In. Proc. of ICPR'02, vol.4, pp.7-11, Quebec, Canada, Aug. 11-15, 2002.

Fusion of range and visual data for the extraction of scene structure information

Haris Baltzakis[†‡] Antonis Argyros[†] Panos Trahanias[†‡]

[†]Institute of Computer Science
Foundation for Research and Technology – Hellas (FORTH)
P.O.Box 1385, Heraklion, 711 10 Crete, Greece

[‡]Department of Computer Science, University of Crete
P.O.Box 1470, Heraklion, 714 09 Crete, Greece
email:{xmpalt,argyros,thahania}@ics.forth.gr

Abstract

In this paper, a method for inferring 3D structure information based on both range and visual data is proposed. Data fusion is achieved by validating assumptions formed according to 2D range scans of the environment, through the exploitation of visual information. The proposed method is readily applicable to robot navigation tasks providing significant advantages over existing methods.

1. Introduction

Laser scanners mounted on mobile robots have recently become very popular for various indoor robot navigation tasks. Their main advantage over vision sensors is that they are capable of providing accurate range measurements in large angular fields and at very fast rates. The acquired range information is compact enough to be processed in real time and encapsulates sufficient information (both in terms of quality and quantity) to enable robots to perform quite confidently a wide class of navigation tasks [3, 6, 13].

However, the quantity of information encapsulated in such 2D profiles, may prove incomplete for specific demanding or crucial robotic tasks such as obstacle detection [5, 9, 12, 14]. The main problem stems from the fact that the profiles produced from laser scans are 2D representations of the 3D space. Various objects common even in the simplest indoor environments, (e.g. chairs, tables, shelves e.t.c) are sometimes invisible to range scanners and thus, absent from the resulting 2D profiles. The potential solution of 3D laser scanners proves to be a quite expensive one. Moreover, the integration of multiple 2D profiles acquired at different heights in the environment is a relatively complex task and, even more importantly, their acquisition requires that the robot does not move for a substantial amount of time. Utilization of other sources of information is required in tasks that depend on real 3D information.

In this paper we propose a method for fusing range with visual information in order to infer 3D structure information. Simple 3D models of the environment, consisting of a flat horizontal floor surrounded by vertical planar walls are initially constructed according to 2D laser range data. Vision is then utilized in order (a) to validate the correctness of the constructed model and (b) to qualitatively and quantitatively characterize inconsistencies between laser and visual data wherever such inconsistencies are detected.

The proposed method employs a laser scanner and a camera that is registered to the laser coordinate system, through a calibration procedure. The method proceeds by exploiting sensory information acquired in two consecutive time instances, as the robot moves in space. At time t_1, the robot acquires a laser range scan R_1 and an image I_1. Based on R_1 the robot builds a 3D model of the environment. The same process is applied at time t_2, resulting to R_2 and I_2. By registering the R_1 and R_2 the robot is able to compute its egomotion. Based on the 3D model derived at time t_1 and the recovered motion, image I_1 is backprojected at the reference frame of image I_2, resulting in image I_2'. Images I_2' and I_2 should be identical in areas where the 3D model is valid and should differ in areas where the model is not valid. Comparison is performed by locally correlating image intensity values. For providing additional depth information in regions with inconsistencies between laser and visual data, image intensity matches along the epipolar line are converted to real word coordinates and accumulated in a 2D occupancy map.

The proposed method has been tested on both synthetic and real data. The results presented in this paper demonstrate its effectiveness.

