3D RECONSTRUCTION OF INDOOR ENVIRONMENTS

Vítor Sequeira, João G. M. Gonçalves, M. Isabel Ribeiro(*)

European Commission, Joint Research Centre,Ispra (VA), Italy
(*) Instituto Superior Técnico/Institute for Systems and Robotics, Lisbon, Portugal
E-mail: [vitor.sequeira, joao.goncalves]@jrc.it; mir@isr.ist.utl.pt

ABSTRACT

This paper presents a new 3D scene analysis system that automatically reconstructs the 3D model of real-world scenes from multiple range images acquired by a laser range finder on board of a mobile robot. The reconstruction is achieved through an integrated procedure including range data acquisition, geometrical feature extraction, planning the next view, registration and integration of multiple views. The system relies only on the acquired data.

Experimental results on real range images are presented in the paper. Direct applications of this technique include 3D reconstruction and/or update of architectural or industrial plans into a CAD model, design verification of buildings, navigation of autonomous robots, and input to virtual reality systems.

1. INTRODUCTION

The goal of 3D scene reconstruction is to provide a 3D model as complete and as accurate as possible, from a real world scene, for which no *a priori* information is available. Some of the current systems for 3D reconstruction acquire single depth images and/or have simple geometric fusion techniques producing inefficient representations of the underlying 3D surfaces. Except for the trivial model-building tasks, it is necessary to locate sensors at several positions in the environment, since all surfaces may not be visible from a single point, nor will data be acquired at sufficient resolution. Mobility is therefore paramount to 3D scene reconstruction. An obvious way of transporting sensors between positions is to have them on a mobile platform. Lebegue and Aggarwal [1] acquire video data from a camera aboard a manually operated mobile platform, aiming at building architectural models to be loaded into a commercial CAD system. The alternative is for the platform to move autonomously in the environment. In most of the work to date, the aim of model building has been for the navigation of the platform itself [2]. As the goal is the 3D description of the free space around the robot for navigation purposes, no occlusion detection is required.

Data acquisition for environment reconstruction, as opposed to robot navigation, has specific requirements: all scene geometry need to be adequately represented, and sufficient visual detail must be acquired for surface rendering at the required resolution. This strongly points out for an active data gathering strategy and pushes towards a requirement for a surface representation where multiple views can be easily fused and entire surfaces represented.

This paper presents a new 3D scene analysis system that automatically reconstructs the 3D model of real-world scenes from multiple range images acquired by a laser range finder on board of a mobile robot. The reconstruction is achieved through an integrated procedure including a) range data acquisition, b) geometrical feature extraction, c) planning the next capture point, d) registration and e) integration of multiple views. The system relies only on the acquired data. A first range image is acquired from an arbitrary or pre-defined capture point. Next capture points are estimated from the already reconstructed model. Fig. 1 is an overview of the iterative procedure for 3D reconstruction.

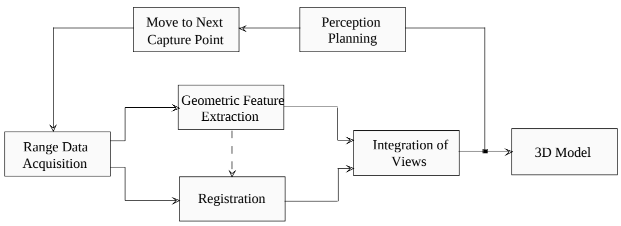


Fig. 1. Iterative procedure for 3D scene reconstruction.

2. 3D SURFACE RECONSTRUCTION FROM MULTIPLE RANGE VIEWS

The 3D scene modelling system combines different algorithms to overcome the specific problems raised by this type of range images. In particular, a new hybrid algorithm for range image segmentation was developed. This algorithm combines edge and region detection approaches to build a high-level description of a scene given a noisy range image [3]. The algorithm puts together the good localisation of range edges with the stability of surfaces.

