Proceedings of the 10th Mediterranean Conference
on Control and Automation - MED2002
Lisbon, Portugal, July 9-12, 2002.

COOPERATION BETWEEN VISUAL AND INERTIAL INFORMATION FOR 3D VISION

Jorge Dias[1], Jorge Lobo[1], Luis A. Almeida[2]

[1] Institute of Systems and Robotics, University of Coimbra, Portugal
e-mail: {jorge,jlobo}@isr.uc.pt.
[2]Instituto Politecnico de Tomar, Tomar,Portugal
e-mail: laa@ipt.pt.

Keywords: Computer Vision, Robotics, Sensor Fusion, Inertial Systems, 3D Depth Maps.

Abstract

Advanced sensor systems, exploring high integrity and multiple sensor modalities, have been significantly increasing the capabilities of autonomous vehicles and enlarging the application potential of vision systems. The article describes the cooperation between two relevant sensors - vision systems and inertial sensors. Vision and inertial sensing are two sensory modalities that can be explored to give robust solutions on segmentation of images and three-dimensional vision. This cooperation between these two sensory modalities may be useful for the elaboration of high-level representations such as multi-modality 3D maps, segmentation of leveled ground or vertical structures.

In this paper we propose a real-time system that extracts information from dense relative depth maps. This method enables the integration of depth cues on higher level processes including segmentation of structures, object recognition, robot navigation or any other task that requires a three-dimensional representation of the physical environment.

1 Introduction

Inertial sensors are a class of sensors useful for internal sensing since they are not dependent on external references. In human and other animals the ear vestibular system gives inertial information essential for navigation, orientation or equilibrium of the body. This is a sensorial modality, which 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, in humans, the information provided by the vestibular system is used during the execution of navigational and visual movements, as described by Carpenter [1]. However the inertial information is also important for head-stabilisation behaviors, including the control of posture and equilibrium of the body.

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 [2] for one example.

A vision system with embedded inertial sensors obtains a partial estimation of *self-motion* or *absolute orientation*. The gravity and the rigid body acceleration can be measured from these sensors as well as its angular instantaneous velocity. From these quantities, the instantaneous velocity, the angular position, and linear translation of the vision system can be obtained. The cooperation between these quantities with visual information can be useful to estimate the instantaneous visual motion, to segment images (for example between moving objects and background), estimation of vision system orientation with respect to the horizontal plane or to segment depth maps based on the estimation of vertical or self-motion.

This article presents our most recent results on the use and integration of those two modalities.

In this article we explore the integration of inertial sensor data in vision systems to segment 3D maps. The depth maps are obtained by a stereo-vision system and the the three-dimensional data is processed to identify specific structures in these 3D maps. The inertial sensor data enable to recover camera pose, and rectify the 3D maps to a common reference ground plane, enabling the segmentation of vertical and horizontal geometric features. The aim of this work is a fast real-time system, so that it can be applied to autonomous robotic systems or to automated car driving systems, for modeling the road, identifying obstacles and roadside features in real-time.

2 Segmentation of Dense Depth Maps using Inertial Data

One of the very important tasks in computer vision is to extract depth information of the world. Stereoscopy is a technique to extract depth information from two images of a scene taken from different view points. This information can be integrated on a single entity called dense depth map.

In this paper we propose a real-time system that extracts information from dense relative depth maps. This method enables the integration of depth cues on higher level processes including segmentation of structures, object recognition, robot navigation or any other task that requires a three-dimensional representation of the physical environment.

Inertial sensors coupled to a vision system can provide important inertial cues for the ego-motion and system pose. The sensed gravity provides a vertical reference. Depth maps obtained from a stereo camera system can be segmented using this vertical reference, identifying structures such as vertical features and leveled planes.

In humans and in animals the vestibular system in the inner ear gives inertial information essential for navigation, orientation, body posture control and equilibrium. In humans this sensorial system is crucial for several visual tasks and head stabilisation. It is well known that the information provided by the vestibular system is used during the execution of visual movements such as gaze holding and tracking, as described by Carpenter [1]. Neural interactions of human vision and vestibular system occur at a very early processing stage [3].

In this work we use the vertical reference provided by the inertial sensors to perform a fast segmentation of depth maps obtained from a stereo real time algorithm.

Nowadays micro-machined low cost inertial sensors can be easily incorporated in computer vision systems. These sensors can perform as an artificial vestibular system, providing valuable data to the vision system. The motivation might be stronger in applications such as walking or flying robots, but in automobiles, due to suspension and system compliance, it is also beneficial to have inertial sensors coupled to the vision system cameras.

