World feature detection and mapping using stereovision and inertial sensors

Jorge Lobo, Carlos Queiroz, Jorge Dias*

Institute of Systems and Robotics, University of Coimbra, Polo II, 3030-290 Coimbra, Portugal

Abstract

This paper explores the fusion of inertial information with vision for 3D reconstruction. A method is proposed for vertical line segment detection and subsequent local geometric map building. Visual and inertial sensing are two sensory modalities that can be explored to give robust solutions on image segmentation and recovery of 3D structure from images, increasing the capabilities of autonomous vehicles and enlarging the application potential of vision systems. From the inertial sensors, a camera stereo rig, and a few system parameters we can recover the 3D parameters of the ground plane and vertical lines. The homography between stereo images of ground points can be found. By detecting the vertical line segments in each image, and using the homography of ground points for the *foot* of each segment, the lines can be matched and reconstructed in 3D. The mobile robot then maps the detected vertical line segments in a world map as it moves. To build this map an outlier removal method is implemented and a statistical approach used, so that a simplified metric map can be obtained for robot navigation. © 2003 Elsevier Science B.V. All rights reserved.

Keywords: Vision; Inertial sensors; Sensor fusion; 3D reconstruction; Outlier removal

1. Introduction

Robot mapping focuses on the problem of acquiring models of the physical environment, and has been considered one of the important subjects for the advance of autonomous mobile robotic systems. Mobile robots will perform given tasks if they can keep track of relevant objects' state in the environment, especially in dynamic environments. This paper explores the fusion of inertial information with vision to recover 3D structures in the environment.

Research started in the 1980s, and the most common representations of space adopted were either *metric* or *topological*. A *metric map* represents the environment according to the absolute geometric position of the objects, capturing the geometric prop-

erties of the environment. A *topological map* is an abstract representation that describes relationships and connectivity among features of the environment, without any absolute reference system.

More recent approaches aim to deal with the uncertainty in the constructed map and are known as probabilistic techniques. Amongst these, the Occupancy Grid map, first introduced by Moravec and Elfes [1], has been widely used [2–7]. An alternative mapping was proposed by Chatila and Laumond [8] to describe the environment geometry by polyhedron. In our previous work [9] we used occupancy grids and vision to represent the environment as a two-dimensional array of cells, each of which indicates the probability of being occupied. The input measurements for these maps are provided by sonar range-finders sensors and a stereo vision system. This type of map is suitable for imprecise range sensors, but for other sensors like laser range finders other probabilistic techniques

---

* Corresponding author.
E-mail address: jorge@isr.uc.pt (J. Dias).

have been proposed [10–13]. Geometric feature maps, based on line features, have been explored by many researchers [14–21].

Many of these techniques have correspondent interest in the computer vision literature and the connections are still rather under-explored [9,22–24]. This paper explores the combination of inertial cues with stereo vision for building a *metric map* of world vertical features.

The proposed method detects vertical line segments in the world and maps them in a *metric map*. From inertial measurements one can only determine an estimate for linear acceleration and angular velocity. By detecting gravity, the inertial sensors provide a vertical reference and an artificial horizon. In the following sections we show how this can be combined with stereo image data to map 3D vertical features and determine the ground plane parameters.

This work is part of ongoing research into the fusion of inertial sensor data in computer vision systems. In [25,26] the inertial sensors and system prototype are described and results presented for ground plane segmentation, in [27] a method is proposed for camera focal distance calibration using a vanishing point and the vertical reference, in [28] the vertical line segmentation method is described in more detail and some preliminary results presented. In this paper the focus is on the 3D reconstruction and mapping of the detected vertical line segments.

The paper is organized as follows: Section 2 describes the system's geometric model. The geometric features extracted from the inertial sensor data, the vertical, the horizon and the homography of ground plane image points, are discussed in this section. Section 3 discusses a method for image segmentation of image lines. In Section 4, we describe how the correspondence problem between image features is solved. In Section 5.1 the feature mapping method and an outlier removal process is described. The experimental setup used and experimental results are presented in Section 6.

2. System parameters and inertial data

The vision system used for this work has an inertial unit at the middle of the stereo camera baseline, as seen in Fig. 1. The system was placed on a wheeled mobile robot for our tests. Our system is completely platform independent, and can be used, for instance, on legged robots, aerial robots and hand-held devices. The cameras' pan is controlled so as to have a symmetric verge angle θ and the system attitude is obtained from the inertial sensor data.

The system's coordinate frame of reference, $\{\mathcal{C}\}$, is defined as having the origin at the center of the baseline, with length b, of the stereo cameras.

2.1. Gravity vector defines vertical reference and image horizon

The measurements $\boldsymbol{a}$ taken by the inertial unit's accelerometers include the sensed gravity vector $\boldsymbol{g}$ summed with the body's acceleration $\boldsymbol{a}_b$. Assuming the system is motionless, $\boldsymbol{a}_b = 0$, the measured acceleration $\boldsymbol{a}$ gives the gravity vector $\boldsymbol{g}$ in the system's
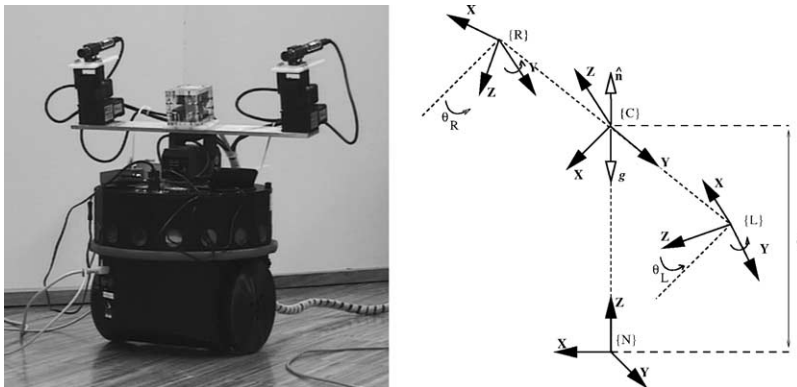

