Automatic Generation of Indoor VR-Models by a Mobile Robot with a Laser Range Finder and a Color Camera

Christian Weiss and Andreas Zell

Universität Tübingen, Wilhelm-Schickard-Institut für Informatik,
Lehrstuhl Rechnerarchitektur, D-72076 Tübingen,
`cweiss,zell@informatik.uni-tuebingen.de`

Abstract. We present an intuitive method to generate visually convincing 3D models of indoor environments from data collected by a single 2D laser range finder and a color camera. The 2D map created from the laser data forms the basis of a line model which serves as floor plan for the 3D model. The walls of this model are textured using the color images. In contrast to panoramic images where the resolution of the textures is substantially decreased by the panoramic mirror, our standard camera can provide high quality textures at moderate image resolutions.

1 Introduction

Traditional 2D maps built using range sensors such as laser range finders or ultrasonic sensors provide enough data for navigation tasks, but lack visual information. This visual information can be very useful, especially when humans have to interpret the data collected by the robot. Textured 3D models of the environment give a faster and more detailed overview of a scene than a 2D map, and therefore can help to simplify tasks such as surveillance or rescue and also provide a way to interchange models of building interiors over far distances.

In this paper, we present a method to create 3D models based on 2D maps and color images, both obtained using a mobile robot. First, we extract a set of line segments from the 2D map to determine walls. This set serves as a floor plan for the 3D model. Then the 3D model is enriched by textures created from the color images. Note that our approach does not intend to include small objects like plants or chairs in the 3D model. On the one hand it is not possible to determine the height of objects with the horizontal laser range finder; objects located entirely below or above the scan plane are not detected anyway. On the other hand, the objects are visible in the textures which suffices for a fast overview.

A method that also uses a single horizontal laser range finder to build 3D models was presented by Biber et al. [1]. They use a panoramic camera to acquire the images from which the textures are created. In combination with a method by Fleck [2], which can create 3D point clouds from pairs of panoramic images, some more 3D information is added to the model [3]. A different approach was presented by Thrun et al. [4] who use a second vertical laser range finder to collect 3D range data. Planar surfaces extracted from this data are textured with color images recorded by a panoramic camera. Due to the wide field of view of the panoramic cameras used in these approaches, the resolutions of the textures are low. Our method uses a standard camera, so we can create detailed textures although the image resolution is moderate (e.g. 640×480 pixels).

The rest of this paper is organized as follows: Section 2 presents the acquisition of the data. Section 3 describes how the 2D map is transformed into a line model. The creation of the 3D model is described in Section 4 and Section 5 presents some experimental results. Finally, Section 6 gives a short conclusion.

2 Data Acquisition

In this work, we use an RWI (Real World Interface) ATRV-Jr mobile robot to collect the laser and image data. The robot is equipped with a SICK LMS 200 laser range finder and a stereo camera system, of which we use only one camera. The camera system is mounted on a pan tilt unit (PTU).

The robot is remotely driven around to build a 2D map using a SLAM (Simultaneous Localization And Mapping) technique proposed by Biber [5]. The final 2D map consists of a set of laser scan points in a global coordinate frame. Fig. 1(a) shows a 2D map generated at our institute. Note that the horizontal left wing is not at a right angle with the vertical main axis corridor in the real environment.

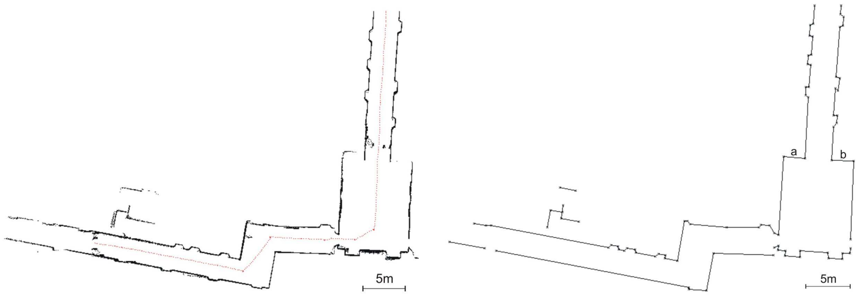
During mapping, the robot stops approximately every meter and takes three color images; one to the front of the robot, one 90° to the left and one 90° to the right. The different orientations of the camera are achieved by turning the PTU. Each image is saved together with the current pose estimate of the robot and the orientation of the PTU relative to the robot's orientation.