2.1 Related Work

The aim of stereo systems is to achieve an adequate throughput and precision to enable video-rate dense depth mapping. The throughput of a stereo machine can be measured by the product of the number of depth measurements per second (pixel/sec) and the range of disparity search (pixels); the former determines the density and speed of depth measurement and the later the dynamic range distance measurement [8], [9], [10], [11].

The CMU Robotics group succeeded in producing a video-rate stereo machine based on the multi-baseline stereo algorithm to generate a dense range map[12]. They use multiple images obtained by multiple cameras to produce different baselines in length and direction. Based on this studies an efficient implementation of area correlation stereo is available: the *SRI Stereo Engine* [13]. The *Stereo Engine* algorithms have been continuously developed by *SRI* and it runs efficiently on many computational platforms, including standard personal computers and embedded processors.The standard development environment, the *Small Vision System* (SVS), runs on personal computers under *Linux* or *MS Windows*. This implementation gives an efficient solution to support camera calibration, 3D reconstruction, and effective filtering. The development of our real-time system for 3D map segmentation is based on some of the routines of this system.

The cooperation of the inertial and visual systems in mobile robot navigation was studied by Viéville and Faugeras. They proposed the use of an inertial system based on low cost sensors for mobile robots [2] and using the vertical cue taken from the inertial sensors [14] [15] [16]. An inertial sensor integrated optical flow technique was proposed by Bhanu *et al.* [17]. Panerai and Sandini used a low cost gyroscope for gaze stabilization of a rotating camera, and compared the camera rotation estimate given by image optical flow with the gyro output [18] [19]. Mukai and Ohnishi studied the recovery of 3D shape from an image sequence using a video camera and a gyro sensor [20].

In our previous work on inertial sensor data integration in vision systems, the inertial data was directly used with the image data. Using just one vanishing point we recovered the camera's focal distance [4]. In a typical indoor corridor scene the vanishing point can also provide an external bearing for the robots navigation frame. Knowing the geometry of a stereo rig, and its pose from the inertial sensors, the homography of level planes can be recovered, providing enough restrictions to segment and reconstruct vertical features [6] and levelled planar patches [5]. In this work we use the inertial data to perform a fast segmentation pre-computed depth maps obtained from the vision system.

3 Depth Maps

In order to describe the tasks involved in depth map construction, let us consider the geometric model in figure 1. This model describes graphically our stereo vision system - see figure 2. The diagram shows the top view of a stereo system composed of two pinhole cameras. The left and right image planes are coplanar and represented by the segments I_l and I_r respectively. C_l and C_r are the centers of projection. The optical axes are parallel: for this reason, the *fixation point* defined as the point of intersection of the optical axes, lies infinitely far from the cameras.

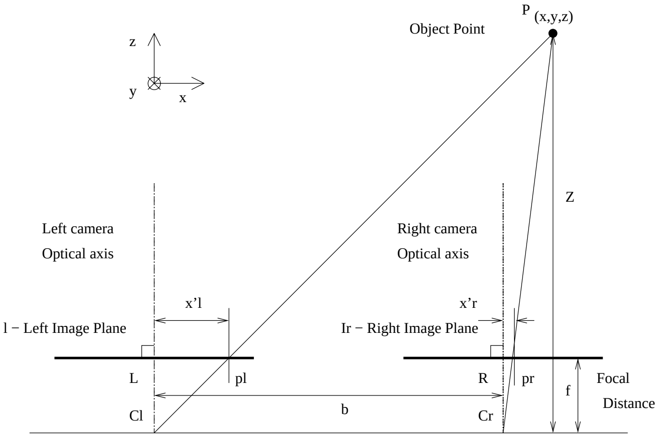

Figure 1: Geometry of front-parallel stereo setup.

The way in which stereo determines the position in space of $P_{(x,y,z)}$ is *triangulation*, that is by intersecting the rays defined by the centers of projection and the images of P and p_l, p_r.

Consider a point $P_{(x,y,z)}$, in three-dimensional world coordinate, on a object and its projections p_l and p_r.

The distance, b, between the centers of projections C_l and C_r, is called *baseline*. Let this point have image coordinates (x'_l, y'_l) and (x'_r, y'_r) in the left and right images plane of the respective cameras. Since the cameras are front-parallel and aligned, we have that $y'_l = y'_l$ Let f be the focal length of both cameras, the perpendicular distance between the lens center and the image plane, the depth z, will be the distance between P and the baseline. Using the left camera as reference and by similar triangles:

$$\begin{aligned} x'_l/f &= x/z \\ x'_r/f &= (x-b)/z \\ y'_l/f &= y'_r/f = y/z \end{aligned} \tag{1}$$

Solving for (x, y, z) gives:

$$\begin{aligned} x &= x'_l * b/(x'_l - x'_r) \\ y &= y'_l * b/(x'_l - x'_r) \\ z &= f * b/(x'_l - x'_r) \end{aligned} \tag{2}$$

The quantity $d = x'_l - x'_r$ which appears of the above equations is called *disparity*. From equation (2) we see that depth is inversely proportional to disparity and disparity can only be measured in pixel differences.

