm for the segmentation of the ground plane can now be presented, based on the solution of (9). Starting with a point of interest in one of the images, say the right image (u_R, v_R), from (6) and (2) the point expressed in the Cyclop referential is given by

$${}^{C_y}\vec{P} = {}^{C_y}T_{C_R}.{}^{C_R}\vec{P}(u_R, v_R, \lambda_R) \tag{10}$$

where λ_R represents an unknown value (depending on depth from camera). Substituting (10) in (8), λ_R can be determined and hence $\vec{g}$ is completely known in the Cyclop referential. Expressing $\vec{P}$ in the $\{C_L\}$ referential by

$${}^{C_L}\vec{P} = {}^{C_L}T_{C_y}.{}^{C_y}\vec{P} = \left({}^{C_y}T_{C_R}\right)^{-1}.{}^{C_y}\vec{P} \tag{11}$$

the correspondent point of interest (u_L, v_L) generated by the projection of P in the left image is given by

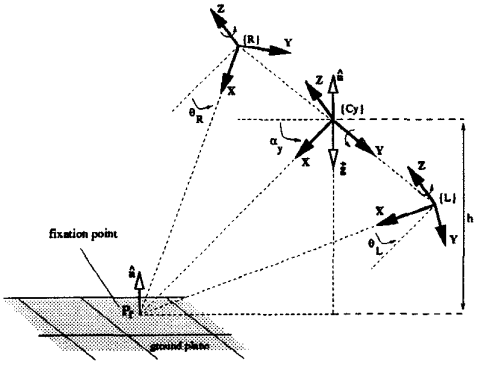

Figure 6: Ground plane point fixated. The point $\vec{P}_f$ in the ground plane is visualised by the active vision system. The geometry of this visualisation corresponds to a state, named visual fixation.

$$u_L = S_u f\frac{X}{Z} \qquad v_L = S_v f\frac{Y}{Z} \tag{12}$$

The correspondent point and its neighbourhood in the left image can then be tested for a match with the original point of interest in the right image. If there is a match, the point belongs to the ground plane. If there is no match the point must be something other than the floor, possibly an obstacle.

3.4 Ground point computation

When visual fixation is obtained for a ground point and assuming symmetric vergence (i.e. $\theta = \theta_R = -\theta_L$) from (9) and (7) we have

$$h = -\hat{n}.\vec{P}_f = \frac{b\sin\alpha_y\cos\theta}{2\sin\theta} \tag{13}$$

as can easily be seen in figure 6 (where α_x is null, but (13) still holds for any α_x).

This value of h will be used to determine if other points in the image belong or not to the level plane passing the image centre point (i.e. fixation point).

Taking λ_R out of (13) and (10) and substituting in (11) and (12) we get a set of equations that allow a quick computation of the algorithm for a given a set of stereo images.

3.5 Ground Detection Results

Figure 7 shows a pair of stereo images where fixation was obtained for a ground plane point. In this example α_x is null and $\alpha_y = 16.05^o$, $\theta = 2.88^o$ and $b = 29.6cm$. Making $h = b.\cot(\theta)sin(\alpha_y)/2 \simeq 81,3cm$.

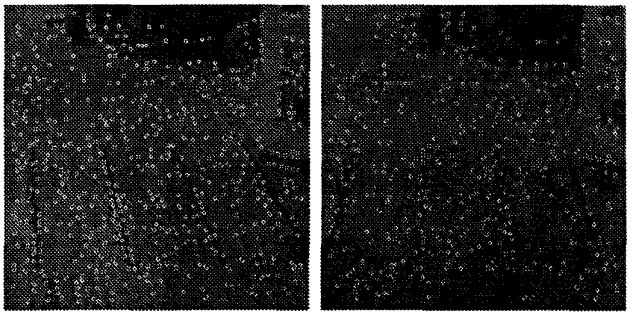

Figure 7: Stereo images with a set of initial points.

The points of interest in the right image can be parsed as described in the previous section. The points shown in figure 7 were obtained using SUSAN [6] corner detector.

Figure 8 and 9 show the matched ground plane points of interest. Grahams Algorithm [7] was used for computation of the convex polygon envolving the

set of points. Adjusting a convex polygon to the set of points can lead to erroneous ground patch segmentation.

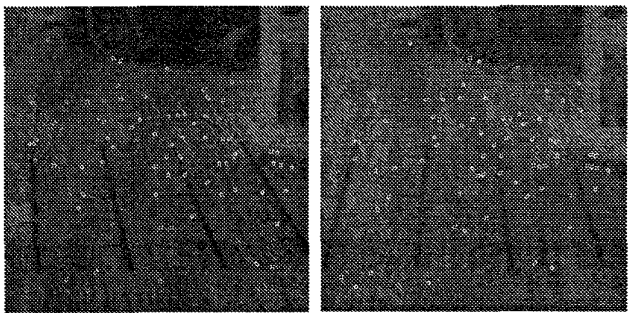

Figure 8: Detected ground points.

We intend to modify the algorithm to allow for non-convexity when points are too far apart and an obstacle could be in the way. A small change has to be made to the algorithm and special cases taken into account, such as having multiple isolated polygons.