Fig. 1. Vision and inertial system on mobile robot, and system frames of reference.

frame of reference. 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

$$^{\mathcal{C}}\hat{\boldsymbol{n}} = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} = -\frac{\boldsymbol{g}}{\|\boldsymbol{g}\|} = \frac{1}{\sqrt{a_x^2 + a_y^2 + a_z^2}} \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} \quad (1)$$

and $^{\mathcal{C}}\hat{\boldsymbol{n}}$ provides a vertical reference and can be mapped to the left and right camera frames of reference $\{\mathcal{L}\}$ $\{\mathcal{R}\}$. From projective geometry we know that any image point belonging to the horizon line has a projective ray parallel to any leveled plane and orthogonal to the vertical reference [29]. The horizon line and the coordinates of the image point are related by

$$n_x u + n_y v + n_z f = 0, \quad (2)$$

where (u, v) are image coordinates and f the camera's focal distance. Any vanishing point obtained from parallel lines of a leveled plane will belong to the horizon. Therefore from (2) the camera focal distance f can be estimated with just one vanishing point [27].

2.2. Robot navigation frame of reference

With the system mounted on the mobile robot, all detected features can be charted onto the robot's world map and updated as the robot moves in its environment. It is not convenient to construct this map in the system frame of reference $\{\mathcal{C}\}$, and a better choice is to convert the points to a robot navigation frame of reference $\{\mathcal{N}\}$, see Fig. 1. The vertical unit vector $\hat{\boldsymbol{n}}$ and the sensor system height d can be used to define $\{\mathcal{N}\}$. If we choose $^{\mathcal{N}}\hat{\boldsymbol{x}}$ to be coplanar with $^{\mathcal{C}}\hat{\boldsymbol{x}}$ and $^{\mathcal{C}}\hat{\boldsymbol{n}}$, in order to keep the same heading, any world point P will have the following coordinate frame transformation

$$^{\mathcal{N}}\boldsymbol{P} = {}^{\mathcal{N}}\boldsymbol{T}_{\mathcal{C}} \cdot {}^{\mathcal{C}}\boldsymbol{P} \quad (3)$$

with

$$^{\mathcal{N}}\boldsymbol{T}_{\mathcal{C}} = \begin{bmatrix} \sqrt{1-n_x^2} & \dfrac{-n_x n_y}{\sqrt{1-n_x^2}} & \dfrac{-n_x n_z}{\sqrt{1-n_x^2}} & 0 \\ 0 & \dfrac{n_z}{\sqrt{1-n_x^2}} & \dfrac{-n_y}{\sqrt{1-n_x^2}} & 0 \\ n_x & n_y & n_z & d \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (4)$$

where d is the system height from the ground plane. In some applications d can be know or imposed by the physical mount, but it can also be determined by the stereo system through a process of visual fixation, see [30,31] for details.

If a heading reference is available, then $\{\mathcal{N}\}$ should not be restricted to having $^{\mathcal{N}}\hat{\boldsymbol{x}}$ coplanar with $^{\mathcal{C}}\hat{\boldsymbol{x}}$ and $^{\mathcal{C}}\hat{\boldsymbol{n}}$, but use the known heading reference. In scenes of man made environments, such as buildings, some vanishing points can provide a heading reference. Using a vanishing point $\hat{\boldsymbol{m}} = (m_x, m_y, m_z)$ as a reference we get

$$^{\mathcal{N}}\boldsymbol{T}_{\mathcal{C}} = \begin{bmatrix} m_x & n_y m_z - n_z m_y & n_x & -n_x d \\ m_y & n_z m_x - n_x m_z & n_y & -n_y d \\ m_z & n_x m_y - n_y m_x & n_z & -n_z d \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1}. \quad (5)$$

The update of $\{\mathcal{N}\}$ as the robot changes its position is beyond the scope of this work, where $\{\mathcal{N}\}$ is just relative to the robot's position, and not truly world fixed. Anyhow, using the robot's odometry, the inertial sensors and landmark matching, the update can be accomplished.

2.3. Homography

The direct mapping or homography of ground plane points between the stereo pair can be obtained by calibration using know ground plane points [32]. But knowing the geometric parameters of the ground plane and the geometry of the stereo rig, we can also compute this homography [28]. Since the points belong to a single plane, the ground plane, we can deduce a mapping between the left and right image points belonging to the ground plane. Considering the right and left camera homographies, $\boldsymbol{H_r}$ and $\boldsymbol{H_l}$, the direct mapping $\boldsymbol{H}$ of ground plane points between the stereo pair can be obtained by

$$s\boldsymbol{p}_{li} = \boldsymbol{H} \cdot \boldsymbol{p}_{ri} = \boldsymbol{H_l} \cdot \boldsymbol{H_r}^{-1} \cdot \boldsymbol{p}_{ri}, \quad (6)$$

where $\boldsymbol{p}_{ri}$ and $\boldsymbol{p}_{li}$ are right and left projective image points, and s an arbitrary scale factor. For our system geometry, $\boldsymbol{H}$ is given by

$$\boldsymbol{H} = \begin{bmatrix} -\dfrac{2n_x b\cos\theta\sin\theta + n_y b + 2d - 4d\cos^2\theta}{-n_y b + 2d} & \dfrac{2bn_z\cos\theta}{-n_y b + 2d} & f\dfrac{-2\cos\theta(2d\sin\theta + bn_x\cos\theta)}{-n_y b + 2d} \\ 0 & 1 & 0 \\ 2\dfrac{2d\sin\theta\cos\theta - n_x b + n_x b\cos^2\theta}{f(-n_y b + 2d)} & \dfrac{2bn_z\sin\theta}{f(-n_y b + 2d)} & -\dfrac{2n_x b\cos\theta\sin\theta + n_y b - 4d\cos^2\theta + 2d}{-n_y b + 2d} \end{bmatrix}.$$

$$(7)$$

This mapping $\boldsymbol{H}$ will be fundamental for the stereo correspondence used in Section 4.