The classical approach to estimate disparities uses two techniques: feature matching and correlation. In a feature-based algorithm, a number of complex tokens are extracted from each left and right images, and then combined according to some constraints. The second technique uses a measure of similarity, correlation for example, to find matching points in two images composing the stereo pair. For each point of the reference image, the corresponding point is selected in the other image by searching for a maximum in similarity measure.

From equation (2) we also see that depth is proportional to the baseline b. If we have a fixed error in determining the disparity then the accuracy of depth determination will be amplified by b. However, as the camera separation becomes larger, difficulties will arise in correlation of the two camera images.

Many stereo camera configurations have vergence and do not comply with the front-parallel geometric model. A stereo configuration with vergence angle can be considerably simplified when the images of interest have been rectified, i.e., replaced by two projectively equivalent pictures with a common image plane parallel to the baseline joining the two optical centers, and equivalent to a front-parallel system as in our system. The *rectification* process can be implemented by projecting the original pictures onto the new image plane [21]. With an appropriate choice of coordinate system, the rectified images have scanlines parallel to the baseline and all the front-parallel geometry of figure 1 can be applied.

3.1 The Stereo Vision System

In order to compute range from stereo images we are using the *SRI Stereo Engine* [13]. It implements an area correlation algorithm for computing range from stereo images and it supports camera calibration, 3D reconstruction, and effective filtering. We are running an implementation of *Stereo Engine* for *Linux* and named *Small Vision System (SVS)*. *SVS* consists of a set of library functions for stereo algorithms optimized for *Pentium* architectures, using *MMX* instructions. It can receive input stereo images from standard cameras and video capture devices. On this particular work we are using a small and compact stereo head developed by *Videre Design* [13] (see figure 2), the STH-V3. This analog vision head can also send a single video signal with the interlaced stereo image pair. *STH-V3* consists of two synchronized cameras modules, mounted on a baseboard, with 320x240 pixels (*NTSC*). The software is running on a *Linux RedHat 7.1* box (PII 350Mhz) with a *Pinnacle Studio PCTV* (Bt878-chip) card as frame-grabber. With a frame size of 160x120, searching 16 disparities and a search window size of 5 x 5 we achieved a frame rate close to 30 Hz.

4 Inertial Data

An *Inertial Measurement Unit (IMU)* coupled to a camera can provide valuable data about camera pose and movement. Figure 2 shows an inertial system prototype built at our lab [7] that was coupled to a stereo camera rig to perform the tests.

Camera calibration was performed using a fixed target and moving the system, recovering the cameras' intrinsic parameters, as well as the target positions relative to the cameras.

By moving the cameras instead of the target, the cameras' position is determined relative to the fixed target. Since the *IMU* is rigidly connected to the camera, R and t shown in figure 3 can be determined from the set of camera positions obtained from the calibration and the correspond-

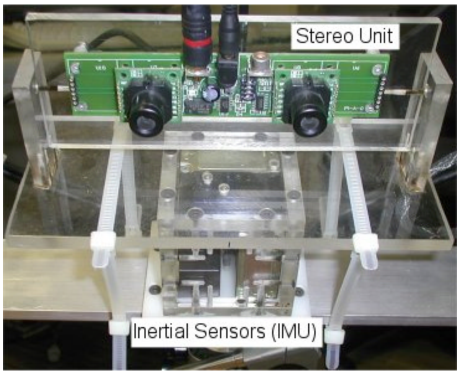

Figure 2: Cameras with Inertial Measurement Unit based on low cost sensors.

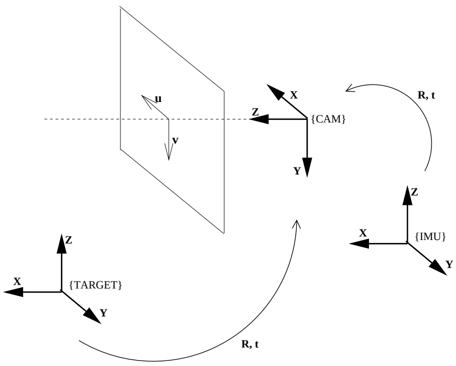

Figure 3: Camera and *IMU* referentials.

ing data from the inertial sensors. By performing a trajectory with the target always in view, the camera calibration can reconstruct the camera pose and position.