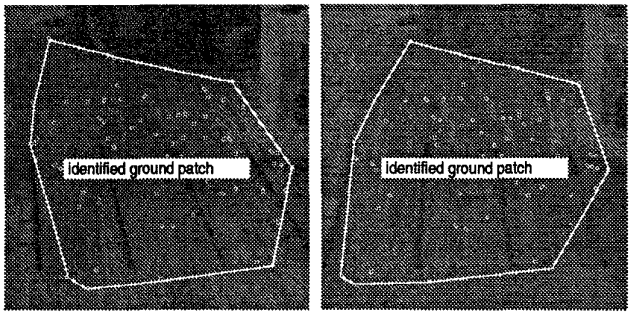

Figure 9: Identified ground patch.

3.6 Viewing with VRML

For visualisation of the detected ground points a VRML (Virtual Reality Modelling Language) world is generated [8]. VRML has become the standard for describing 3D scenes on the Internet. A web browser with the appropriate plug-in lets the user see the images from the robots viewpoint, or change the viewpoint and "fly" around the 3D world viewed by the vision system. The identified ground plane patch can be mapped onto the 3D scene, as seen in figure 10.

The images were generated using Netscape Communicator 4.04 web browser with the Cosmo Player 2.0 plug-in from Silicon Graphics, Inc.

To update the VRML world on-the-fly, only the ground patch vertex points need to be sent, so that

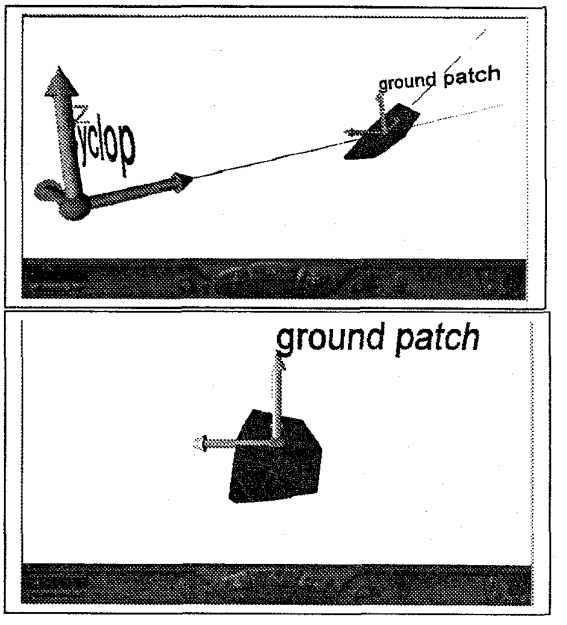

Figure 10: Two views of ground patch with detected points in VRML.

the polygon can be rendered. When bandwidth is not a problem the segmented image patch can also be sent and placed onto the polygon. VRML opens many other possibilities such as tele-operation or path-planning environments.

4 Conclusions

The integration of inertial systems in autonomous systems opens a new field for the development of applications based or related with inertial information. Together with the information provided by an artificial vision system, several applications could be implemented. In this article we described a system that integrates inertial sensors and active vision systems used in autonomous vehicles.

Information about the ground plane extracted from the inertial sensors is used to identify the floor plane using the simulation of visual fixation with an active vision system. By fixating the visual system on a point in the ground plane, the 3D position of the plane in space is obtained. Any other correspondent point in the stereo image can be identified as belonging or not to the ground plane. Segmentation of the image is therefore accomplished. Some ground detection results were presented. For on-the-fly visualisation of the detected points a VRML world is generated which

can be viewed in any web browser with the appropriate plug-in. VRML opens many other possibilities such as tele-operation or path-planning environments.

Acknowledgments

This work was sponsored by PO-ROBOT project from NATO Science for Stability program.

References

[1] H. Carpenter. *Movements of the Eyes.* London Pion Limited, 2nd edition, 1988. ISBN 0-85086-109-8.

[2] T. Viéville and O.D. Faugeras. Computation of Inertial Information on a Robot. In Hirofumi Miura and Suguru Arimoto, editors, *Fifth International Symposium on Robotics Research*, pages 57–65. MIT-Press, 1989.

[3] Jorge Lobo and Jorge Dias. Integration of Inertial Information with Vision towards Robot Autonomy. In *Proceedings of the 1997 IEEE International Symposium on Industrial Electronics*, pages 825–830, Gimaraes, Portugal, July 1997.

[4] Jorge Dias, Carlos Paredes, Inácio Fonseca, and A. T. de Almeida. Simulating Pursuit with Machines. In *Proceedings of the 1995 IEEE Conference on Robotics and Automation*, pages 472–477, Japan, 1995.

[5] Carlos Paredes, Jorge Dias, and A. T. de Almeida. Detecting Movements Using Fixation. In *Proceedings of the 2nd Portuguese Conference on Automation and Control*, pages 741–746, Oporto, Portugal, September 1996.

[6] S.M. Smith and J.M. Brady. SUSAN - a new approach to low level image processing. *Int. Journal of Computer Vision*, 23(1):45–78, May 1997.

[7] Joseph O'Rourke. *Computational Geometry in C.* Cambridge University Press, 1993. ISBN 0-512-22592-2.

[8] A. L. Ames, D. R. Nadeau, and J. L. Moreland. *VRML 2.0 Sourcebook.* John Wiley and Sons, 2nd edition, 1997. ISBN 0-471-16507-7.