3. Image line segmentation

Knowing the vertical, the vanishing point of all image lines that correspond to world vertical features is known. This vanishing point is at infinity when there is no tilt, and the vertical lines are all parallel in the image. For small tilt values, the vertical lines can be taken as parallel. Based on this assumption, the vertical line segments found in the image will be parallel to the local image vertical $\hat{\boldsymbol{n}}_i$, the normalized image projection of the vertical $\hat{\boldsymbol{n}}$. The image vertical reference corresponds to the unit sphere projection of the vanishing point of all 3D vertical lines in the image plane.

In order to detect vertical line segments we extracted the edges in the image using a modified Sobel filter proposed by Jahne [33] that uses different coefficients to obtain a lower angle error in the gradient. By choosing an appropriate threshold for the gradient magnitude, the potential edge lines can be identified.

To obtain the vertical edges we compare the pixel gradient with the vertical. The dot product of the gradient with the vertical should be null, so by setting a tolerance threshold value the detected edge points can be taken as vertical or not.

$$\boldsymbol{\mathcal{D}} \cdot \hat{\boldsymbol{n}}_i < \text{tolerance}. \tag{8}$$

In order to extract the vertical lines in the image, all edge points that satisfied (8) were mapped to a rectified image table using (9), so that continuity could be tested along the vertical edge direction. Each edge point $\boldsymbol{p}_i = (u, v)$ contributed to the table at position

$$\text{vert\_points}(x, y) = (\boldsymbol{p}_i \cdot \hat{\boldsymbol{h}}, \boldsymbol{p}_i \cdot \hat{\boldsymbol{n}}), \tag{9}$$

where $\hat{\boldsymbol{h}}$ is the horizontal unit vector, perpendicular to $\hat{\boldsymbol{n}}$ in the image plane, i.e.,

$$\hat{\boldsymbol{n}}_i \cdot \hat{\boldsymbol{h}}_i = 0. \tag{10}$$

The minimum line length and allowable gaps is set and each column of the table parsed. The end result is a set of lines, given by their end-points in the original image. The parameters that need to be set are the gradient magnitude and angle tolerance thresholds, and the minimum line length and tolerated gap size.

We implemented this method with our system, working real time at 10 frames per second. Fig. 2 shows an example of the results obtained. The results showed that the method performs well in man made environments where vertical lines segments are abundant, but required some parameter adjustment to have good results with different types of scenes. The gradient magnitude threshold used to identify edges is sensitive to image lighting and contrast. The line length and gap size that work well in an indoor structured environment might not work well in highly textured but less structured outdoor environments.

Fig. 2. Vertical line detection.

4. 3D reconstruction

In the previous section a method was proposed for vertical image line detection. But in order to find out if they truly correspond to 3D vertical line segments they have to be matched across the stereo image pair. Making the assumption that the relevant vertical line segments start from the ground plane, and since we know the homography of the ground plane image points, a common unique point is identified. The lower point or *foot* of each vertical feature in one image should map to the corresponding *foot* in the other image.

An algorithm for the 3D reconstruction of vertical line segments can now be presented. For each detected line in the right image, map its *foot* to the other image using (7). The correspondent point and its neighborhood in the left image can then be tested for a match with the original point of interest in the right image by parsing all the left image lines, determining the corresponding left image point p_{li} with

$$p_{li} = H \cdot p_{ri}, \tag{11}$$

where p_{ri} is the *foot* of the right image line, and testing an allowed neighborhood window $p_{li} \pm \delta$ for the *foot* of a left image vertical line. If there is a match, the point belongs to the ground plane and must be the *foot* of a true 3D vertical world line segment. The 3D position $^{\mathcal{N}}P$ of the *foot* of this vertical line segment is given by

$$^{\mathcal{N}}\tilde{P} = H_r^{-1} p_{ri}, \tag{12}$$

where $^{\mathcal{N}}\tilde{P} = [\, X \quad Y \quad 1 \,]$ and $^{\mathcal{N}}P = [\, X \quad Y \quad 0 \quad 1 \,]$.

With the system mounted on the mobile robot, the vertical line segments can be charted on a world map, constructed as the robot moves in its environment. This map is constructed in the robot's navigation frame of reference $\{\mathcal{N}\}$ previously defined.
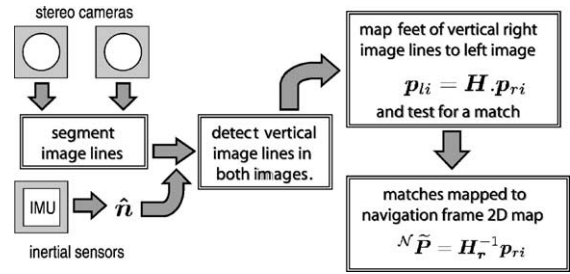

Fig. 3. 3D reconstruction algorithm.

Fig. 3 summarizes the implemented method, and Fig. 4 shows an example of the results obtained.

4.1. Error analysis

The mapped features' accuracy depends on the estimated homography parameters and the detected lower end points of the image vertical line segment. From (7) it is clear that H will be sensitive to small variations of the vertical reference provided by the inertial sensors. When analyzing estimates of directions or surface normals, the statistic distribution on a sphere provides a good model, see [34] and sensor statistics study in Section 6.3.