The combination of range images requires the registration of partially overlapping images, aiming at finding the geometrical transformation between the co-ordinate frames of data acquired at different locations. In the present case, the problem is to register, as precisely as possible, two partially overlapping surfaces that may lack significant features. Based on that, a hierarchical registration scheme working directly at the pixel level was implemented for higher accuracy [4]. The integration scheme combining the different views works, instead, at a higher representation level, i.e., edges and surfaces. For constructing the model resulting from the combination of the multiple views, an average of the samples that are in the overlapping surfaces of the views is performed. The surface patches that overlap are detected and then the surface parameters that fit to the fused data are adjusted. The merged surface is then expanded to the edge points, that were previously merged, and the boundary is reconstructed. The iterative integration of surface descriptions obtained for each view by the hybrid algorithm is summarised in Fig. 2.

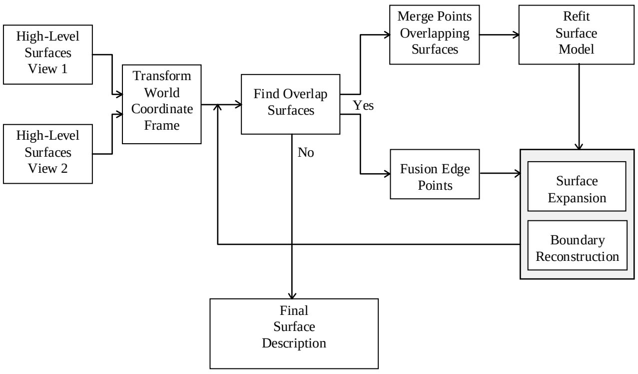


Fig. 2: Integration of Multiple High-Level Surface Descriptions

3. PERCEPTION PLANNING

The problem of determining the next capture point depends on how much *a priori* information about the scene is available. Other works tackled the problem of sensor positioning when i) the fully environment is known [5], ii) for 3D object recognition [6], or iii) for the reconstruction of unknown small objects [7,8].

In the present work the problem is formulated as follows: given the already reconstructed model up to a time instant, which is the best capture point to perform the next range acquisition. The selection of the next capture point should consider:

a) The pieces of environment which have not yet been modelled.
b) The resolution of occlusions within the already acquired model.

There is hence the need to:

i) detect the occurrence of occlusions;
ii) evaluate the next most appropriate capture point;
iii) determine the next scanning parameters.

A strategy for planning the next view based on occlusions, was developed using the objective function:

$$F(x,y,z) = \sum_{i=1}^{M} F_i(x,y,z), \qquad (x,y,z) \in \bigcup_{i=1}^{n_v} V_i \tag{1}$$

with

$$F_i(x,y,z) = A_i \cdot \eta_i(x,y,z) \cdot \varepsilon_i(x,y,z)\,, \tag{2}$$

where M represents the total number of occlusions at a given time, A_i is the area of the i^{th} occlusion resolved with the current position (benefit), and $\eta_i\,(x,y,z)$ is the cost function associated with the angle between the viewing direction and the occlusion plane. Finally, $\varepsilon_i\,(x,y,z)$ is the cost function (Normal density) associated with $d_i\,(x,y,z)$, the distance between the acquisition point and the origin of the occlusion co-ordinate system.

The motivation for the objective function comes from the fact that there are some locations where data can be acquired with higher precision (e.g., the best acquisition position is perpendicular to the surface being scanned). Furthermore, other locations minimise the number of capture points, and maximise the occlusion areas to be resolved.

The selection of the next capture point is thus treated as an optimisation problem [9]. The solution to this problem takes also into consideration the environment that is incrementally being built and the associated constraints: topological (imposed by the objects being scanned) and operational (imposed by the already reconstructed environment and by the acquisition system).