If both *IMU* and cameras are perfectly aligned, we have a simple translation change of axis, and

$$
{}^{\mathcal{C}}\boldsymbol{T}_{\mathcal{CAM}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & b/2 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}
$$

where b is the baseline, $\{\mathcal{CAM}\}$ the camera referential and $\{\mathcal{C}\}$ the camera system referential, with origin at the center of the baseline.

Having determined the rigid transformation between the camera and the *IMU*, the sensed acceleration and rotation are mapped to the camera system referential.

4.1 Gravity Vector

The measurements $\boldsymbol{a}$ taken by the inertial unit's accelerometers include the sensed gravity vector $\boldsymbol{g}$ summed with the body's acceleration $\boldsymbol{a}_b$:

$$
\boldsymbol{a} = \boldsymbol{g} + \boldsymbol{a}_b \tag{4}
$$

Assuming the system is motionless, then $\boldsymbol{a}_b = 0$ and the measured acceleration $\boldsymbol{a} = \boldsymbol{g}$ gives the gravity vector in the system's referential. So, with a_x,a_y and a_z being the accelerometer filtered measurements along each axis, the vertical unit vector will be given by

$$
\hat{\boldsymbol{n}} = -\frac{\boldsymbol{g}}{\|\boldsymbol{g}\|} = \frac{1}{\sqrt{a_x^2+a_y^2+a_z^2}} \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} \tag{5}
$$

This vertical reference, given in the systems frame of reference, will be used in segmenting the depth maps obtained from the stereo algorithm.

Consider a point given in the camera system referential ${}^{\mathcal{C}}\boldsymbol{P}$ that belongs to the ground plane. The plane equation is given by

$$
{}^{\mathcal{C}}\hat{\boldsymbol{n}}.{}^{\mathcal{C}}\boldsymbol{P} + d = 0 \tag{6}
$$

where d is the distance from the origin to the ground plane, *i.e.*, the system height. In some applications it can be known or imposed by the physical mount.

5 Depth Maps in Inertial Reference Frame

In our experimental setup, the stereo algorithm provides depth maps in the left camera frame of reference. Using the vertical reference provided by the inertial sensors, $\hat{\boldsymbol{n}}$, the depth maps can be rotated and aligned with the horizontal plane. The points obtained in the camera referential, $\{\mathcal{C}\}$, can be converted to a world frame of reference $\{\mathcal{W}\}$. The vertical unit vector $\hat{\boldsymbol{n}}$ and system heigh d can be used to define $\{\mathcal{W}\}$, by choosing ${}^{\mathcal{W}}\hat{\boldsymbol{x}}$ to be coplanar with ${}^{\mathcal{C}}\hat{\boldsymbol{x}}$ and ${}^{\mathcal{C}}\hat{\boldsymbol{n}}$ in order to keep the same heading, we have

$$
{}^{\mathcal{W}}\boldsymbol{P} = {}^{\mathcal{W}}\boldsymbol{T}_{\mathcal{C}}.{}^{\mathcal{C}}\boldsymbol{P} \tag{7}
$$

where

$$
{}^{\mathcal{W}}\boldsymbol{T}_{\mathcal{C}} = \begin{bmatrix} \sqrt{1-n_x^2} & \frac{-n_x n_y}{\sqrt{1-n_x^2}} & \frac{-n_x n_z}{\sqrt{1-n_x^2}} & 0 \\ 0 & \frac{n_z}{\sqrt{1-n_x^2}} & \frac{-n_y}{\sqrt{1-n_x^2}} & 0 \\ n_x & n_y & n_z & d \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8}
$$

System height d can be known *apriori* or inferred from the subsequent segmentation process, using an initial null value.

If a heading reference is available, then $\{\mathcal{W}\}$ should not be restricted to having ${}^{\mathcal{W}}\hat{\boldsymbol{x}}$ coplanar with ${}^{\mathcal{C}}\hat{\boldsymbol{x}}$ and ${}^{\mathcal{C}}\hat{\boldsymbol{n}}$, but use the known heading reference. Using a heading reference given by the unit vector $\hat{\boldsymbol{m}} = (m_x, m_y, m_z)$ we get

$$
{}^{\mathcal{C}}\boldsymbol{T}_{\mathcal{W}} = \begin{bmatrix} m_x & n_y m_z - n_z m_y & n_x & -n_x d \\ m_y & n_z m_x - n_x m_z & n_y & -n_y d \\ m_z & n_x m_y - n_y m_x & n_z & -n_z d \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}
$$

We are therefor able to have $\{\mathcal{W}\}$ coherent with the inertial vertical and the available scene heading. The gyros included in the inertial measurement unit can be used to keep a heading without external references, but they accumulate drift over time. Visual land marks or a magnetic compass provide an external heading reference to reset the drift. In scenes of man made environments image vanishing points from detected edge lines can also provide a heading reference. The segmented depth maps can also be used, by identifying features such as walls in the points mapped to the inertial reference frame and above the ground plane.