3 Line Model

The next step is to transform the 2D map, which is a set of unrelated 2D points, into a set of line segments. Each line segments represents one wall of the 3D model, i.e. each line segment is the projection of a wall onto the floor.

To extract lines from the 2D map, we use the Hough Transform [6,7]. Additionally, each line is associated with the set of 2D scan points which contribute to its creation. These point are orthogonally projected onto the line. The lines are then split into line segments at every location where two neighbouring points on the line have a distance larger than 10 cm.

After that, line segments which seem to belong to the same linear structure of the environment are merged. We use adaptions of methods proposed by Sack [8] and Schröter [9]. As Sack uses operations that are geometrically more exact and involve non-linear regression, his method is five to 12 times slower than Schröter's technique, but creates results that are more suitable for our purposes.



(a) 2D laser map consisting of laser scan points. The path of the robot is indicated by a dotted line.

(b) Final line model of the same environment.

Fig. 1. 2D laser map and corresponding line model.

More suitable means that the environment can be represented by fewer but longer line segments.

Next, the line model is further improved to meet our requirements. The steps to improve the model are the following, where steps 1 and 4 are adaptions of techniques proposed in [8]:

1. We adjust the length of segments whose endpoints are near each other. This ensures that walls forming the corner of a room meet in a common point.
2. Line segments shorter than 20 cm are deleted; for line segment that are not connected to another one, the minimal length is 80 cm. This step deletes many line segments that represent small structures like plants or chairs.
3. The orientation of each wall is determined, i.e. we decide which side of the wall faces the corridor. We use the scan points associated with the line segment and the directions towards the global poses of the corresponding scans to solve this task. This step is very important as the orientation of the texture on the wall is determined by the orientation of the wall.
4. Short line segments are inserted to close gaps. Only gaps smaller than 60 cm are closed.
5. Segments that share a common endpoint with more than one other segment are deleted. This step implements the constraint of the real environment that no more than two walls meet in the corner of a room.
6. Certain sequences of connected line segments are deleted. The overall length of these sequences must be below a threshold (1.5 m) and the individual segments must be short.

The improvement steps are fully automatic and provide a good representation of the walls in the environment. As it is possible that some walls are fully occluded during the acquisition process and therefore are not present in the laser map, the user is given the possibility to manually insert missing walls or to remove undesired ones. Fig. 1(b) shows the final line model for the map shown in Fig. 1(a). The line segments a and b were manually inserted as the corresonding walls never were inside the range of the laser range finder during data acquisition.

4 3D Model

The line model serves as a floor plan for the 3D model, where each line segment represents one wall. As the horizontal laser range finder cannot determine the height of the walls automatically, the height must be given as parameter. The images taken by the color camera are used to texture the walls.

The outline of the texture creation process is as follows:

1. Long walls are split into shorter subwalls.
2. For each subwall, a suitable image is determined from which the texture can be created.
3. The texture for the wall is determined as the section of the image which shows nothing but the wall.
4. The intensities of neighbouring textures are adjusted to hide seams.

In contrast to Biber et al. [1], who merge several panoramic images by multiresolution blending to create textures for long walls, we split long walls into shorter subwalls of fixed length l. If l is chosen properly, the texture for a subwall can be created from a single image which shows the entire subwall. In corridors, l is around 1.5 meters, in wider rooms it can be larger.

To be suitable as a texture for a wall, an image must meet at least three conditions. First, the photograph position p of the image must be located such that p is on the corridor or room side of the wall. This fact can be verified using the orientation of the wall. Second, the wall must not be occluded by other walls when looking from p towards the wall. Third, the wall must be seen entirely on the image. This can be verified by projecting the corner points of the wall onto the image and checking if they are inside the image. Finally, it is beneficial if the wall is displayed nearly frontally on the image; the angle between the viewing direction of the image and the orientation of the wall provides a measure for this condition. For such images, the perspective does not have to be corrected much. A perspective correction can drastically decrease the quality of the image in some image regions. If no frontal image is available, this condition is ignored. If no image is suitable at all, i.e. does not meet at least the first three conditions, the wall is left untextured. In situations where there is more than one suitable image, e.g. when the robot frontally approaches a wall, the image with the closest distance to the wall is selected.