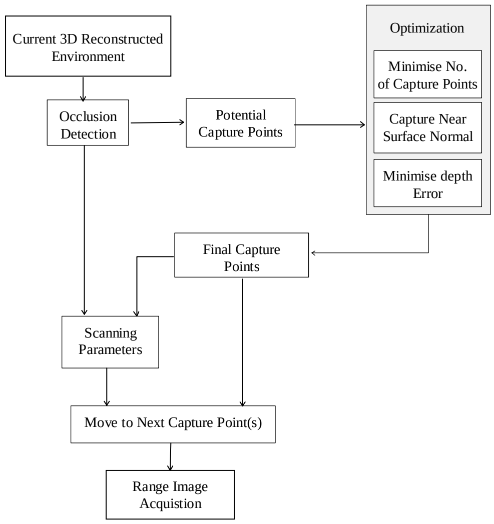


Fig. 3: Perception Planning

Fig. 3 represents the perception planning scheme. The method starts by detecting the occlusions on the current 3D reconstructed environment, followed by the evaluation of the set of potential capture points from which all the occlusions can be resolved. This set is fed into an optimisation procedure aiming at:

- minimising the number of capture point,
- selecting those points from which the occlusions areas are captured, as much as possible, along the normal to the occluded plane,
- selecting those points from which the distance to the occlusions leads to smaller errors on range acquisition, which is a function of the used laser.

The output of this procedure is the set of final point(s) that, when combined with the occlusions that are visible from that point(s), determine the scanning parameters (e.g., image field of view, view direction and area to be covered in one scan). The acquisition system is then moved to the next capture point and a set of new views is acquired with the evaluated scanning parameters. Whenever there is the need of more than one capture point for resolving all the occlusions on the current model, the acquisition system is moved to each capture point, a set of new images is acquired at each point, and afterwards, a single model is constructed.

4. EXPERIMENTAL RESULTS

Range images were acquired using a time-of-flight laser range finder mounted on a computer controlled pan and tilt unit aboard of a mobile robot (see Fig. 4). Distances are measured on a grid with equally spaced pan and tilt angles. Fig. 5 shows three different representations of a typical range image with 140 by 140 samples, covering a solid angle of 60° by 60°.

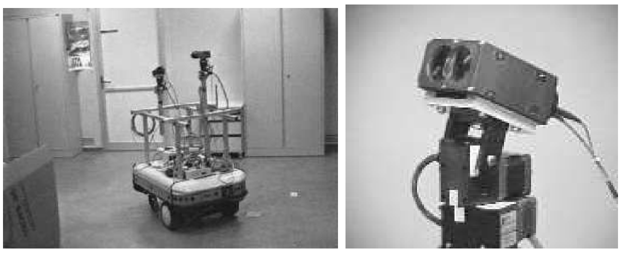


Fig. 4: LRF on board of the mobile robot.

The first image is acquired from an arbitrary or pre-defined capture point. The next capture point(s) are estimated based on the information extracted in the previous images.

A complete 3D surface model resulting from the integration of five range images of a laboratory (scanned area of 10x4 m) to which Fig. 4 and 5 belongs, is presented in Fig. 6. This example is a sub-set of the full room description which has been reconstructed from 12 images. The final result is a 3D model of the laboratory described in terms of planar and bi-quadratic surfaces with explicit boundary information. Range images with different spatial and range resolutions were acquired from four different positions of the mobile robot. The reconstruction was achieved without any human intervention to correct the final model. The algorithm's thresholds have been kept constant for the processing of the different range images, this proving the robustness of the algorithms due to the presence of different size and shape objects. It should be stressed that reconstructed models must extract meaningful information from noisy sensory data while making no assumptions about observed shapes.

Current work progresses towards the integration of other sensory data such as reflectance. The aim is to make the overall process more robust and reliable, and to have more realistic representations by rendering the geometric models with natural texture. Results using the texture information from the infrared reflectance image are shown in Fig. 7. This is a new concept called virtualized reality, that turns a real world scene into a virtual one. This differs very significantly from traditional virtual reality, in the sense that rather than generating a virtual world using CAD graphic models, it starts with the real world and virtualize it.

5. CONCLUSIONS

The paper describes a full integrated procedure for efficient 3D scene reconstruction from range acquisition to the final model, including occlusion detection and resolution. The algorithms were fully tested in real world environments with different shape complexity and surface types (e.g., planar and convex or concave curved surfaces).

Previous work on sensor placement was done aiming at the reconstruction of unknown small objects. This work presents a new technique for sensor placement aiming at scene modelling of large and complex environments taking into account, at each step, the already reconstructed environment and associated constraints.