6 Segmented Depth Maps

Using the vertical reference, the depth maps can be segmented to identify horizontal and vertical features. The aim in on having a simple algorithm suitable for a real-time implementation. Since we are able to map the points to an inertial reference frame, planar levelled patches will have the same depth z, and vertical features the same xy, allowing simple feature segmentation using histogram local peak detection.

Using the stereo depth algorithm we obtain a set of points ${}^{\mathcal{CAM}}P_i$ in the left camera referential. Using the previous equations we can map them to the world referential as

$$
{}^{\mathcal{W}}P_i = {}^{\mathcal{W}}\boldsymbol{T}_{\mathcal{C}}.{}^{\mathcal{C}}\boldsymbol{T}_{\mathcal{CAM}}.{}^{\mathcal{CAM}}P_i \tag{10}
$$

In order to detect the ground plane point, an histogram is performed for each point depth.

$$
hist_z(n) = \sum (P_i \mid floor(z_{P_i}) = n) \tag{11}
$$

The histogram's lower local peak z_{gnd} is used as the reference depth for the ground plane. The detected points can than be parsed and segmented as being a ground plane point, or some feature above ground. Points bellow the ground plane can be ignored or not, depending on the application.

$$
P_{gnd} = P_i \mid z_{gnd} - \delta \leq floor(z_{P_i}) \leq z_{gnd} + \delta \tag{12}
$$

$$
P_{above} = P_i \mid floor(z_{P_i}) \geq z_{gnd} + \delta \tag{13}
$$

were δ is the allowed tolerance.

The points above ground can be projected in the XY plane, and further segmentation performed to identify vertical features.

7 Results

A simple indoor scene was used to test our method. The stereo pair seen in figure 5 was obtained with the experimental setup shown in figure 4. Figure 6 shows the disparity image and reconstructed 3D points obtained with the 3D reconstruction based *SVS* package routines [13].

Using the vertical reference provided by the inertial sensors, in this case $\boldsymbol{n} \approx (-0.456, -0.022, 0.890)$, the 3D points were transformed to a world aligned frame of reference as previously described. Figure 7 shows a 3D view of the points in the world frame of reference.

In order to detect the ground plane, an histogram was done for all depths, and the peak used as a reference value, as seen in figure 8

Figure 4: Experimental setup with inertial sensors and vision system, and scene used for the test.

Figure 5: Stereo rectified image pair obtained with SVS [13] system.

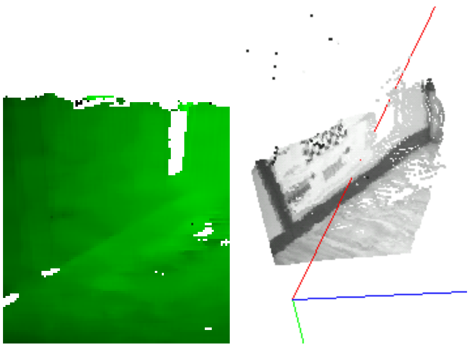

Figure 6: Disparity image obtained with SVS [13], and reconstructed 3D points

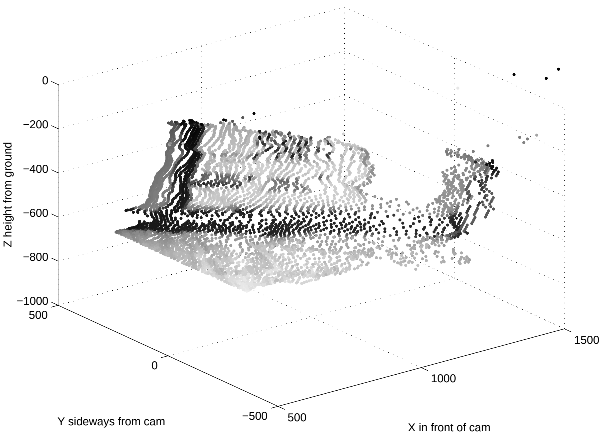

Figure 7: 3D view of the points.

