MVA'94 IAPR Workshop on Machine Vision Applications Dec. 13-15, 1994, Kawasaki

3D MODELLING OF IN-DOOR SCENES USING LASER RANGE SENSING

Vítor Sequeira†, João G. M. Gonçalves

European Commission
Joint Research Centre, 21020 Ispra (VA), Italy
e-mail: [vitor.sequeira,joao.goncalves]@cen.jrc.it

M. Isabel Ribeiro

Dept. Eng. Electrotécnica e Computadores
Instituto Superior Técnico, Lisboa, Portugal
e-mail:mir@kappa.ist.utl.pt

ABSTRACT

This paper describes a 3D scene analysis system that is capable of modelling real-world scenes, based on data acquired by a Laser Range Finder on board of a mobile robot. Laser range images directly provide access to three dimensional information, as opposed to intensity images. This characteristic makes range data suited to build 3D models of objects, and in general real world scenes. Results from the complete modelling procedure are presented, starting with the extraction of surface characteristics from range images, approximation of the surface geometry by a 3D triangular surface mesh and the merging of the surface descriptions from different viewpoints into a consistent 3D scene model. Direct applications of this technique include retrieval of architectural and/or industrial plans into a CAD model, design verification of buildings and plants, and update of a CAD model with the real world ("as-built").

INTRODUCTION

Most computer vision research has concentrated on replicating human vision using digitised grey-scale intensity images as sensor data. This approach proved to be most difficult to understand and describe real world images in a general purpose way. An important problem is the lack of explicit depth information in intensity images, capable of directly relating the image to the real world. In recent years digitised range data became available from both active and passive sensors. Range data is produced in the form of a rectangular array of numbers (range image), where the numbers quantify the distances from the sensor plane to the surfaces within the field of view.

Active range sensing (e.g. Laser Range Finder) has several advantages over passive range sensing: a) it provides explicit range information (without the computation overhead of conventional passive techniques such as stereo vision), b) independence from illumination conditions, and c) largely independent from the reflectivity of the objects being sensed. Intensity image problems with shadows and surface markings do not occur.

This paper describes the automatic creation of a 3D volumetric model (i.e. a description of the topology and the contents) of an unknown environment, e.g. an office scene, based on data acquired by a Laser Range Finder (LRF) on board of a mobile robot. Such a description is most useful for visualisation, scene understanding, mobile robot navigation and verification purposes [1,2]. Constructing the model requires sensing and interpretation, ranging from low-level data collection to high level scene modelling.

Most works described in literature deal with small part modelling usually for industrial applications. The range acquisition system is normally based on structured light techniques and the objective is the identification and/or recognition of a given object so that it can be counted, sorted or grasped by a manipulator arm. The major novelty of the 3D scene analysis system presented is that is capable to model real-world complex scenes (distance range from 0.5 to 20 meters), based on data acquired by a Laser Range Finder.

Scene modelling is at the end of a long chain of processing steps. This chain starts with range data acquisition and conversion to a format that is independent from the scanning system. This is followed by the extraction of surface characteristics from range images, and the approximation of the surface geometry by a 3D triangular surface mesh. If a complete volumetric model is required, there is the need for acquiring range data from multiple viewpoints, to solve all the ambiguities due to occlusions. This leads to the problem of registration and integration of range images into a single model. A complete model is more than the sum of individual models, and integrate features from individual views.

The paper is structured along the computational steps described above, focusing mainly on the problem of merging surface descriptions from different viewpoints into a consistent 3D scene model.

RANGE IMAGE ACQUISITION/PROCESSING

A large variety of measurement techniques can be used to create a range image, including structured light, time-of-flight lasers, radar, sonar, and several vision based methods such as depth from stereo, shading, texture, motion and focus [3]. In the present case, range images were captured using a time-of-flight Laser Range Finder (LRF). To have a range image of a real world scene, it is necessary to scan the LRF in both horizontal and vertical directions. The mechanical scan is achieved by rotating the LRF by means of a computer controlled pan-and-tilt unit. Fig. 1 sketches this scanning principle.