A measurement set from the accelerometers gives the acceleration vector in Cartesian coordinates. Spherical polar coordinates (θ, ϕ) can be used to represent the direction of the acceleration vector. The correspondent direction cosines l, m and n are given by

$$l = \sin\theta\cos\phi, \qquad m = \sin\theta\sin\phi,$$
$$n = \cos\theta, \quad 0 < \theta < \pi, \ 0 < \phi < 2\pi. \tag{13}$$

Let (l_i, m_i, n_i), $i = 1, \ldots, k$, be k observed directions. The spherical mean direction $(\bar{l}_0, \bar{l}_0, \bar{l}_0)$ is
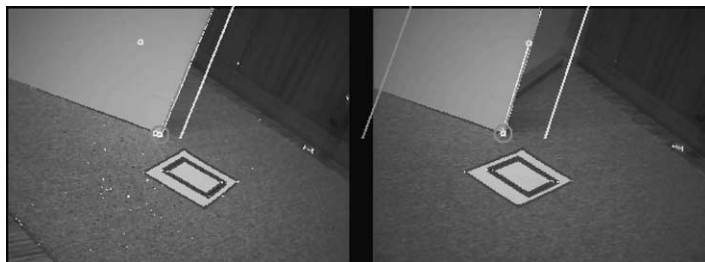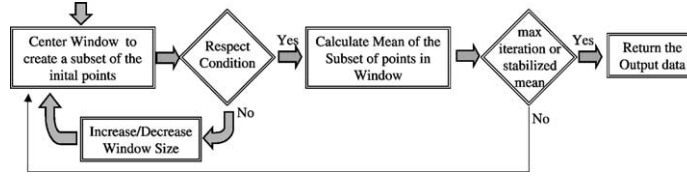

Fig. 4. 3D vertical line segment reconstruction.

Fig. 5. Outlier removal algorithm.

given by

$$\bar{l}_0 = \sum_{i=1}^{k} \frac{l_i}{R}, \qquad \bar{m}_0 = \sum_{i=1}^{k} \frac{m_i}{R}, \qquad \bar{n}_0 = \sum_{i=1}^{k} \frac{n_i}{R}, \quad (14)$$

where R is the length of the resultant given by

$$R = \left\{ \left(\sum l_i\right)^2 + \left(\sum m_i\right)^2 + \left(\sum n_i\right)^2 \right\}^{1/2}. \quad (15)$$

The value R is a measure of concentration about a mean direction, and the spherical variance is defined as

$$S^* = \frac{k - R}{k}, \quad 0 \le R \le k, \ 0 < S^* < 1. \quad (16)$$

R and S can then be used to describe the error statistics of the vertical reference provided by the inertial sensors. These two values are critical for our application and results of our tests are shown in Section 6.

5. Mapping of detected features

The vertical line segment detection method produces some outliers. Before using it to update the world map, the detected feature data must be filtered to remove the outliers.

5.1. Outlier removal

To deal with the outliers, a fast computational method was developed in this work. This method is iterative but with a simple computational implementation for real time performance. Since the system detects several vertical line segments in the scene, the method starts by clustering the measurements on several sub-sets of vertical line segments. For each of these sub-sets the method removes all outlier measurements.

The process starts by searching the complete measurement set M to determine a sub-set of measures S, with the biggest cardinality possible that is within an initial window size, providing the currently best estimate of one vertical line segment. After removing the sub-set S from measurement set M, the process continues until the measurement set is empty or has minimal cardinality. Fig. 5 summarizes the outlier removal method.

The measurements in the set M are points $^{\mathcal{N}}\tilde{P} = [\, X \quad Y \quad 1 \,]$ obtained from (12). The outlier removal method explores the bi-dimensionally of the data and searches the measurement set M by using a *window*. The *window* location changes its position and dimension during the searching process and all points in the measurement set M are tested in each iteration.

When two vertical features fall within the same initial windows size, they are detected as a single feature. If the distance between two vertical lines is less than a threshold the outlier removal method assumes there is only one vertical line. The threshold used is given by a predefined ratio between two distances: from the system to the verticals, and between the two vertical lines.

This method was compared with the RANSAC [35] method, and the results of both methods are similar but the RANSAC method is computationally slower.

After the outlier filtering the different sub-sets of measurements are mapped in the world map representation. Each sub-set represents an estimation of one vertical feature.

6. Results

6.1. Experimental setup

For the experiments we used the system in Fig. 6 briefly described in Section 2. In this system an inertial system prototype built at our lab was used [26]. The system is based on low-cost inertial sensors and is
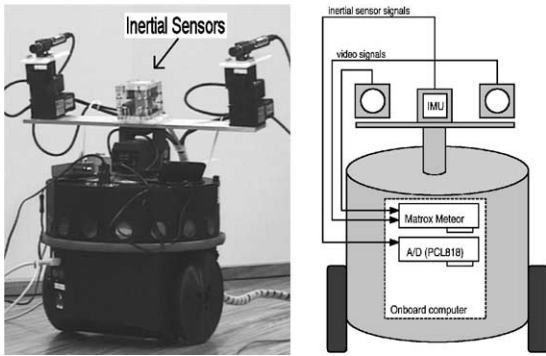
Fig. 6. Vision system with inertial sensors, and system architecture.

intended for robotic applications. The sensors used in the prototype system include three accelerometers and three gyroscopes.

The inertial system prototype was mounted onto the active vision system as seen in Fig. 6. This figure shows the architecture of the complete system, with a

large baseline to obtain better precision on the stereo reconstruction.