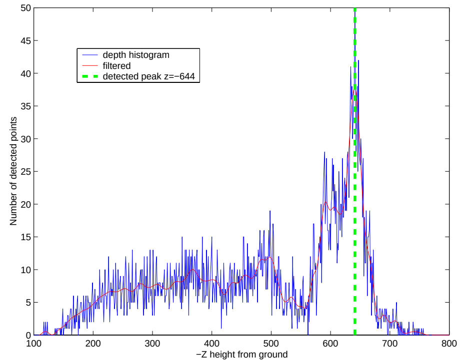

Figure 8: Depth histogram with detected peak.

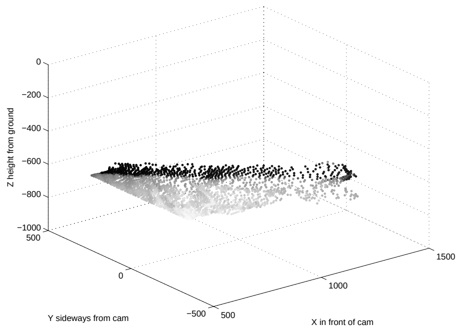

Figure 9: Ground plane points segmented from global map seen in figure 7.

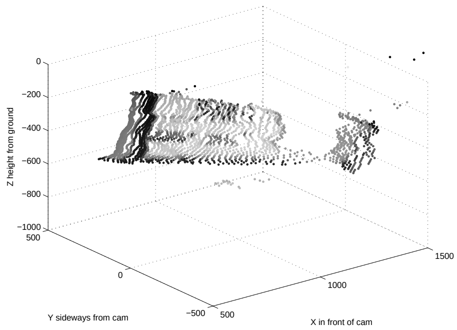

Figure 10: Points above the floor, walls or obstacles, segmented from global map seen in figure 7.

The points were than parsed and segmented as ground plane points, figure 9, and points above ground, figure 10.

A linear line fit was done using the points above ground, ignoring their depth, to reconstruct the wall orientation in the test scene. Figure 11 shows the result. More complex scenes require a previous point clustering stage, so that a simplified world model can be built, but this only has to be done in 2D.

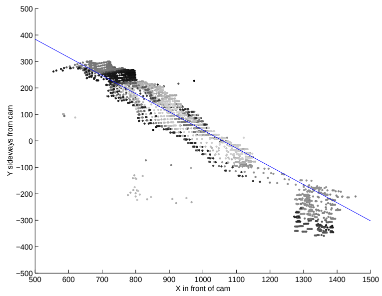

Figure 11: Top view of all points above the floor, and line fit for wall orientation.

Figures 12, 14 and 14 present some more results. Since low resolution images are being used to achieve real-time performance, the depth maps obtained are not very precise, and a large tolerance threshold has to be used to segment ground plane points.

8 Conclusions

Depth maps obtained from a stereo camera system were segmented using a vertical reference provided by inertial sensors, identifying structures such as vertical features and level planes. Rectifying the maps to a reference ground plane enables the segmentation of vertical and horizontal geometric features. Preliminary results were presented that show the validity of the method.

The aim of this work is a fast real-time system, avoiding 3D point clustering methods that are not suitable for real-time implementations. It can be applied to an automated car driving system, modeling the road, identifying obstacles and roadside features.

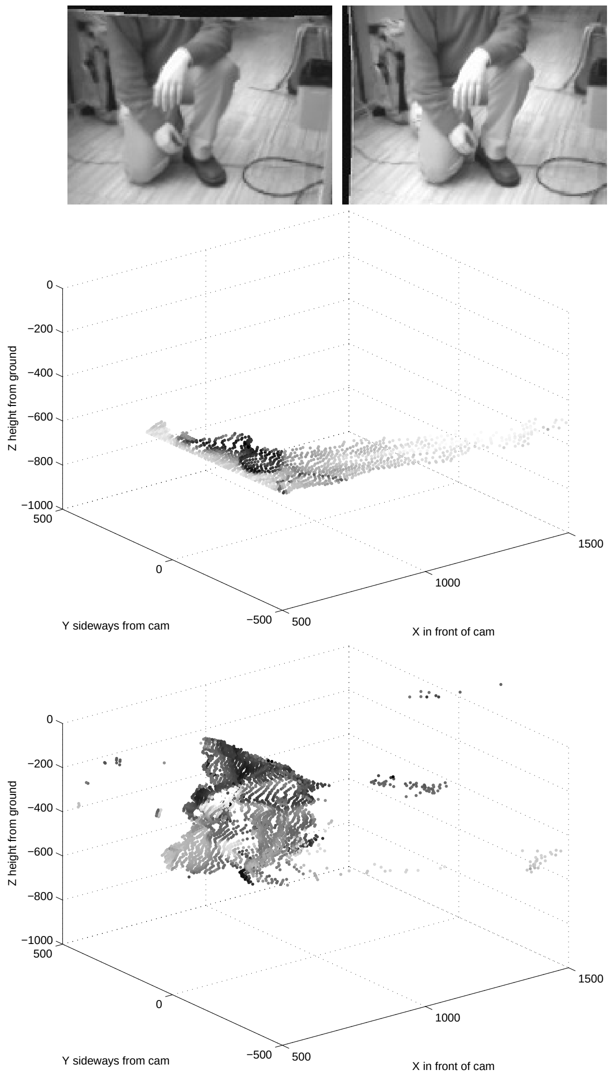

Figure 12: Stereo rectified image pair; Ground plane points; Points above the floor, walls or obstacles.

