Proceedings of the 2003 IEEE
International Conference on Robotics & Automation
Taipei, Taiwan, September 14-19, 2003

# Registration and Segmentation for 3D Map Building
## A Solution based on Stereo Vision and Inertial Sensors

Jorge Lobo, Luis Almeida, João Alves and Jorge Dias
Institute of Systems and Robotics,
University of Coimbra, Portugal,
{jlobo,laa,jalves,jorge}@isr.uc.pt

*Abstract*— **This article presents a technique for registration and segmentation of dense depth maps provided by a stereo vision system. The vision system uses inertial sensors to give a reference for camera pose. The maps are registered using a modified version of the *ICP - Iterative Closest Point* algorithm to register dense depth maps obtained from a stereo vision system. The proposed technique explores the integration of inertial sensor data for dense map registration.**

**Depth maps obtained by vision systems, are very point of view dependent, providing discrete layers of detected depth aligned with the camera. In this work we use inertial sensors to recover camera pose, and rectify the maps to a reference ground plane, enabling the segmentation of vertical and horizontal geometric features and map registration. We propose a real-time methodology to segment these dense depth maps, including segmentation of structures, object recognition, robot navigation or any other task that requires a three-dimensional representation of the physical environment.**

**The aim of this work is a fast real-time system, that can be applied to autonomous robotic systems or to automated car driving systems, for modelling the road, identifying obstacles and roadside features in real-time.**

## I Introduction

The registration of 3D surfaces has applications varying from building terrain maps, or depth maps of sea floor, for autonomous robots, to recognition of objects or to reconciling various medical imaging modalities. This article describes a technique for dense maps registration based on data from inertial sensors and depth maps provided from a stereo vision system, within the context of autonomous robots. The field of robot mapping and localization has been a very active domain but only a few applications have used dense data from vision sensors. Recently many computer vision researchers explored techniques to combine images obtained from different points-of-view, but the relation between techniques from both domains is still somewhat underexploited [1][2]. This article proposes a technique suitable for robot mapping based on data obtained by computer vision techniques. 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 humans and in animals the vestibular system in the inner ear gives inertial information essential for navigation, orientation, body posture control and equilibrium. Neural interactions of human vision and vestibular system occur at a very early processing stage [3][4]. 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.

In our previous work on inertial sensor data integration in vision systems, the inertial data was directly used with the image data [5][6][7]. In this work we use the inertial data to perform a fast segmentation of pre-computed depth maps obtained from the vision system. The map registration technique proposed in this article uses a modified version of the $ICP - IterativeClosestPoint$ algorithm [8]. The technique is a modified approach that explores the inertial sensor data for dense map registration.

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 [9][10][11][12]. The group of T. Kanade at CMU [13] succeeded in producing a video- rate stereo machine based on the multi-baseline stereo algorithm

to generate a dense range map. SRI has developed an efficient implementation of area correlation stereo, the SRI Stereo Engine filtering [14]. We are using this system to obtain real-time depth maps.

The use of inertial systems and vision have been studied by several researchers. Viéville and others proposed the use of an inertial system based on low cost sensors for mobile robots [15][16][17][18]. An inertial sensor integrated optical flow technique was proposed by Bhanu *et al.* [19]. Panerai and Sandini used a low cost gyroscope for gaze stabilization of a rotating camera [20][21]. Mukai and Ohnishi studied the recovery of 3D shape from an image sequence using a video camera and a gyro sensor [22].

In our previous work we have explored the integration of inertial sensor data in vision systems [5][6][7]. In this article we extended this previous work for 3D mapping, using information provided from stereo vision and inertial sensors.

## II Rectified Depth Maps

Before fusing the depth maps, they must be registered to a common referential. This can be done using data fitting alone, or aided by known parameters or restrictions on the way the measurements were made. In order to obtain depth maps with known vision system pose, the stereo vision system was mounted onto an Inertial Measurement Unit (IMU), as shown in fig. 1.

### A. Depth Maps

A stereo vision system can compute depth form triangulation by matching points across the stereo pair. If the cameras are front-parallel and aligned, the disparities can be measured along a single scan line, and the triangulation geometry is simplified.