6.2. 3D vertical line detection and mapping

With this experimental setup, the 3D vertical line segment detection algorithm was tested. Fig. 7 shows the stereo image pair with the identified image vertical lines and the matched *feet* of the true 3D vertical lines for a structured real scene. Fig. 8 shows the map obtained without moving the robot. The robot is represented by the semi-circle on the left side of the map, the points represent the *feet* of the detected vertical line segments. The solid lines represent the ground truth obstacles in the world (all ground truth distances were tape measured). The *clouds* of points at the line corners represent the *feet* of the detected vertical lines. These sets of points are characterized by small ellipses representing the uncertainty of the set. Fig. 9 shows the results of the robot moving along a
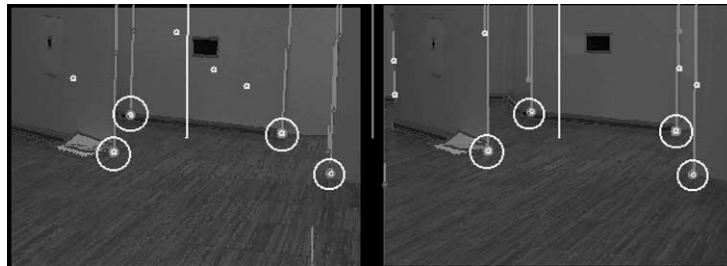

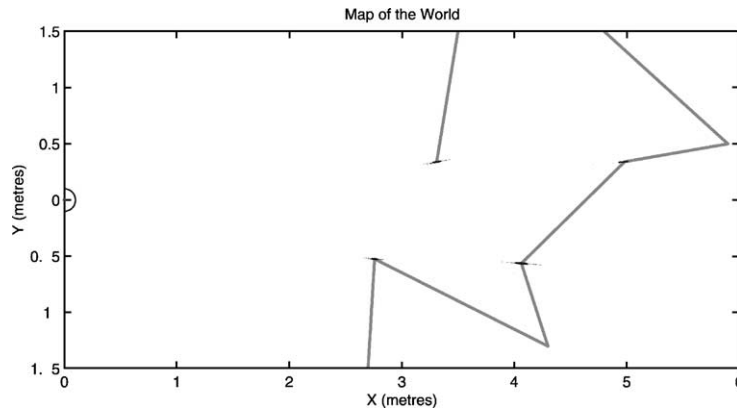Fig. 7. Stereo images with detected edges and vertical lines.

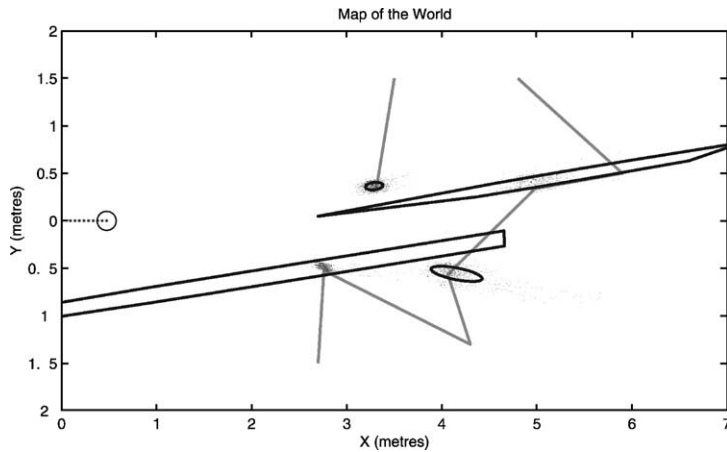Fig. 8. World map with robot motionless.

Fig. 9. World map with robot moving without outlier removal.

trajectory without outlier removal. As expected the results are more unstable due to vibration and oscillation.

The uncertainty increases when the robot moves along a trajectory, because the vision system suffers mechanical vibrations, and the robot's odometry introduces additional errors, as can be seen in Fig. 9. The distribution pattern clearly suggests that the *X* coordinate is more unstable. This is due to the accelerometer sensitivity and noise, and implies that more filtering of the signal is required. The vision algorithm is also more prone to errors in the *X*-direction due to the stereo geometry. From the results we ob-

serve that the uncertainty ellipses of the vertical line segments are quite large. This is due to the fact that there are many data outliers generated when the robot moves along a trajectory.

To remove the outliers we applied the iterative method described in Section 5.1. The results, after removing the outliers with this method, are shown in Fig. 10. The results are better since the uncertainty ellipses are smaller.

Fig. 11 shows a second test made in a less structured real scene. In this test, we only intend to see how our system works in a less structured scene and with the robot stopped.
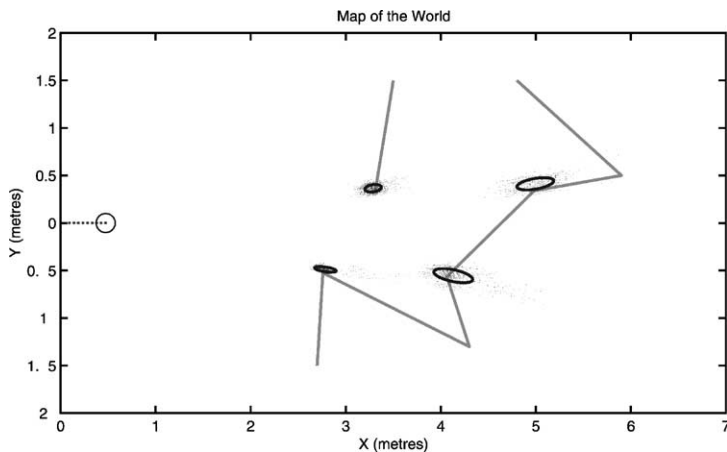

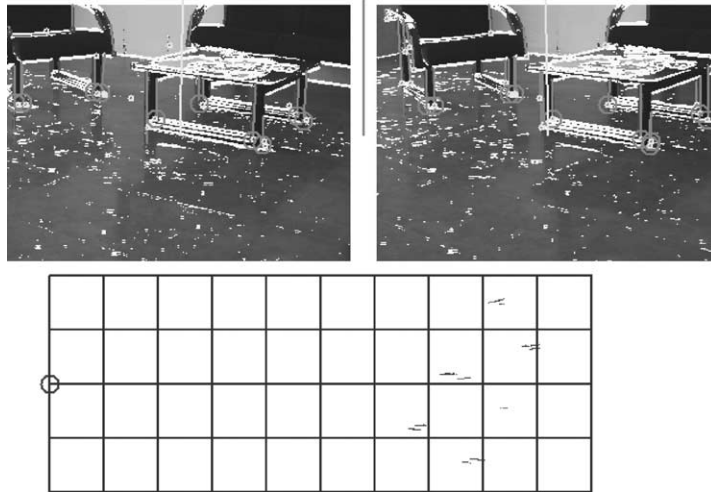Fig. 10. World map with robot moving with outlier removal.