Fig. 2 a) represents a grey level representation of a range image from an office scene. Lighter areas correspond to shorter distances. It can be seen that some straight lines, e.g., the corner between the wall and the

†PhD program supported by "Programa Ciência"-JNICT - Portugal

ceiling, look curved. This effect is due to the fact that the scanning mechanism has a cylindrical geometry and the image is being displayed on a plane. A 3D view of the same range image with geometrically corrected data is shown in Fig. 2 b). Some details of the scene illustrate the potentialities of the range image acquisition system for representing small objects: i) the drawers' handles; ii) a small thermostat on the wall; iii) the door frame and handle.

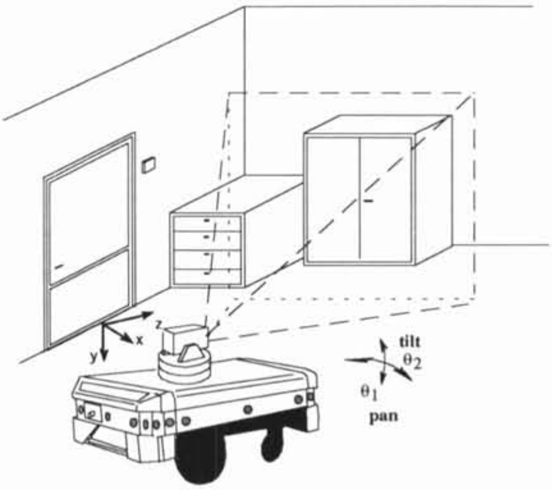


Fig. 1 Acquisition of a range image by scanning a LRF

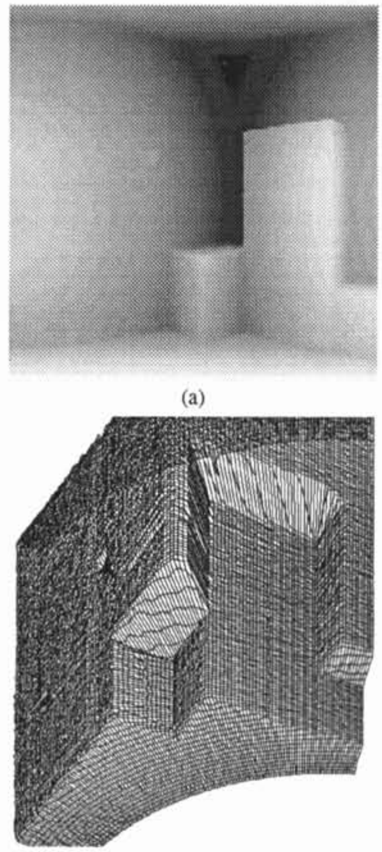
Some requirements for the extraction of geometric features for geometric modelling is that the localisation of edges, especially those corresponding to surface discontinuities (step edges), must be as precise as possible, and that the classification of edges must be correct. Based on these criteria, it was decided to use edge-based methods due to the poor ability of region-based methods in edge localisation and classification [2]. Most objects, both in indoor and outdoor urban environments, can be represented by planar patches bounded by linear edges. The edge-detection is based on geometric model fitting [2,4]. The basic idea is to fit pre-defined edge models to data, and then make a decision to either accept or reject the presence of such models.

3-D SURFACE RECONSTRUCTION

The scene reconstruction process starts from range segments, and is based on a 2D constrained Delaunay triangulation [5]. This is done in the image plane and results in a triangular faced piecewise linear description of scene surfaces. Since the input data is a range image, a 2D triangulation method can be used to construct a surface triangulation.

The construction of a surface triangulation constrained to the pre-extracted surface features is the same as having a 2D triangulation constrained to the projection of these features in the image domain. The computed 2D mesh is then back-projected into the 3D space using the corresponding 3D segments endpoints evaluated during the edge-detection phase. The result of the whole procedure is a triangular-faced piecewise linear surface, in which segments are preserved. An interesting feature of this approach is its fairly low computational cost, considering that most of the processing is done in 2D. The surface reconstruction procedure consists of three phases:

i) Reconstruction of range segments;
ii) Constrained Delaunay triangulation in the image plane;
iii) Back projection of the 2D tessellation onto the 3D space.