The classical approach to estimate disparities uses two techniques: feature matching and correlation. In a feature-based algorithm, a set of complex tokens is 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.

Many stereo camera configurations have vergence and do not comply with the front-parallel geometric model. In that case the disparity measurement can be related with the *horopter*. Coombs [23] characterizes horopter as being the surface in three dimensional space defined by the points that stimulate exactly corresponding points (i.e., that has zero

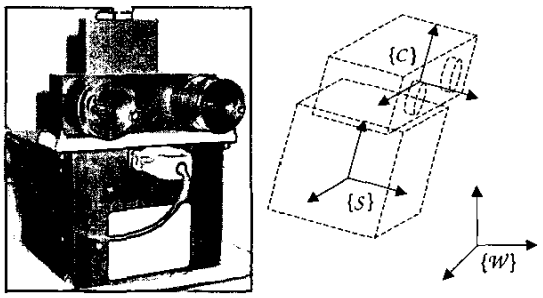


Fig. 1: Stereo Vision System with Inertial Measurement Unit, and frames of reference.

stereo disparity) in the two cameras. Small disparities correspond to small deviations in depth from the horopter. Such disparities can be measured by simple local neighbourhood operators to build up a dense surface map of the environment near the horopter [24][25]. 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. The *rectification* process can be implemented by projecting the original pictures onto the new image plane [26]. With an appropriate choice of coordinate system, the rectified images have scanlines parallel to the baseline and the front-parallel geometry can be applied.

In order to compute range from stereo images we are using the SRI Stereo Engine [14] with the Small Vision System (SVS) Software and the MEGA-D Digital Stereo Head, shown in fig. 1.

### B. Inertial Data

As seen in fig.1, the inertial measurement unit (IMU) was coupled to the stereo vision system, providing valuable data about camera pose and movement. 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, the rotation from the inertial sensors frame of reference, $\{S\}$, to the camera system frame of reference, $\{C\}$, shown in fig. 1, can be determined from the set of camera positions obtained from the calibration and the corresponding data from the inertial sensors.

Having determined the rigid transformation be-

tween the camera and the IMU, the sensed acceleration and rotation are mapped to the camera system frame of reference.

### C. 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 $\boldsymbol{a}_b$:

$$\boldsymbol{a} = -\boldsymbol{g} + \boldsymbol{a}_b \tag{1}$$

Notice that the accelerometer will measure the reactive (upward) force to gravity. Assuming the system is motionless, then $\boldsymbol{a}_b = 0$ and the measured acceleration $\boldsymbol{a} = -\boldsymbol{g}$ gives the gravity vector in the sensor system frame of reference $\{\mathcal{S}\}$. 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

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

By performing the rotation update using the IMU gyro data, gravity can be separated from the sensed acceleration. In this case $\hat{\boldsymbol{n}}$ is given by the rotation update, but must be monitored using the low pass filtered accelerometer signals, for which the above equation still holds, to reset the accumulated drift. This vertical reference, given in the camera system frame of reference, will be used in segmenting the depth maps obtained from the stereo algorithm.

### D. Inertial Frame of Reference

In our experimental setup, the stereo algorithm provides depth maps in the left camera frame of reference, $\{\mathcal{C}\}$. 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 height $d$ 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. For a given point $P$, we have the following coordinate frame transformation

$${}^{\mathcal{W}}\boldsymbol{P} = {}^{\mathcal{W}}\boldsymbol{T}_{\mathcal{C}}.{}^{\mathcal{C}}\boldsymbol{P} \tag{3}$$

where

$${}^{\mathcal{W}}\boldsymbol{T}_{\mathcal{C}} = \begin{bmatrix} \sqrt{1-n_x^2} & \frac{-n_x n_y}{\sqrt{1-n_x^2}} & \frac{-n_x n_z}{\sqrt{1-n_x^2}} & 0 \\ 0 & \frac{n_z}{\sqrt{1-n_x^2}} & \frac{-n_y}{\sqrt{1-n_x^2}} & 0 \\ n_x & n_y & n_z & d \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where $d$ is the system height.