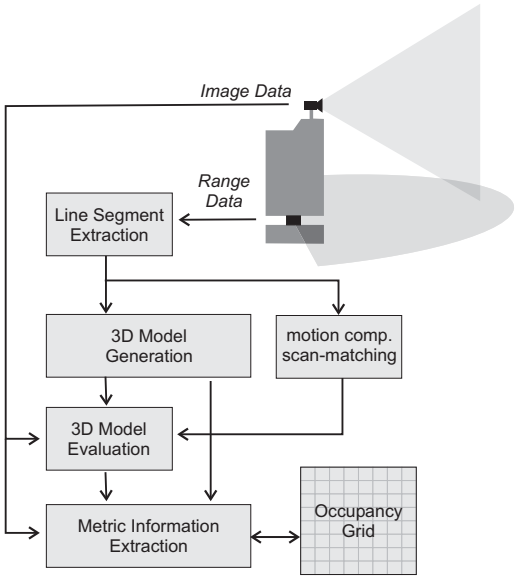


Figure 1. Block diagram of the proposed method

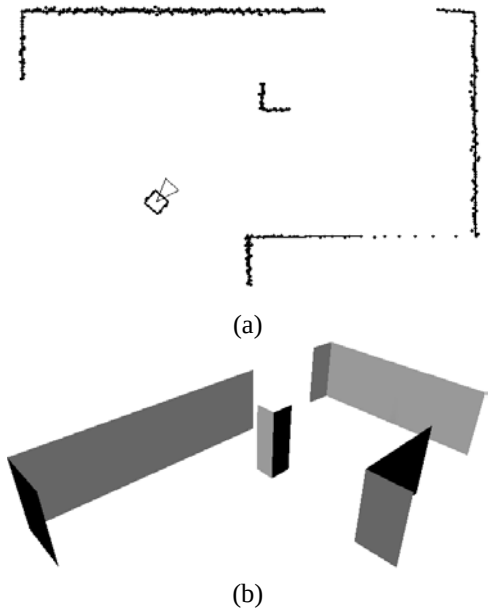


Figure 2. 3D Model definition process (a) range data and line segments defined and (b) resulting 3D model.

2. Method Description

In this section, the proposed method for fusing laser and visual data is described in detail. Figure 1 provides a block diagram of the method.

2.1. Line Segment Extraction and 3D-model Generation

In order to build a 3D model of the environment, range measurements have to be grouped into line segments. For line segment extraction, a three-stage algorithm has been implemented. Range measurements are initially grouped to clusters of connected points according to their Sphere-of-Influence graph [10]. Clusters are then further grouped to line segments by utilizing the Iterative-End-Point-Fit (IEPF) algorithm [4, 2]. Finally, after range points have been segmented into groups of collinear points, line segment parameters are re-estimated by a line fitting procedure.

For generating the local 3D model of the environment, an infinite horizontal plane (floor) is generated right below the robot, at a known distance from the robot's coordinate system (the position of the range finding device). Then, line segments defined in the previous step are extended to form rectangular vertical surfaces of infinite height. More specifically, for each line segment, the plane that is perpendicular to the floor and contains the line segment is inserted to the 3D model. The coordinate system of the generated 3D model is assumed to coincide with the coordinate system of the robot. Figure 2(a) shows the line segments as extracted by the algorithm for a simple artificial environment. Laser measurements are also depicted in the image. A rendered view of the corresponding 3D model is depicted in Fig.2(b).

2.2. Coordinate System Registration

Camera positions (with respect to the 3D model) need to be known at the time when the two images are captured. Provided that the relative position of the camera with respect to the laser range scanner is fixed and can be obtained through calibration procedures, camera positions can be recovered if the motion of the robot is known or can be computed.

For calculating the motion of the robot, an iterative, scan-matching algorithm based on the Hausdorff distance metric [1, 7] has been employed. The algorithm is applied directly to the line segments extracted as described in the previous section.

Given two sets of line segments L_1, L_2 corresponding to range scans acquired at two different time instants t_1 and t_2, the goal of the algorithm is to find the transformation $T = (dx, dy, df)$ (robot's motion) that, when applied to L_1, produces a set L_2' as similar to L_2 as possible. As a measure

of similarity, the directional Hausdorff distance, given as

$$H(L_2, L_2') = \max_{l_2 \in L_2} \min_{l_2' \in L_2'} ||l_2 - l_2'|| \quad (1)$$

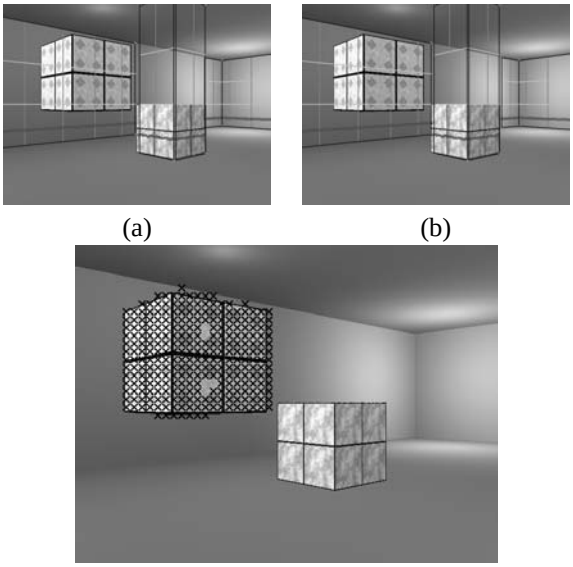
has been utilized.