(a)

(b)

Fig. 2: Representation of a range image from an office scene: a) Grey level display b) 3D perspective view

COMBINING MULTIPLE RANGE IMAGES

One range image is only a portion of the information present in a 3-D working domain. To resolve the ambiguities that are caused by occlusions (see left side of the cupboard in Fig. 2), images from different viewpoints are required.

There are two main issues in creating a single model from multiple range images: registration and integration. Registration refers to computing a rigid transformation that brings the points of one range image into alignment with the overlapping parts of a second range image. Integration is the process of creating a single representation from the sample points from several range images. In the present case, the problem is registering two partially overlapping surfaces that may lack significant surface features. The registration is required to be as precise as possible. From these two considerations, it can be seen that methods based on distance minimisation [6,7] are more appropriate to the present case.

The iterative registration algorithm that was implemented assumes that an initial estimate of the 3-D rigid transformation between two partially overlapping range images is available. It can be obtained by a high level registration process or provided by the data acquisition process. The registration algorithm is divided into two different steps:

i) Finding the corresponding point
ii) Estimating the 3D Rigid transformation

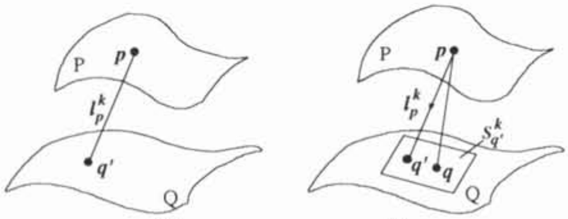


Fig. 3: Corresponding point.

The problem of finding the nearest point **q** on a geometric shape Q to a given point **p** has been discussed by Besl and McKay [6] where Q can be represented by a set of segments, implicit curves, parametric curves, triangular faceted surfaces, implicit surfaces, or parametric surfaces. In the present case Q is a digital surface arranged in a range image. If Q is big, the exhaustive search for the nearest point would be very time consuming. Another approach to this search problem is to relax the demand for finding the exact nearest point **q** on Q to **p**. This idea has been used by Potmesil [8] and Chen and Medioni [7].

The approach presented here for finding the corresponding point is based on the approach used by Chen and Medioni [7]. The idea is to approximate Q using its tangent $S_{q'}^k$ plane at **q'** (see Fig. 3b). That is:

$$\mathbf{q} = \min_{\mathbf{q}'' \in S_{q'}^k} d^2(\mathbf{p}, \mathbf{q}'') \qquad (1)$$

with:

$$S_{q'}^k = \left\{ \mathbf{q}'' \middle| \mathbf{n}_{q'}^k \cdot (\mathbf{q}' - \mathbf{q}'') = 0 \right\}, \quad \mathbf{q}' = l_p^k \cap Q \qquad (2)$$

where d is the Euclidean distance from a point to a plane, $l_p^k = \left\{ \mathbf{a} \middle| (\mathbf{p} - \mathbf{a}) \times \mathbf{n}_p = 0 \right\}$ is the line normal to P at **p** (see Fig. 3a), **q'** is the intersecting point of Q with l_p^k, and $n_{q'}^k$ is the normal to surface Q at **q'**. In order to reduce the computational cost, only a subset of the surface $\mathbf{p} \in S_p$, called the control points, is used in the registration process. Details about the algorithm can be found elsewhere [2].

Given a set of corresponding points, the least squares rotation matrix and the translation vector can be estimated using either the singular value decomposition (SVD) technique [9] or the quaternion method [10]. In both methods, the estimation of translation is based on the estimated rotation, i.e., the translation vector is a function of the optimal rotation matrix and the data measurements. Problems with such methods stem from cumulated errors in calculating the translation due to the previous errors in computing the rotation. This problem is solved using the concept of dual number quaternions [11].