The system operates autonomously with no *a priori* knowledge about the scene to be reconstructed and with no human intervention, and relies only on the acquired data.

6. ACKNOWLEDGEMENTS

V. Sequeira, Ph.D. student of Instituto Superior Técnico, acknowledges the PRAXIS XXI programme from JNICT, Portugal, for his Ph.D. grant. This research was partially funded by the ACTS programme of the European Union (RESOLV project).

7. REFERENCES

[1] X. Lebegue, J. K. Aggarwal - Generation of Architectural CAD Models Using a Mobile Robot, Proc. IEEE Int.. Conf. on Robotics and Automation., pp. 711-717, May 1994.

[2] M. Hebert, T. Kanade, I. Kweon - 3-D Vision Techniques for Autonomous Vehicles, Analysis and Interpretation of Range Images (R.C Jain, A.K. Jain Editors), pp 273-337, Springer-Verlag 1990.

[3] V. Sequeira, J.G.M. Gonçalves, M.I. Ribeiro - High-Level Surface Descriptions from Composite Range Images, in Proc. IEEE International Symposium on Computer Vision, Florida (USA), November 1995, pp. 163-168.

[4] V. Sequeira, J.G.M. Gonçalves, M.I. Ribeiro - 3D Environment Modelling Using Laser Range Sensing, in Journal of Robotics and Auton. Systems - Elsevier, 16(1), Nov. 1995, pp. 114-127.

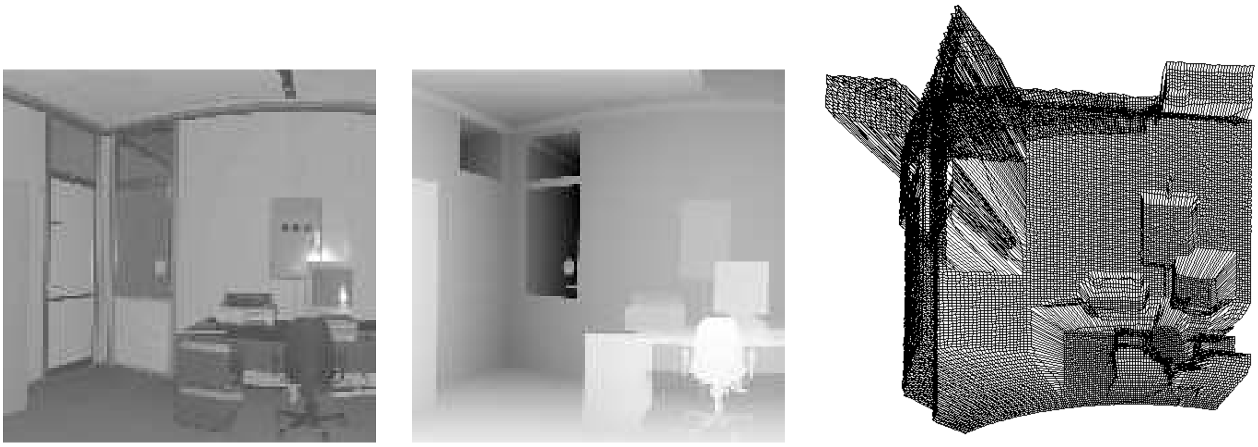
[5] K. Tarabanis, R.Y. Tsai, Viewpoint Planning: the visibility constraint, Proceedings of the Image Understanding Workshop, May 1989, pp. 893-903.

[6] S.A. Hutchinson, A.C. Kak - Planning Sensing Strategies in a Robot Work Cell with Multi-Sensor Capabilities, IEEE Trans. on Robotics and Automation, 5(6), pp. 765-783, 1989.

[7] J. Maver, R. Bajcsy - Occlusions as a Guide for Planning the Next View, IEEE Trans. Pattern Anal. Mach. Intell. 15(5), May 1993, pp. 417-433.