2.3. Model Evaluation

Let M be the 3D model built as described in section 2.1, according to range data acquired at the time instant t_1, and let I_1 be an image acquired by a camera c_1 at the same time instant.

For each image point $p_1 = (x_1, y_1)$ of I_1, a 3D coordinate $P = (X, Y, Z)$ can be found by ray-tracing it to the model M. Let I_2 be a second image, acquired by the same camera at a different time instant t_2. Since the coordinate system of c_2 with respect to the coordinate system of M is also known, for each point $P = (X, Y, Z)$ in M, the projected point $p_2 = (x_2, y_2)$ can also be calculated. By ray-tracing points of I_1 to find 3D world coordinates and projecting them to I_2, we are able to calculate analytically correspondences between I_1 and I_2. If the assumptions made in order to form the model M are correct, corresponding image points would actually be projections of the same world object points and, thus, they will share the same attributes (color, intensity values, intensity gradients etc). The normalized cross correlation metric [4] is employed to evaluate the correctness of the calculated point correspondences. Low values of the calculated cross correlation correspond to regions of the environment that do not conform with the 3D model constructed from laser data.

Figures 3(a) and 3(b) show two consecutive frames of a synthetic scene captured at the positions corresponding to the range data depicted in Fig.2(a). For convenience, wireframes of the 3D-model extracted according to the procedure described in section 2.1, as well as range finder points, are projected on the images. The scene contains two cubes; one lying at approximately 1m above the floor (on the left part of the image) while the other (on the right) is placed directly on the floor. Figure 3(c) demonstrates the results of the evaluation process. Regions with inconsistencies are marked with "x"s. As it can be easily observed, the algorithm succeeds in correctly detecting the cube on the left of the image that is "invisible" to the range finder since it is floating above its scanning plane. On the other hand, the cube on the right of the image, does not yield any unmatched areas because it is visible to the range scanner.



(a) (b)

(c)

Figure 3. Example of the 3D model evaluation process. (a),(b) two frames of an artificial scene containing two cubes. The 3D model and the range finder points are also projected on the images, (c) results of the evaluation process, projected on the second image.

3. Extraction of Metric Information

In the previous section we utilized vision in order to evaluate the correctness of 2D range information provided by range scanners. Having identified regions of inaccurate range information, vision is further employed to qualitatively and quantitatively characterize these inconsistencies.

Let's assume the camera configuration depicted in Fig. 4. The 3D point P lying on the model M is projected to the point p_1 on the left image (I_1) and to the point p_2 on the right image (I_2). The epipolar plane π created by point P and the camera centers C_1 and C_2 intersects the image planes in lines l_1 and l_2, the epipolar lines [8].

Suppose that the image coordinates of a point p_1 are known, and that neither the location of the corresponding 3D point P nor the corresponding point p_2 in the second image is known. It has already been shown that by ray-tracing point p_1 to the 3D model we can compute the 3D coordinates of point P and by projecting the later to the second camera plane we can define the corresponding point p_2 in the second image. Suppose that by locally correlating pixel intensity values at the positions of p_1 and p_2, we discover a dissimilarity and we conclude that the 3D model M is inaccurate and hence the 3D coordinates of point P as implied by the model, are not correct. This raises the question whether point P actually lays behind the model M (further to the camera than assumed) or in front of it; the latter making it a potential obstacle invisible by the range finder.

Whatever the depth of point P may be, its projection on

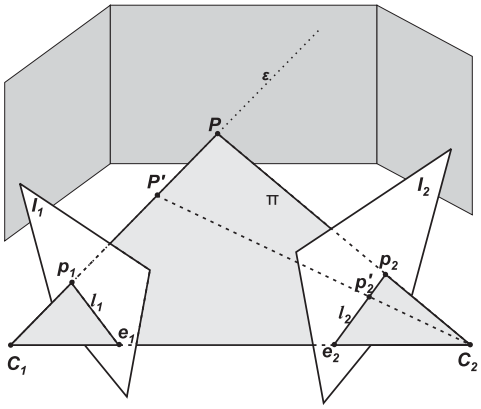


Figure 4. Epipolar geometry for the two cameras.