Integration of different views can be conducted at various abstraction levels, for example at pixel level or at surface patch level. The problem with pixel level fusion is that the image structure may be lost once the fusion is performed. This reduces the use of efficient geometric techniques for extracting surface characteristics. Fusion at surface patch level is very difficult because it is hard to obtain an identical surface patch segmentation in two neighbouring views. The integration method used is based on a contour propagation technique. After having registered two neighbouring views, first a surface description is constructed from one view. The bounding contours are then transported to the second view using the 3D rigid transformation obtained from the registration process. These transported contours are embedded as constraints in the surface description of the second view.

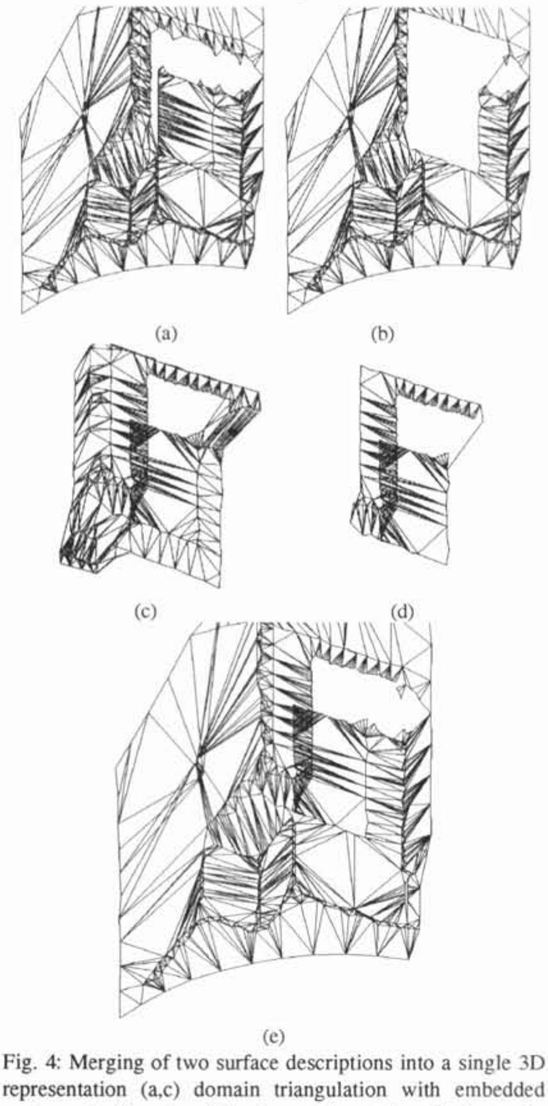


Fig. 4: Merging of two surface descriptions into a single 3D representation (a,c) domain triangulation with embedded contours (b,d) parts of the two surface triangulation to be fused (e) integrated view.

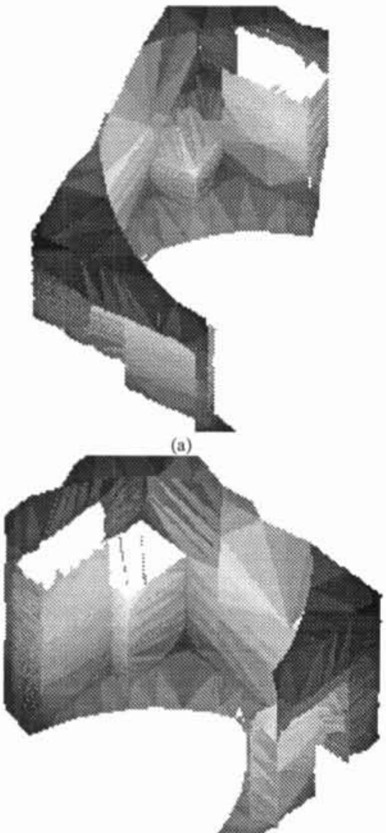
EXPERIMENTAL RESULTS

In the experimental part, results on the registration of three views are presented. Moreover, model construction resulting from the integration of the three different views is shown using real sensor data. In the experiments, the approximate rotations used as initial estimates are set

10 degrees off from the actual value. Typically the number of control points used was 120, and 8 iterations are required to obtain sufficient accuracy. The errors in pairwise registrations among the 3 views are given in table 1. These results are significant given the quantization noise of the LRF (typical resolution in depth is 2.5 cm), and *mixed-point* errors [2] distorting the image at specific locations: the top edge of the cupboard in Fig. 2 is not a straight line, but rather appears as a sawtooth.