This is obtained as follows. Consider a frame of reference $\{\mathcal{W}_c\}$ with the same origin as $\{\mathcal{C}\}$ and ${}^{\mathcal{W}_c}\hat{\boldsymbol{x}}$ coplanar with ${}^{\mathcal{C}}\hat{\boldsymbol{x}}$ and ${}^{\mathcal{C}}\hat{\boldsymbol{n}}$ in order to keep the same heading. A simple rotation $R$ maps the two frames of reference as follows

$${}^{\mathcal{C}}\boldsymbol{P} = \begin{bmatrix} R & 0 \\ 0 & 1 \end{bmatrix}.{}^{\mathcal{W}_c}\boldsymbol{P} = \begin{bmatrix} \hat{\boldsymbol{r}}_1 & \hat{\boldsymbol{r}}_2 & \hat{\boldsymbol{r}}_3 & 0 \\ & 0 & & 1 \end{bmatrix}.{}^{\mathcal{W}_c}\boldsymbol{P} \tag{5}$$

where $\hat{\boldsymbol{r}}_1$, $\hat{\boldsymbol{r}}_2$ and $\hat{\boldsymbol{r}}_3$ are the X, Y, and Z axis of $\{\mathcal{W}_c\}$ given in the camera $\{\mathcal{C}\}$ frame of reference. But the Z axis of $\{\mathcal{W}_c\}$ is just the vertical given by the inertial sensors:

$${}^{\mathcal{W}_c}\hat{\boldsymbol{z}} = \hat{\boldsymbol{r}}_3 = {}^{\mathcal{C}}\hat{\boldsymbol{n}} = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} \tag{6}$$

But there are infinite possibilities for the $\{\mathcal{W}_c\}$ X and Y axis, since $\hat{\boldsymbol{n}}$ only defines the $XY$ plane, but no heading within this plane. The $X$ axis of $\{\mathcal{W}_c\}$ can be chosen to be coplanar with $\{\mathcal{C}\}$ $X$ and $\hat{\boldsymbol{r}}_3$ axis, keeping the same heading, so we have:

$$\hat{\boldsymbol{r}}_1 = a.\hat{\boldsymbol{x}} + b.\hat{\boldsymbol{r}}_3 = a \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} \tag{7}$$

since $\hat{\boldsymbol{r}}_1$ is a unit vector we have:

$$\|\hat{\boldsymbol{r}}_1\| = a^2 + 2abn_x + b^2 = 1 \tag{8}$$

and since $\hat{\boldsymbol{r}}_1$ is orthogonal to $\hat{\boldsymbol{r}}_3$ we have:

$$\hat{\boldsymbol{r}}_1.\hat{\boldsymbol{r}}_3 = 0 = \begin{bmatrix} a+bn_x \\ bn_y \\ bn_z \end{bmatrix}^T \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} = n_x a + b = 0 \tag{9}$$

From the above equation we get:

$$\hat{\boldsymbol{r}}_1 = \begin{bmatrix} \sqrt{1-n_x^2} \\ \frac{-n_x n_y}{\sqrt{1-n_x^2}} \\ \frac{-n_x n_z}{\sqrt{1-n_x^2}} \end{bmatrix} \tag{10}$$

Finally we have that $\hat{\boldsymbol{r}}_2$ is orthogonal to both $\hat{\boldsymbol{r}}_1$ and $\hat{\boldsymbol{r}}_3$, and is obtained with the external product:

$$\hat{\boldsymbol{r}}_2 = \hat{\boldsymbol{r}}_3 \times \hat{\boldsymbol{r}}_1 = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} \times \begin{bmatrix} \sqrt{1-n_x^2} \\ \frac{-n_x n_y}{\sqrt{1-n_x^2}} \\ \frac{-n_x n_z}{\sqrt{1-n_x^2}} \end{bmatrix} \tag{11}$$

and so the transformation matrix is given by:

$${}^{\mathcal{C}}\boldsymbol{T}_{\mathcal{W}_c} = \begin{bmatrix} \sqrt{1-n_x^2} & 0 & n_x & 0 \\ \frac{-n_x n_y}{\sqrt{1-n_x^2}} & \frac{n_z}{\sqrt{1-n_x^2}} & n_y & 0 \\ \frac{-n_x n_z}{\sqrt{1-n_x^2}} & -\frac{n_y}{\sqrt{1-n_x^2}} & n_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{12}$$