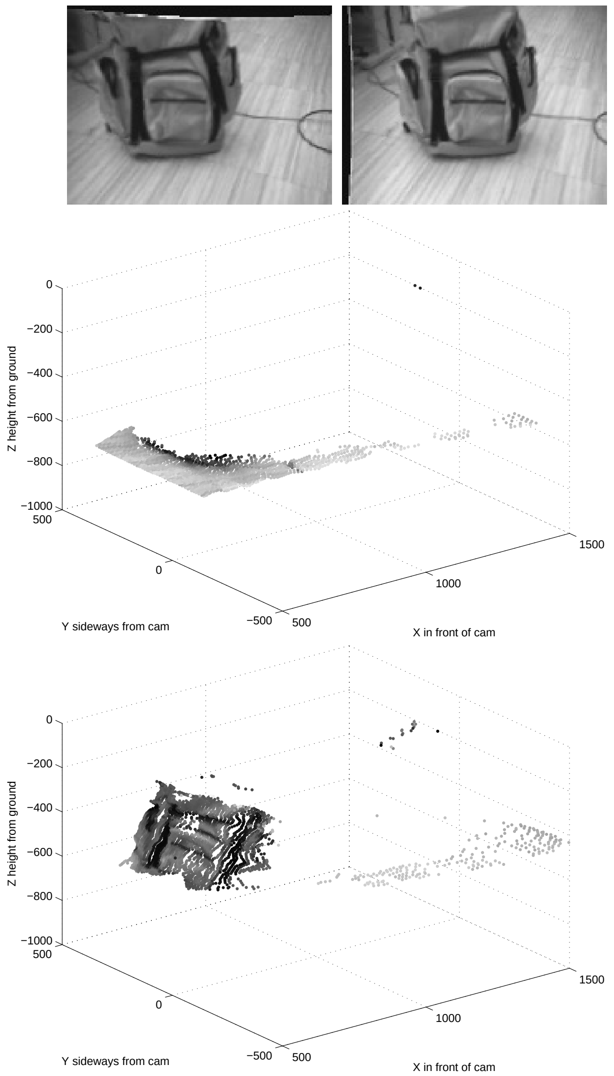

Figure 13: Stereo rectified image pair; Ground plane points; Points above the floor, walls or obstacles.

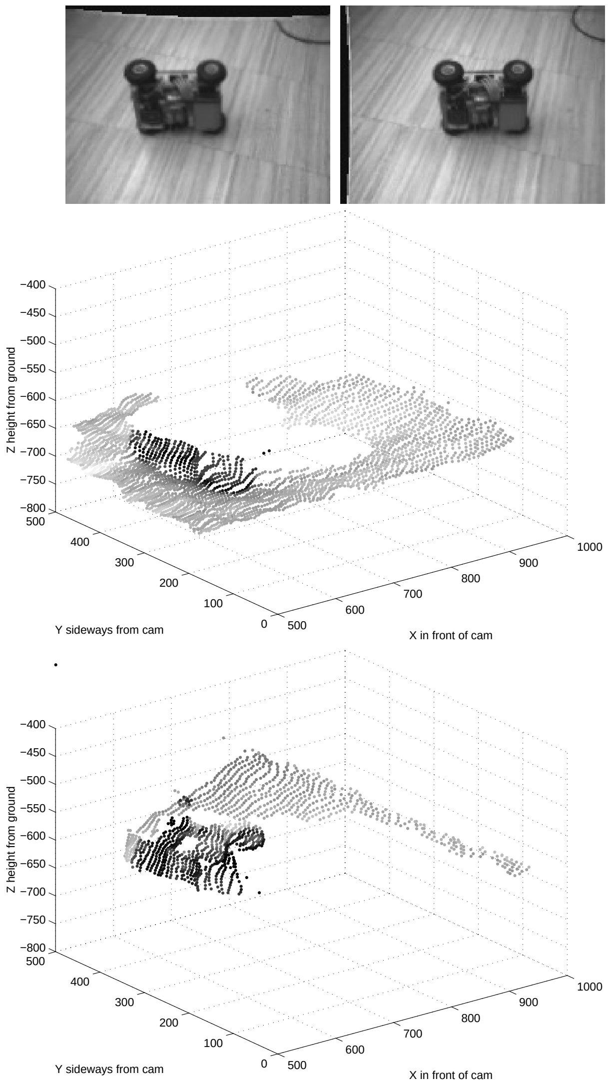

Figure 14: Stereo rectified image pair; Ground plane points; Points above the floor, walls or obstacles.