After a suitable image for a wall has been found, we calculate the texture as the section of the image which shows nothing but the wall. This calculation is done by projecting the 3D corner points of the wall onto the image and cropping out the resulting image section. Fig. 2 illustrates the procedure. Note that in

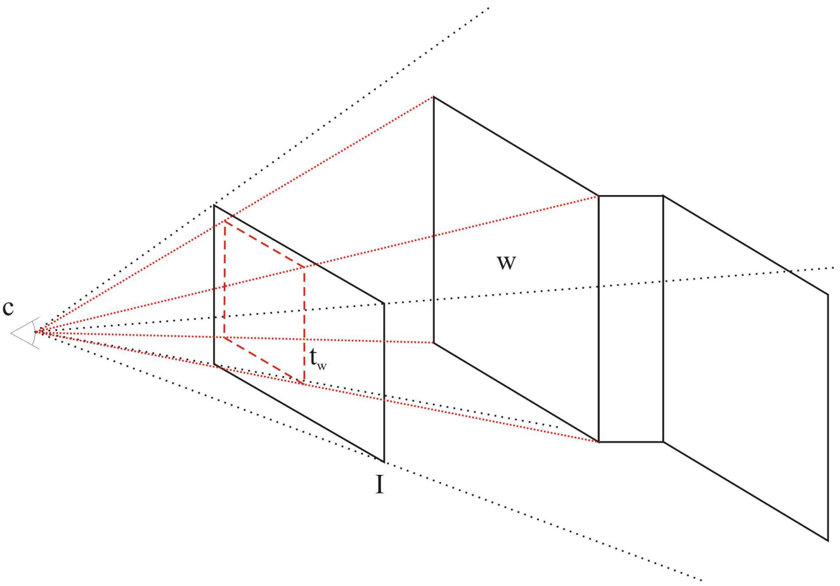


Fig. 2. Generation of the texture for a wall w. The image I, shot from camera position c, was identified as the most suitable image for w. The section t_w of I is cropped out and serves as texture for w. Note that in general, t_w is not a rectangle but a trapezoid.

general, the image section is a trapezoid. A rectification of this trapezoid corrects the perspective.

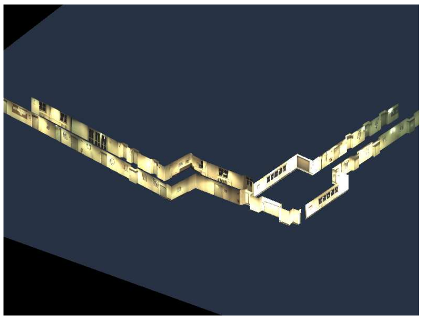
To build the projection matrix, we need the intrinsic and extrinsic parameters of the camera. The intrinsic ones are obtained by camera calibration; the extrinsic ones are given by the pose of the robot, the height of the camera above the floor and the orientation of the PTU at the time the image was taken.

5 Results

Fig. 3(a) partly shows the 3D model of a single room, the robots laboratory; the 36 color images on which the textures are based have a resolution of 1280×960 pixels. The creation of the line and 3D models took a total of about 10.4 s on a 3Ghz Pentium 4 PC (512MB of RAM). Once the 3D model is created, real-time walkthroughs are possible. The same holds for the next data set.

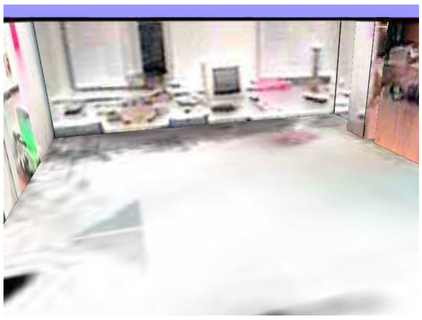
This data set is based on the map shown in Fig. 1. It contains a total of 195 images (640×480 pixels), which were taken at 65 different locations. The resulting 3D model was created from the laser map in about 43.2 s using the adaption of [8] for line merging. Fig. 3(b) shows a bird's eye view of the 3D model and Fig. 3(c) shows a detailed view. As our camera is not capable of adjusting its exposure automatically, the intensities of the textures differ throughout the 3D model due to different lighting conditions in the real environment. In Fig. 3(c) you can also see that in reality, indentations for doors are only as high

(a) 3D model of the robot's lab.