The robot navigation frame of reference $\{\mathcal{W}\}$ is just $\{\mathcal{W}_c\}$ translated by $\left[\begin{array}{cccc} 0 & 0 & d & 1 \end{array}\right]^{\mathsf{T}}$, as presented in equation (4).

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. Proceeding as above, but replacing (7) with the heading reference given by the unit vector $\hat{\boldsymbol{m}}$ we have

$$\hat{\boldsymbol{r}}_1 = \hat{\boldsymbol{m}} = \left[\begin{array}{c} m_x \\ m_y \\ m_z \end{array}\right] \tag{13}$$

as before, since $\hat{\boldsymbol{r}}_2$ is orthogonal to both $\hat{\boldsymbol{r}}_1$ and $\hat{\boldsymbol{r}}_3$ , we have:

$$\hat{\boldsymbol{r}}_2 = \hat{\boldsymbol{r}}_3 \times \hat{\boldsymbol{r}}_1 = \left[\begin{array}{c} n_y m_z - n_z m_y \\ n_z m_x - n_x m_z \\ n_x m_y - n_y m_x \end{array}\right] \tag{14}$$

and so the transformation matrix using the heading reference is given by:

$${}^{\mathcal{C}}\boldsymbol{T}_{\mathcal{W}_c} = \left[\begin{array}{cccc} m_x & n_y m_z - n_z m_y & n_x & 0 \\ m_y & n_z m_x - n_x m_z & n_y & 0 \\ m_z & n_x m_y - n_y m_x & n_z & 0 \\ 0 & 0 & 0 & 1 \end{array}\right] \tag{15}$$

Translating $\{\mathcal{W}_c\}$ as before, we have

$${}^{\mathcal{W}}\boldsymbol{T}_{\mathcal{C}} = \left[\begin{array}{cccc} m_x & n_y m_z - n_z m_y & n_x & -n_x d \\ m_y & n_z m_x - n_x m_z & n_y & -n_y d \\ m_z & n_x m_y - n_y m_x & n_z & -n_z d \\ 0 & 0 & 0 & 1 \end{array}\right]^{-1} \tag{16}$$

We are therefor able to have $\{\mathcal{W}\}$ coherent with the inertial vertical and the available heading reference. The gyros included in the inertial measurement unit can be used to keep a heading without external references. Visual land marks or a magnetic compass provide an external heading reference to reset the drift accumulated over time, by the gyros.

### *E. Depth Maps in Inertial Frame of Reference*

Using the vertical reference, the depth maps can be segmented to identify horizontal and vertical features. The aim is 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. Fig. 2 summarizes the proposed depth map segmentation method.

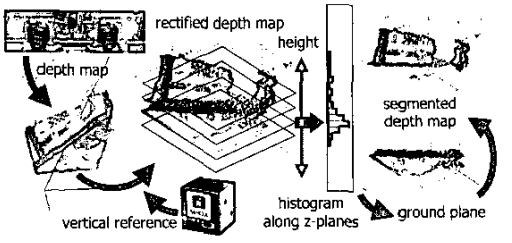


Fig. 2: Summary of implemented method.

The depth map points are mapped to the world frame of reference. In order to detect the ground plane, an histogram is performed for the different heights. The histogram's lower local peak, $z_{gnd}$, is used as the reference height for the ground plane. Fig. 4 shows some results of ground plane detection and depth map rectification. Details are given in [27].

## III Map Registration and Fusion

Map registration can be roughly partitioned into three phases: choice of the representation for the *rigid-body transformation*, the definition of the *similarity criterion and matching process* and *global optimization*. These three phases occur in the *ICP-Iterative Closest Point* algorithm [8].