Acknowledgments

The authors would like to acknowledge the support from the project financed by PRAXIS /P/EEI 11218/98, from the Portuguese Foundation for Science and Technology.

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", *Fifth International Symposium on Robotics Research*, ed. H. Miura, S. Arimoto, pg. 57-65, MIT-Press,1989

[3] A. Berthoz, *The Brain's Sense of Movement*, Havard University Press, 2000, ISBN: 0-674-80109-1.

[4] J. Lobo and J. Dias, "Fusing of image and inertial sensing for camera calibration", in *Proc. Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems*, Baden-Baden, Germany, August 2001, pp.103-108.

[5] J. Lobo and J. Dias, "Ground plane detection using visual and inertial data fusion", in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Victoria, Canada, October 1998, pp.912-917.

[6] J. Lobo, C. Queiroz, and J. Dias, "Vertical world feature detection and mapping using stereo vision and accelerometers", in *Proc. 9th Int. Symp. on Intelligent Robotic Systems*, Toulouse, France, July 2001, pp.229-238.

[7] J. Lobo and J. Dias. "Integration of Inertial Information with Vision towards Robot Autonomy", in *Proc. IEEE Int. Symp. on Industrial Electronics*, Guimaraes, Portugal, July 1997, pp 825-830.

[8] H.K. Nishihara, "Real-time implementation of a signcorrelation algorithm for image-matching", (Draft) Teleos Research, February, 1990.

[9] J. Webb, "Implementation and Performance of Fast Parallel Multi-baseline Stereo Vision", in *Proc. Image Understanding Workshop*, 1993, pp.1005-1012.

[10] L.H. Matthies, "Stereo vision for planetary rovers: stochastic modelling to near real time implementation", *Int. Journal of Computer Vision*, 1992, 8(1), pp.71-91.

[11] O. Faugeras et al, "Real time correlation based stereo: algorithm, implementations and applications", Research Report 2013, INRIA Sophia-Antipolis, 1993.

[12] T. Kanade, H. Kano, and S. Kimura, "Development of a video-rate stereo machine", in *Proc. Int. Robotics and System Conf.*, Pittsburg (PA), 1995.

[13] K. Konolige, "Small Vision Systems: Hardware and Implementation", *8th Int. Symp. on Robotics Research*, Hayama, Japan, October 1997.

[14] T. Viéville and O.D. Faugeras, "Cooperation of the Inertial and Visual Systems", in T.C. Henderson, editor, *Traditional and NonTraditional Robotic Sensors*, volume F 63 of *NATO ASI*, SpringerVerlag, 1990, pp.339-350.

[15] T. Viéville, et. al., "Autonomous navigation of a mobile robot using inertial and visual cues", in M. Kikode, T. Sato, and K. Tatsuno, editors, *Intelligent Robots and Systems*, Yokohama, 1993.

[16] T. Viéville, E. Clergue, and P.E.D. Facao, "Computation of ego-motion and structure from visual an inertial sensor using the vertical cue", in *ICCV93*,1993, pp. 591-598.

[17] B. Bhanu, B. Roberts, and J. Ming, "Inertial Navigation Sensor Integrated Motion Analysis for Obstacle Detection", in *Proc. IEEE Int. Conf. on Robotics and Automation*, Cincinnati, Ohio, USA, 1990, pp.954-959.

[18] F. Panerai and G. Sandini, "Oculo-Motor Stabilization Reflexes: Integration of Inertial and Visual Information". *Neural Networks*, 11(7-8), 1998, pp.1191-1204.

[19] F. Panerai, G. Metta, and G. Sandini, "Visuo-inertial stabilization in space-variant binocular systems". *Robotics and Autonomous Systems*, 30(1-2), 2000, pp.195-214.

[20] T. Mukai and N. Ohnishi, "Object Shape and Camera Motion Recovery Using Sensor Fusion of a Video Camera and a Gyro Sensor", *Information Fusion*, 1(1), 2000, pp.45-53.

1992.

[21] G. Xu and Z. Zhang, *Epipolar Geometry in Stereo, Motion and Object Recognition - A unified approach*, Kluwer Academic publishers, 1996, ISBN 0-7923-4199-6.