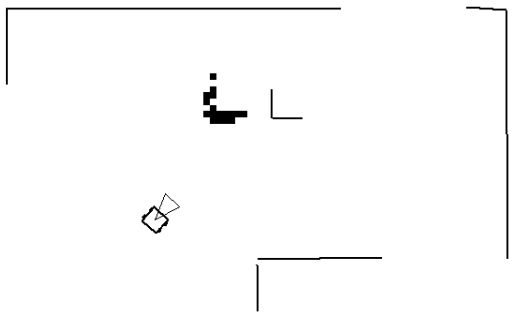


Figure 5. Extraction of metric information by utilization of visual data.

the second camera will comply to the epipolar constraint; that is, it will lay on the epipolar line l_2. The important observation is that the shortest the depth of point P actually is, the closer its projection p_2 to the epipolar point e_2 will be. That is, if point p_1 actually corresponds to a 3D point P' closer to the first camera than point P, its projection $p_2{}'$ on the second camera will lay on the epipolar line l_2, between point p_2 and the epipole e_2. If point P', lays further to the first camera than P, its projection $p_2{}'$ will also lay on l_2 but this time outwards the direction of e_2. If the exact location of point $p_2{}'$ corresponding to p_1 were known, computation of the intersection of the line passing through points C_1 and P with the line passing through points C_2 and $p_2{}'$ would yield the exact 3D location of point P'. However, since exact computation of $p_2{}'$ is not always possible, only guesses about the position of P' can be made. Relying on the assumption that for robots that move on a planar surface, the projection of the obstacles on the 2D surface of motion suffices for navigation, we alleviate the problem of spurious range evidence by accumulating range estimates in a 2D occupancy grid [11] in order to accumulate evidence about the location of P'. The exact algorithm is as follows:

- Initialize accumulation occupancy grid
- For each unmatched point pair p_1-p_2 repeat:
  For each point $p_2{}'$ lying close to p_2 along the epipolar line l_2, repeat:
  - compute the correlation of the intensity values near in the vicinities of p_1 and $p_2{}'$
  - If the correlation is above a threshold, compute the location of point P' as the intersection of the lines C_1-P and C_2-$p_2{}'$ and add the correlation result on the cell of the occupancy grid corresponding to the position of P'

Figure 5 demonstrates the results of the procedure described above for the synthetic data set used in the previous sections. As it can be easily observed, the location of the left cube shown in Figs. 3, although not visible by the range finder, is correctly identified on the resulting occupancy grid.

4. Results

The proposed method has been implemented and assessed on a robotic platform of our laboratory, namely an iRobot-B21r, equipped with a SICK-PLS laser range finder and a digital camera operating at a standard resolution of 640 x 480 pixels. The range finder is capable of scanning 180 degrees of the environment, with an angle resolution of one measurement per degree and a range measuring accuracy of 5cm. An internal calibration procedure has been applied prior to testing our methodology, so that the relative positions of the both sensors were known as well as their intrinsic parameters. Extensive tests have been performed with real and simulated data. In all cases the proposed framework was verified to operate accurately, provided that the outcome of the calibration procedure was also accurate.

Figure 6 demonstrates the operation of the proposed method in a corridor structure outside our lab. A pair of images acquired sequentially by the robot's camera are shown in Figs. 6(a) and 6(b). Projections of range finder data as well as of the resulting 3D model are also overlayed on the images. The results of the evaluation process, projected on the second image are shown in Fig. 6(c). Regions with inconsistencies are marked with an "x". As it can be verified, various structures laying on the walls of the corridor, invisible by the range finder were correctly identified by the evaluation process. Results of the metric information extraction