Fig. 11. Stereo images with detected edges and vertical lines, and raw map of vertical world features.

The detected verticals for the scene were mapped as shown in Fig. 12. The lines represent the ground truth position of the table and chair in the world. In Fig. 12 we observe that the uncertainty levels are bigger than in the previous situation. These results were expected, because in less structured scenes there is more noise in edge detection than in a structured real scene.

The last test used an unstructured scene seen in Fig. 13. A chair was placed in a natural environment with plants and vases. The robot was initially at rest and then set in motion, the results are shown in Figs. 14 and 15. The lines show the ground truth chair position, the nearby circle a plant vase that occluded one of the chair legs. The robot was placed on higher ground, and

the detected features were very near, therefore the error spread along the *X*-direction is less than in previous tests. Fig. 16 shows the result after the outlier removal process and a slight reduction of the uncertainty ellipse can be observed.

6.3. Sensor statistics

To study the effect of sensor noise we characterized the noise model of the accelerometers. The statistical distribution of the accelerometer data on a sphere was analyzed as described in Section 4. Samples were taken with the robot stopped and in motion. Table 1 shows the results.
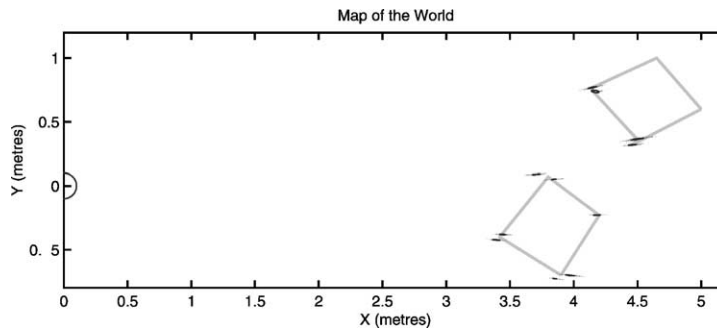

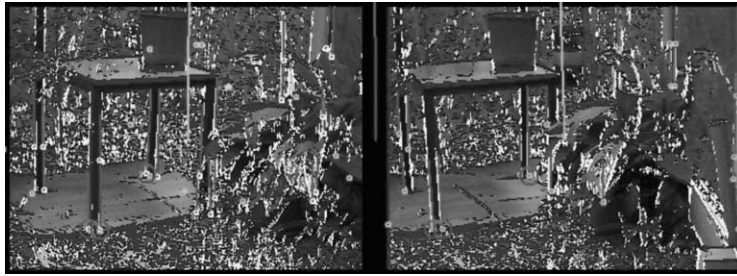Fig. 12. World map with the robot motionless.

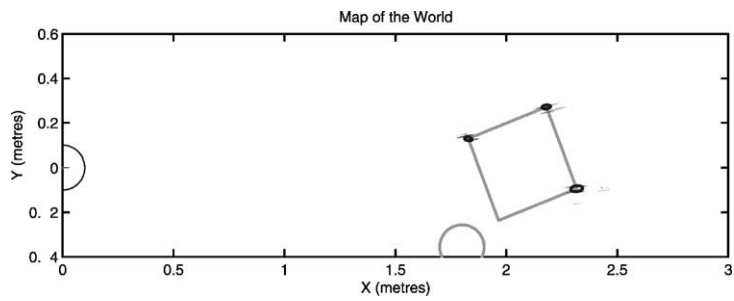Fig. 13. Stereo images with detected edges and vertical lines.

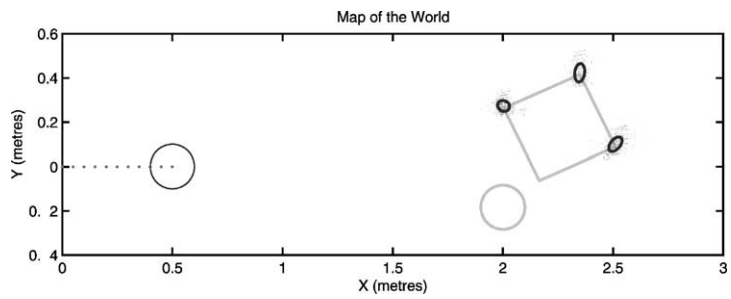Fig. 14. World map with robot motionless.

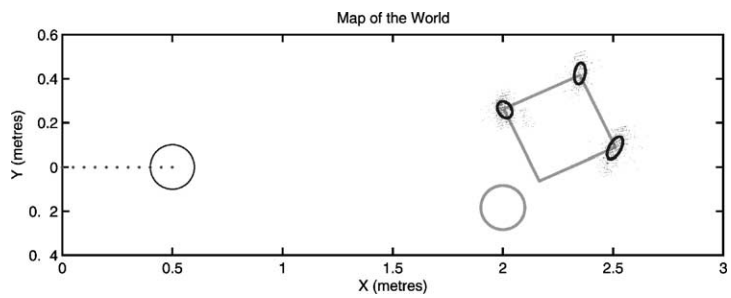Fig. 15. World map with robot moving without outlier removal.