(b) 3D model based on Fig. 1.

(c) 3D model based on Fig. 1.



(d) 3D model created by Biber.

Fig. 3. Different views of automatically generated 3D models. (a)-(c): generated using our method. (d): taken from a 3D model presented by Biber et al. in [1]. Thanks to Peter Biber for the permission to use the image of his 3D model.

as the doors themselves. Our horizontal laser range finder cannot detect such situations, so the geometry of the virtual walls is wrong, but the 3D model still looks good.

However, in more complex environments with many objects in the middle of a room or occluding entire walls, our approach would fail to detect the walls. Here, additional 3D information from another source like a second laser range finder or a stereo camera is needed.

Fig. 3(d) shows a detailed view of a small part of the 3D model presented by Biber et al. in [1]. In the picture, you can see a desk and a monitor standing on it, but the rest is not recognisible. In our model, small objects like ring binders or the checkerboard used for camera calibration are identifiable (Fig. 3(a) on the right). In Fig. 3(c), posters are visible in detail. Note that in Biber's 3D model, the textures of most other walls are worse than the ones shown in Fig. 3(d).

6 Conclusion

We presented a method to create visually convincing 3D models of indoor environments. The 3D model is based on a 2D map, which is built using a single laser range finder. We then transform the point-based 2D map into a set of line segments, where each segment represents a rectangular wall. Textures for the walls are created from images obtained by a standard color camera. The relatively high-quality textures provide more visual detail than the textures used in other work. Furthermore, the creation of the 3D model is fast and needs no or few manual intervention.

As the depth information gathered by the horizontal laser range finder is bound to a single plane, our method is best suited to corridor-like environments and rooms with few objects occluding the walls. Objects different from walls are not geometrically represented in the 3D model but visible in the textures. To overcome this limitation, further 3D information could be used to improve the model. This additional information could be obtained e.g. by a stereo camera system or a second laser range finder.

References

1. Biber, P., Andreasson, H., Duckett, T. and Schilling, A.: 3D Modeling of Indoor Environments by a Mobile Robot with a Laser Scanner and Panoramic Camera. Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), 2004
2. Fleck, S., Busch, F., Biber, P., Andreasson, H. and Straßer, W.: Omnidirectional 3D Modeling on a Mobile Robot using Graph Cuts. Proc. IEEE Int. Conf. on Robotics and Automation (ICRA), 2005
3. Biber, P., Fleck, S., Busch, F., Wand, M., Duckett, T. and Straßer, W.: 3D Modeling of Indoor Environments by a mobile Platform With a Laser Scanner And Panoramic Camera. Proc. 13th European Signal Processing Conference (EUSIPCO), 2005
4. Thrun, S., Martin, C., Liu, Y., Hähnel, D., Emery-Montemerlo, R., Chakrabarti, D. and Burgard, W.: A Real-Time Expectation Maximization Algorithm for Acquiring Multi-Planar Maps of Indoor Environments with Mobile Robots. IEEE Trans. on Robotics and Automation, 20(3), pp. 433-442, 2003
5. Biber, P. and Straßer, W.: The Normal Distributions Transform: A New Approach to Laser Scan Matching. Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), 2003
6. Hough, P.C.V.: Methods and Means for Recognising Complex Patterns. US Patent 3069654, 1962
7. Ballard, D. H. and Brown C. M.: Computer Vision. Prentice Hall, Englewood Cliffs, NJ, USA, 1982
8. Sack, D.: Lernen von Linienmodellen aus Laserscannerdaten für mobile Roboter. Diplomarbeit, Albert-Ludwigs-Universität Freiburg, Institut für Informatik, Arbeitsgruppe Autonome Intelligente Systeme, 2003 (german)
9. Schröter, D., Beetz, M. and Gutmann, J.-S.: RG Mapping: Learning Compact and Structured 2D Line Maps of Indoor Environments. Proc. Int. Workshop on Robot and Human Interaction Communication (ROMAN), 2002