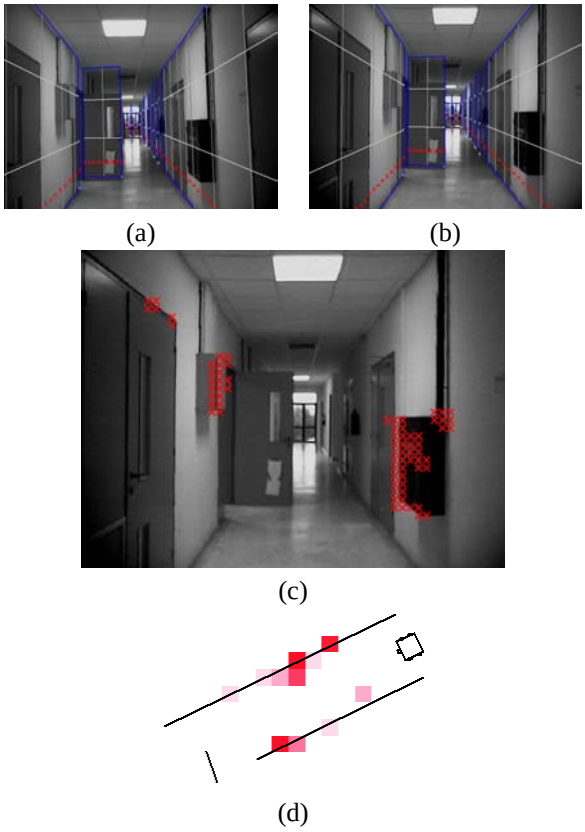


Figure 6. Demonstration of the proposed framework in a corridor environment

algorithm, in the form of an occupancy grid map, applied to these areas of range data inconsistency, are depicted in Fig. 6(d). For convenience, line segments used for constructing the 3D model are also overlayed.

5. Conclusions

In this paper a new method for fusion of range and visual data for the extraction of 3D structure has been proposed. Visual information is used to detect regions where vision and laser data are mutually inconsistent. Moreover, vision is utilized to provide additional metric information in such regions. Since pixel displacements are computed analytically by rendering image points to the model, their direct computation is not necessary. The proposed method requires two views of the environment. These can be acquired either by one moving camera at two different points in time, or by a stereoscopic vision system that simultaneously acquires two views of a scene.

Besides its obvious applicability to obstacle detection, the general idea presented in this paper can be utilized by range based mapping and navigation algorithms in order to make them more accurate and robust. We believe that fusion of data provided by range and vision sensors constitutes a proper framework for mobile robotic platforms to perform demanding navigation tasks. It is in our intention to further-investigate the applicability of the presented methodology in this area.

Acknowledgments

This work has been partially supported by EU IST projects WebFair (IST-2000-29456) and ActIPret (IST-2001-32184).

References

[1] V. Ayala-Ramirez and C. Parra and M. Devy. Active Tracking Based on Hausdorff Matching. In *Proc. IEEE ICPR*, Vol IV: pages 706–709, 2000.

[2] G. Borges and M. Aldon. A split-and-merge segmentation algorithm for line extraction in 2-d range images. In *Proc. IEEE ICPR*, Vol I: pages 441–444, 2000.

[3] J. Castellanos, J. Tardós, and J. Neira. Constraint-based mobile robot localization. In *Proc. Intl. Workshop Adv. Robotics and Intell. Machines*, Apr. 1996.

[4] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. Wiley-Interscience, New York, 1973.

[5] P. Fornland. Direct obstacle detection and motion from spatio-temporal derivatives. In *Proc. Intl. Conf. Comp. Analysis of Images and Patterns*, pages 874–879, Sep. 1995.

[6] J.-S. Gutmann and K. Konolige. Incremental mapping of large cyclic environments. In *Proc. IEEE Intl. Symp. on Computational Intelligence in Robotics and Automation*, Monterey, CA, 2000.

[7] D. P. Huttenlocher and G. A. Klanderman and W. J. Rucklidge., Comparing Images Using the Hausdorff Distance. *IEEE Trans. on PAMI.*, 15(9):850-863, 1993.

[8] K. Kanatani. *Geometric Computation for Machine Vision*. Clarendon Press, Oxford, 1993.

[9] L. Matthies and P. Grandjean. Stochastic performance modeling and evaluation of obstacle detectability with imaging range sensors. *Robotics and Automation*, 10:783–792, 1994.

[10] T. Michael and T. Quint. Sphere of influence graphs in general metric spaces. *Mathematical and Computer Modelling*, 29:45–53, 1994.

[11] H. P. Moravec and A. Elfes. High resolution maps from wide angle sonar. In *Proc. IEEE ICRA*, pages 116–121, St. Louis, Missouri, 1985.

[12] C. Stiller, J. Hipp, C. Rossig, and A. Ewald. Multisensor obstacle detection and tracking. *Image and Vision Comput.*, 18(5):389–396, Apr. 2000.

[13] S. Thrun, A. Buecken, W. Burgard, D. Fox, T. Froehlinghaus, D. Hennig, T. Hofmann, M. Krell, and T. Schmidt. Map learning and high-speed navigation in RHINO. Tech. Rep. IAI-TR-96-3, Dept. of Computer Science, Univ. of Bonn, Jul. 15, 1996.

[14] Z. Zhang, R. Weiss, and A. Hanson. Obstacle detection based on qualitative and quantitative 3d reconstruction. *IEEE Trans. on PAMI.*, 19(1):15–26, Jan. 1997.