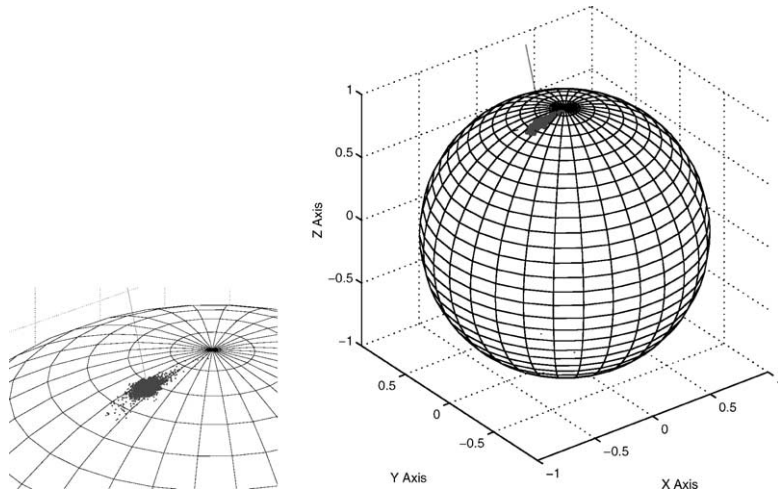Fig. 16. World map with robot moving with outlier removal.

Fig. 17. Accelerometer data—Robot moving.

Table 1
Accelerometer data

Robot	R/samples	$S^*(\times 10^{-4})$
Stopped	0.99972	2.7935
Moving	0.99898	9.9038

From Table 1 it is clear that the robot motion adds noise to the measurements, degrading the accuracy of the vertical feature mapping. From Fig. 17 we can see that with the system motionless the error distribution is close to isotropic, but when the robot is in motion a longer error spreads along the X-direction. This will propagate to the error along the same direction in the final mapping of the detected features, as shown in Section 6.

7. Conclusions

This paper presented our recent results on the use and integration of two different sensor modalities, inertial sensing and vision, for 3D reconstruction and map-building. The integration of inertial and visual information is used to detect world vertical features. The horizontal plane given by the inertial sensors was used to establish correspondence between stereo image vertical lines, enabling the detection of true 3D world vertical line segments. These features are used to build a *metric map* useful to improve mobile robot autonomy and navigation.

This approach permits a real-time implementation. The results show that it works with less structured environments focusing well on the few structured scene elements.

Future work will aim at improving the robustness of the feature detection method. By using the vertical vanishing point and not just the vertical reference, the method will work for any camera rotation. Other restrictions on the detected vertical features can be explored, such as testing for different heights, and not just there *feet*. Using the gyro sensor data to de-couple body acceleration and gravity, the static restriction can also be removed. The map building process can be improved so that it can be used autonomously by the robot. Other applications for inertial and vision sensor fusion will be explored, such as walking robots or wearable sensing devices for human–machine interface and augmented reality.

References

[1] H. Moravec, A. Elfes, High resolution maps from wide angle sonar, in: Proceedings of the IEEE International Conference on Robotics and Automation, 1985, pp. 116–121.
[2] S. Wei, Y. Yagi, M. Yachda, Building local floor map by use of ultrasonic and omni-directional vision sensor, in: Proceedings of the IEEE International Conference on Robotics and Automation, 1998, pp. 2548–2553.

[3] B. Yamauchi, A. Schultz, W. Adams, Mobile robot exploration and map-building with continuous localization, in: Proceedings of the IEEE International Conference on Robotics and Automation, 1998, pp. 3715–3720.

[4] J. Borenstein, Y. Koren, The vector field histogram—fast obstacle avoidance for mobile robots, IEEE Journal of Robotics and Automation 7 (3) (1991) 278–288.

[5] J. Buhmann, W. Burgard, A.B. Cremers, D. Fox, T. Hofman, F. Schneider, J. Strikos, S. Thrun, The mobile robot Rhino, AI Magazine 16 (1) (1995).

[6] W. Burgard, A.B. Cremers, D. Fox, D. Hahnel, G. Lakemeyer, D. Schulz, W. Steiner, S. Thrun, Experiences with an interactive museum tour-guided robot, Artificial Intelligence 114 (1–2) (1999) 3–55.

[7] B. Yamauchi, P. Langley, Place recognition in dynamic environments, Journal of Robotics Systems 14 (2) (1997) 107–120.

[8] R. Chatila, J.-P. Laumond, Position referencing and consistent world modelling for mobile robots, in: Proceedings of the IEEE International Conference on Robotics and Automation, 1985.

[9] A. Silva, P. Menezes, J. Dias, H. Araujo, Grid-Based framework for sensorial data integration in mobile robots, in: Proceedings of the SPIE's International Symposium on Intelligent Systems & Advanced Manufacturing—Special Session on Sensor Fusion and Decentralized Control in Autonomous Robotic Systems, Pittsburgh, PA, October 14–15, 1997.

[10] P. Newman, On the structure and solution of the simultaneous localisation and map building problem, Ph.D. Thesis, Australian Centre for Field Robotics, University of Sidney, Sidney, Australia, 2000.

[11] J.J. Leonard, H.F. Durrant-Whyte, I.J. Cox, Dynamic map building for an autonomous mobile robot, International Journal of Robotics Research 11 (4) (1992) 89–96.

[12] H. Durrant-Whyte, S. Majumder, S. Thrun, M. deBattista, S. Scheding, A Bayesian algorithm for simultaneous localization and map building, in: Proceedings of the 10th International Symposium on Robotics Research (ISRR'01), Lorne, Australia, 2001.