The process described in section *E.* maps the points in the inertial reference frame, with planar levelled patches at the same depth $z$. These maps simplify the registration process and reduce the rotation and translation required to merge the two maps to a bi-dimensional translation vector and rotation matrix.

Assuming a general rigid-body transformation expressed as combination of a rotation and a translation, two maps represented by point set ${}^{\mathcal{A}}\boldsymbol{p}_i$ and ${}^{\mathcal{B}}\boldsymbol{p}_i$ with $i = 1, 2, ..., n$ are related by

$${}^{\mathcal{B}}\boldsymbol{p}_i = {}^{\mathcal{B}}\boldsymbol{R}_{\mathcal{A}} . {}^{\mathcal{A}}\boldsymbol{p}_i + {}^{\mathcal{B}}\boldsymbol{t} \tag{17}$$

where ${}^{\mathcal{B}}\boldsymbol{R}_{\mathcal{A}}$ is a rotation matrix and ${}^{\mathcal{B}}\boldsymbol{t}$ a translation vector.

The rigid-body registration computes the values for $\boldsymbol{R}$ and $\boldsymbol{t}$ which minimize

$$min \sum_{i=1}^{n} ||{}^{\mathcal{B}}\boldsymbol{p}_i - ({}^{\mathcal{B}}\boldsymbol{R}_{\mathcal{A}} . {}^{\mathcal{A}}\boldsymbol{p}_i + {}^{\mathcal{B}}\boldsymbol{t})||^2. \tag{18}$$

Since all depth map points are mapped to a inertial world frame, all maps have identical frames of reference with equal pose and heigh to the ground plane $z_{gnd}$. The vertical projection of points from these maps is used to compute the values for $\boldsymbol{R}$ and

$t$. Since the maps are referenced to a inertial frame, these quantities can be computed in 2D and they are equivalent to computing just one rotation angle and two planar coordinates.

The minimization (18) can be formulated as the computation of $t$, followed by that of $R$, by referring the coordinates to the centroids of each map. This leads to the minimization

$$min \sum_{i=1}^{n} ||^{\mathcal{B}}\boldsymbol{p}_i^* - (^{\mathcal{B}}\boldsymbol{R}_{\mathcal{A}}.^{\mathcal{A}}\boldsymbol{p}_i^*)||^2 \qquad (19)$$

where

$$^{\mathcal{A}}\boldsymbol{p}_i^* = ^{\mathcal{A}}\boldsymbol{p}_i - \frac{1}{n}\sum_{i=1}^{n} ^{\mathcal{A}}\boldsymbol{p}_i. \qquad (20)$$

and

$$^{\mathcal{B}}\boldsymbol{p}_i^* = ^{\mathcal{B}}\boldsymbol{p}_i - \frac{1}{n}\sum_{i=1}^{n} ^{\mathcal{B}}\boldsymbol{p}_i. \qquad (21)$$

The translation is given by the difference of centroids:

$$\boldsymbol{t} = \frac{1}{n}.\sum_{i=1}^{n} ^{\mathcal{B}}\boldsymbol{p}_i - ^{\mathcal{B}}\boldsymbol{R}_{\mathcal{A}}.\frac{1}{n}\sum_{i=1}^{n} ^{\mathcal{A}}\boldsymbol{p}_i. \qquad (22)$$

The algorithm for registration is iterative, at each iteration the set of points $X = \{^{k+1}\boldsymbol{p}_i\}$ is shifted and rotated a little closer towards the other set $Y = \{^{k}\boldsymbol{p}_i\}$. The following steps are applied until convergence is achieved:

1. Compute a sub-set $Y$ from $X$ of closest points of $U$. For each point $u$ from $U$ determine a point $x$ from $X$ which is closest in Euclidean distance to $u$.
2. Compute the rigid-body transformations using expressions from 19 to 22 to obtain the registration vector $\boldsymbol{q} = [\boldsymbol{t}, \boldsymbol{R}]$ which best aligns $Y$ with $U$.
3. Apply the translation and rotation contained in $\boldsymbol{q}$ to $U$.
4. Compute the mean-squared error.
5. Terminate if this error falls bellow a predetermined threshold, otherwise go to step 1.