Image Pair	Rotation	Translation
1-2	0.25	-0.18 0.46 -0.23
1-3	0.31	-0.37 -0.24 0.64

Table 1: Registration results: Errors in rotation (degrees) and translation for two images pairs.



(a)

(b)

Fig. 5: 3D shaded model of an office seen: (a) right view (b) left view.

In order to illustrate the integration of the previously registered images, the merging of two surface descriptions into a single 3D representation is shown in Fig. 4. 3D edge segments were first extracted from each image, and a curve was traced on the overlapping part of the two images. This curve was then mapped onto the second image coordinate system using the 3D rigid transformation obtained from the registration process. Triangular based surface descriptions were then constructed for both images, with the traced curve and its transformation embedded in the two surface descriptions as shown in Fig. 4a. Fig. 4b shows the parts of the two surface triangulation to be fused. A 3D shaded model of an office scene, resulting from the integration of surface descriptions from three views, is seen in Fig. 5.

CONCLUSIONS

The paper presents a 3D scene analysis system that is capable to model real-world scenes, based on data acquired by a Laser Range Finder on board of a mobile robot (distance range from 0.5 to 20 meters). In range images, surfaces and physical structural properties are captured with minimum distortion, unlike intensity images, in which depth is induced. This characteristic makes range data suited to build 3D models of objects, architectural features and in general real world scenes.

The quality of the final 3D model depends highly on the precision of the extracted features, and on the quality of the registration between the different images. Sensor resolution and problems of geometrical nature influence the quality of the extracted features. Both the extracted features and the registration process are also dependent on the type of environment that is used. Current results were obtained for an office environment. For this kind of scenes the modelling algorithm has been successfully applied. The registration working directly on the range data provided accurate registration.

REFERENCES

[1] J.G.M. Gonçalves, V. Sequeira - Application of Laser Range Images to Design Information Verification, Proc. of IAEA Symposium on International Safeguards, Vienna , 1994.

[2] V. Sequeira, J.G.M. Gonçalves, M.I. Ribeiro - 3D Environment Modelling Using Laser Range Sensing, Proc. of Int. Symposium on Intelligent Robotic Systems'94, Grenoble (France), pp 238-245, July 1994.

[3] X. Lebegue, J.K. Aggarwal - Automatic Creation of Architectural CAD Models, Proc. Second CAD-Based Vision Workshop, pp. 82-89, February 1994.

[4] Y.G. Leclerc, S.W. Zucker - The Local Structure of Image Discontinuities in One Dimension, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 9, n. 3, pp. 341-355, 1987.

[5] O.D. Faugeras, E. Le Bras-Mehlman, J.D. Boissonnat - Representing Stereo Data with the Delaunay Triangulation, Artificial Intelligence, vol. 44, pp. 41-87, 1990.

[6] P.J. Besl, N.D. McKay - A Method for Registration of 3-D Shapes, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-14, n. 2, pp. 239-256, February 1992.

[7] Y. Chen, G. Medioni - Object Modelling by Registration of Multiple Range Images, International Journal of Image and Vision Computing, vol. 10, n. 3, pp. 145-155, April 1992.

[8] M. Potmesil - Generating Models of Solid Objects by Matching 3D Surface Segments, Proc. 8th Int Joint Conf. Artificial Intelligence, Karlsruhe, W. Germany, pp. 1089-1093, August 1983.

[9] K. Arun, T. Huang, S. Blostein - Least-Squares Fitting of Two 3-D Points Sets, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-9, n. 5, pp. 698-700, 1987.

[10] O.D. Faugeras, M. Hebert - The Representation, Recognition, and Locating of 3-D Objects, Intern. Journal. of Robotics Research, vol. 5, n. 3, pp. 27-52, 1986.

[11] M.W. Walker, L. Shao, R.A. Volz - Estimating 3-D Location Parameters Using Dual Number Quaternions, CVGIP:Image Understanding, vol. 54, n. 3, pp. 358-367, 1991.