[13] J. Guivant, E. Nebot, H. Durrant-Whyte, Simultaneous localization and map building using natural features in outdoor environments, in: Proceedings of the International Symposium on Intelligent Autonomous Systems 6 (IAS-6), Venice, Italy, IOS Press, 2000, pp. 581–588. ISBN 1 58603 078 7

[14] J. Vandorpe, H. Van Brussel, H. Xu, Exact dynamic map building for a mobile robot using geometrical primitives produced by a 2D range finder, in: Proceedings of the IEEE International Conference on Robotics and Automation, 1996, pp. 901–908.

[15] J. Gonzalez, A. Ollero, A. Reina, Map building for a mobile robot equipped with a 2D laser rangefinder, in: Proceedings of the IEEE International Conference on Robotics and Automation, 1994, pp. 1904–1909.

[16] D.J. Kriegman, E. Triendl, T.O. Binford, Stereo vision and navigation in buildings for mobile robots, IEEE Transactions on Robotics and Automation 5 (6) (1989) 792–803.

[17] J.L. Crowley, P. Bobet, K. Sarachik, Dynamic world modeling using vertical line stereo, Journal of Robotics and Autonomous Systems, Elsevier Press, June 1991.

[18] K.O. Arras, N. Tomatis, B.T. Jensen, R. Siegwart, Multisensor on-the-fly localization: precision and reliability for applications, Robotics and Autonomous Systems 34 (2001) 131–143.

[19] J.A. Castellanos, J.M.M. Montiel, J. Neira, J.D. Tardos, The SPmap: a probabilistic framework for simultaneous localization and map building, IEEE Transactions on Robotics and Automation 15 (5) (1999) 948–953.

[20] L. Mathies, S.A. Shafer, Error modeling in stereo navigation, IEEE Journal Robotics Automation 3 (3) (1987) 239–248.

[21] S. Shah, J.K. Aggarwal, Modeling Structured Environments Using Robot Vision, Lecture Series in Computer Science: Recent Progress in Computer Vision, 1995, pp. 113–128.

[22] P.K. Allen, I. Stamos, Integration of range and image sensing for photorealistic 3D modeling, in: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2000, pp. 1435–1440.

[23] R. Bajcsy, G. Kamberova, L. Nocera, 3D Reconstruction of environments for virtual reconstruction, in: Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision, 2000.

[24] H. Shum, M. Han, R. Szeliski, Interactive construction of 3D models from panoramic mosaics, in: Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR), 1998.

[25] J. Lobo, J. Dias, Integration of inertial information with vision towards robot autonomy, in: Proceedings of the IEEE International Symposium on Industrial Electronics, Guimaraes, Portugal, July 1997, pp. 825–830.

[26] J. Lobo, J. Dias, Ground plane detection using visual and inertial data fusion, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems—IROS'98, Victoria, Canada, October 1998, pp. 912–917.

[27] J. Lobo, J. Dias, Fusing of image and inertial sensing for camera calibration, in: Proceedings of the International Conference on Multisensor Fusion and Integration for Intelligent Systems, MFI 2001, Baden-Baden, Germany, August 2001, pp. 103–108.

[28] J. Lobo, C. Queiroz, J. Dias, Vertical world feature detection and mapping using stereo vision and accelerometers, in: Proceedings of the Ninth International Symposium on Intelligent Robotic Systems—SIRS'01, Toulouse, France, July 2001, pp. 229–238.

[29] K. Kanatani, Geometric Computation for Machine Vision, Oxford University Press, Oxford, 1993. ISBN 0-19-856385-X.

[30] J. Dias, C. Paredes, I. Fonseca, A.T. de Almeida, Simulating pursuit with machines, in: Proceedings of the IEEE Conference on Robotics and Automation, Japan, 1995, pp. 472–477.

[31] J. Dias, C. Paredes, I. Fonseca, H. Araujo, J. Baptista, A.T. de Almeida, Simulating pursuit with machine experiments with robots and artificial vision, IEEE Transactions on Robotics and Automation 3 (1) (1998) 1–18.

[32] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, 2000. ISBN 0521623049.

[33] B. Jahne, Digital Image Processing, Springer, Berlin, 1997. ISBN 3-540-62724-3.

[34] K.V. Mardia, Statistics of Directional Data, Academic Press, New York, 1972. ISBN 0-12-471150-2.

[35] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (1981) 381–395.

Carlos Queiroz is a Teaching Assistant at the Department of Informatics in the Polytechnic Institute of Tomar, Portugal, and a Researcher at the Institute of Systems and Robotics, Coimbra, Portugal. He graduated in Electrical Engineering (specialization on Computers) from the University of Coimbra in 1999. He is concluding his M.Sc. degree at the Department of Electrical and Computer Engineering at the University of Coimbra. His current research interests focus on computer vision, sensor fusion, 3D reconstruction, robotics and legged robots.

Jorge Lobo was born on 23 September 1971, in Cambridge, UK. In 1995 he completed his 5-year undergraduate Course in Electrical Engineering at Coimbra University. He was a junior teacher at the Computer Science Department of the Coimbra Polytechnic School, and later joined the Electrical and Computer Engineering Department of the Faculty of Science and Technology of the University of Coimbra where he currently works as a Teaching Assistant. In April 2002 he completed his M.Sc. and his current research is carried out at the Institute of Systems and Robotics, University of Coimbra, and is aimed at the fusion of inertial information with vision systems in mobile robots.

Jorge Dias was born on 7 March 1960, in Coimbra, Portugal. He completed his Electrical Engineering degree (specialization on Computers) by the Faculty of Science and Technology from the University of Coimbra on July 1984 and Ph.D. degree on Electrical Engineering by the University of Coimbra, specialization in Control and Instrumentation, November 1994. His main research area is Computer Vision and Robotics, with activities and contributions on the field since 1984. His current research is carried out at the Institute of Systems and Robotics, University of Coimbra.