## IV Results

A simple indoor scene was used to test our method. Fig. 3 shows the disparity image and reconstructed 3D points obtained with the SVS package [14] for this scene. Using the vertical reference provided by the inertial sensors, the 3D points were transformed to a world aligned frame of reference as previously described.

In order to detect the ground plane, an histogram was done for all depths, and the peak used as a reference value. The points were than parsed and segmented as ground plane points and points above ground. Fig. 4 shows the graphical front-end of the implemented system working realtime at 10 frames per second. On the left the height histogram is shown.

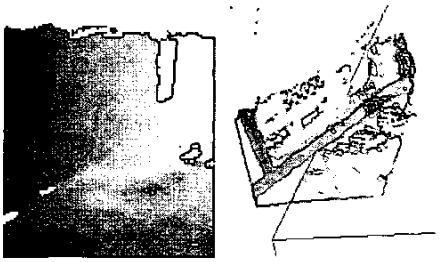

Fig. 3: Disparity image obtained with SVS [14], and reconstructed 3D points.

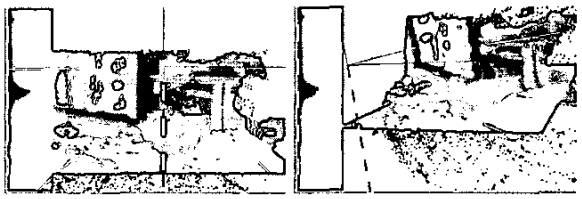

Fig. 4: Graphical front-end with height histogram and segmented depth map.

The results from the registration process are visualized by a rendering system specially developed for such purpose, shown in fig. 5. Each point has an assigned colour, obtained by the vision system. Decimation is used on the dense set of points, in order to reduce point density due to map overlap. This process decreases the time for scene image rendering, since there are fewer points to process.

Fig. 5: Two sets of point clouds obtained from different camera positions, before and after applying the full ICP.

## V 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. Segmenting out the ground plane points, the vertical projection of points are used to perform the map registration in 2D. The rotation and translation required to merge the two maps is translation vector and rotation matrix in 2D.

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, modelling the road, identifying obstacles and roadside features.

## VI ACKNOWLEDGMENTS

This work was sponsored by *FCT-Fundação para a Ciência e Tecnologia*, Portugal.

## References

[1] R. Bajcsy, G. Kamberova, and Lucien Nocera, "3D reconstruction of environments for virtual reconstruction" In Proc. of the 4th IEEE Workshop on Applications of Computer Vision, 2000.

[2] P.K. Allen and Ioannis Stamos, "Integration of range and image sensing for photorealistic 3D modeling", In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), pages 1435- 1440, 2000.

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

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

[5] 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.

[6] 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.

[7] 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.

[8] P.J. Besl, N.D.McKay, "A Method for Registration of 3-D Shapes",IEEE Transactions on Pattern Analysis and Machine Intelligence, VOL.14, NO.2, February 1992.

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

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

[11] 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.

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

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

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

[15] T. Viéville and O.D. Faugeras, "Computation of Inertial Information on a Robot", in H. Miura and S. Arimoto, editors, *5h Int. Symp. on Robotics Research*, MIT-Press, 1989, pp.57-65.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

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

[21] 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.

[22] 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.

[23] D.J. Coombs, "Real-time Gaze Holding in Binocular Robot Vision", Ph.D. Dissertation, Dept. Computer Science, Univ. Rochester, June, 1992.

[24] M.Jenkin, J. Tsotos, and G. Dudek, "The horopter and active cyclotorsion", in *Proc. IEEE International*, 1994.

[25] C.F.M. Weiman, "Log-polar vision system", Technical report, NASA,1994.

[26] 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.

[27] J. Lobo, L. Almeida and J. Dias, "Segmentation of Dense Depth Maps using Inertial Data. A real-time implementation", in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, Lausanne, Switzerland, October 2002, pp.92-97.