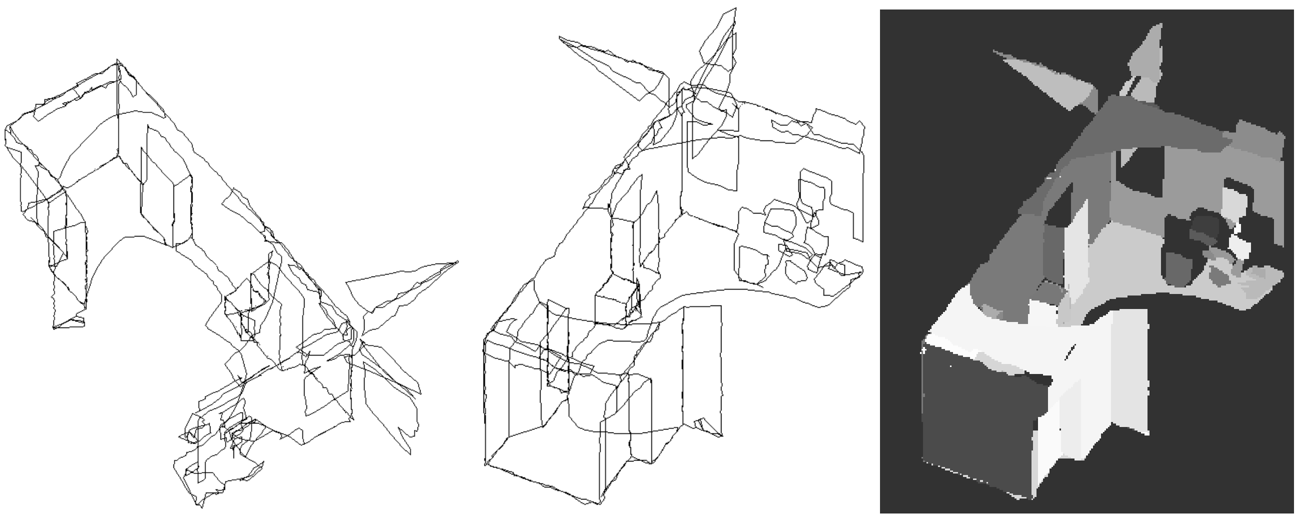
[8] P. Whaite, F.P. Ferrie - From Uncertainty to Visual Exploration, IEEE Trans. on Pattern Anallysis and Machine Intelligence PAMI-13 (10), 1991, pp. 1038-1049.

[9] V. Sequeira, J.G.M. Gonçalves, M.I. Ribeiro - Active View Selection for Efficient 3D Scene Reconstruction - Proc. 13th ICPR'96 - Computer Vision, Vienna (Austria), August 1996.



(a) b) (c)

Fig. 5. Range image from an office scene: a) infrared reflectance b) grey level; c) 3D perspective view



(a) different views of boundary after merging (b) final surface description

Fig. 6: High-level surface description from composite range images

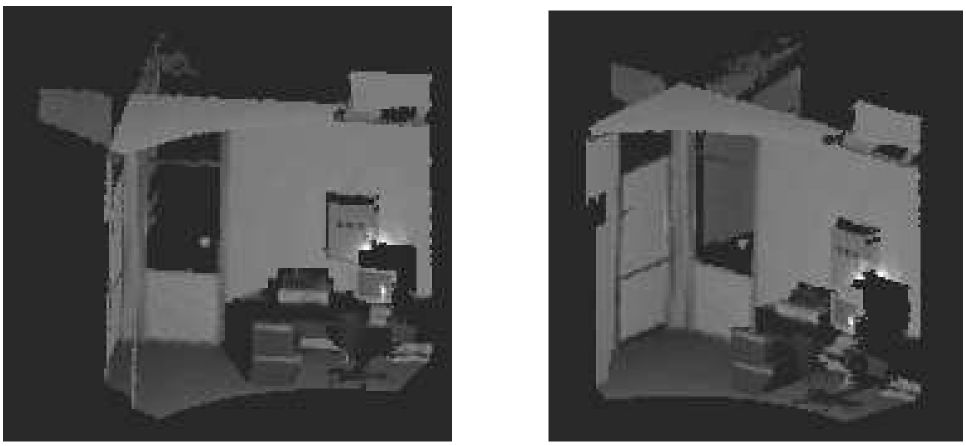


Fig. 7. 3D model with texture information from reflectance image.