Automatic 3D Reconstruction of Indoor Manhattan World Scenes using Kinect Depth Data*

Dominik Wolters

Institute of Computer Science, Kiel University, Germany

Abstract. This paper discusses a system to reconstruct indoor scenes automatically and evaluates its accuracy and applicability. The focus is on the realization of a simple, quick and inexpensive way to map empty or slightly furnished rooms. The data is acquired with a Kinect sensor mounted onto a pan-tilt head. The Manhattan world assumption is used to approximate the environment. The approach for determining the wall, floor and ceiling planes of the rooms is based on a plane sweep method. The floor plan is reconstructed from the detected planes using an iterative flood fill algorithm. Furthermore, the developed method allows to detect doors and windows, generate 3D models of the measured rooms and to merge multiple scans.

1 Introduction

Today there are no simple and cost-effective systems available to map interiors. Acquisition is usual performed either manually or with 3D laser scanners. Laser scanner achieve a high level of accuracy, however, the acquisition is costly.

The purpose of this research is the development of a low-cost solution for automatic mapping of empty or slightly furnished interiors. For a few years, with the Kinect sensor, a new, light and cost-effective sensor is available that can provide both color and depth images of the environment in real time. The used active structured light method ensures the acquisition of depth data also in poorly textured and dark areas. An automatic movement of the sensor performed by a pan-tilt unit guarantees a full coverage of the environment.

Related Work. Since floor plans are often needed for robotic navigation tasks, much research has been done on mapping of building interiors. The maps are usually captured with laser scanners during the movement of the robots. Example approaches are region growing using surface normals [6], plane sweeping [2, 7] or extracting walls using a Hough transform [10, 1].

Geometry estimation of indoor environments from single Kinect images has been studied in several works [12, 8]. In the last years 3D reconstruction from multiple images with a handheld Kinect has been a popular task [9, 4].

* Recommended for submission to YRF2014 by Prof. Dr.-Ing. Reinhard Koch

Contribution. The developed method allows an automatic acquisition and generation of floor plans. It is based on an approach proposed by Budroni and Böhm [2], which was adapted and extended in this work for use with data from the Kinect sensor. The described plane sweep approach for the detection of walls has been extended by an edge-based method that allows the detection of smaller and partially covered wall surfaces. The floor plan is determined from the detected planes using a novel iterative flood fill algorithm. Furthermore, a method for the detection of doors, including their opening direction, and windows is introduced. A three-dimensional reconstruction allows the detection of block-like objects in the room. To enable the reconstruction of larger rooms a histogram-based merging of multiple scans is presented.

2 Method

2.1 Equipment Overview

The equipment consists of a tripod on which a pan-tilt unit is mounted with the Kinect sensor on the head. The system is intended to capture the entire area from a single point. For this purpose, a gradual horizontal rotational movement is performed and at each position images with different angles of inclination are recorded. The data processing is performed on a connected laptop.

Based on the known rotation angle of the pan-tilt unit and Kinect sensor as well as the transformation between the camera coordinate system and the pan-tilt coordinate system (determined by a hand-eye calibration), a point cloud of the environment is created.

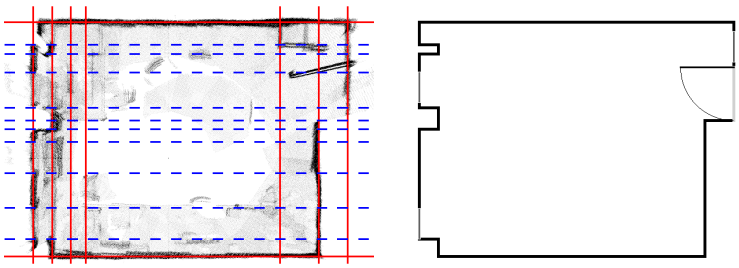
2.2 Modeling of Floor Plans

Plane-Sweep. The point cloud is used as input for the plane sweep method according to Budroni and Böhm [2]. First the alignment of the point cloud is determined using the Manhattan world assumption [3], which states that the scene contains three orthogonal, dominant directions. Then the main wall planes are detected by plane sweeping along this directions.

The rotational sweep used for the alignment proved to be error-prone in case of the noisy data of the Kinect sensor and additional spatial structures that do not correspond to the Manhattan world. Therefore, this step was replaced by an entropy based method similar to the one used by Gallup et. al. [5].

Edge-based Plane Detection. The developed edge-based plane detection allows to detect planes that can not be detected with the plane sweep approach due to a small size or because of occlusion (Fig. 1 (a)).

To determine the edges, histograms of the point coordinates distributions along the detected planes are created. For each histogram bin, the gradient is determined by discrete differences. Areas with high gradients represent potentially the beginning or end of a wall surface, which indicates that orthogonal walls are linked.



(a) (b)

Fig. 1. (a) Projection of the point cloud with planes detected by the plane sweep (red/solid) and edge-based plane detection (blue/dashed). (b) Determined floor plan

Orthogonal planes are determined for all significant minima and maxima of the gradient. The planes are added to the previously detected planes, sorted by the strength of the gradient. However, to prevent duplicate or closely spaced planes, they are only added if there is not already a plane present in a given neighborhood.

Iterative Interior Detection. For the determination of the interior, an iterative flood fill algorithm has been developed. The algorithm employs the rectangular cells, caused by cutting the detected planes. In the first step, the cell which contains the recording system is labeled as interior. Then the edges to the 4-neighbors are examined. If the number of points around the edges is below a threshold, the neighboring cell is considered as part of the interior. Similarly, the edges of the newly added cells are examined (Fig. 2). The edges to neighboring cells are only examined if they have a greater distance to the starting cell than the currently considered cell. The purpose is that only edges are examined, which can be seen from the recording location.

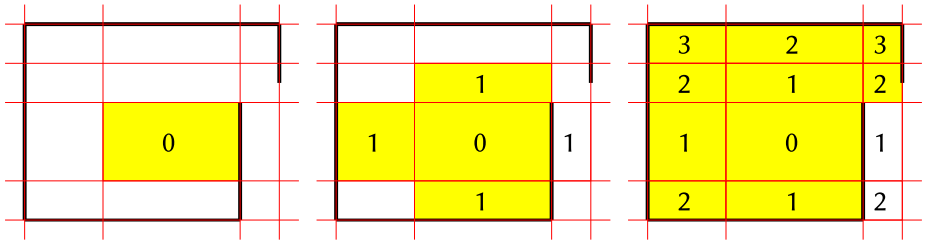


Fig. 2. Process flow of the iterative interior detection. The current interior cells are highlighted in yellow. The numbers indicate the distance to the starting cell

Door and Window Detection. The technique to detect door and windows is an extension of the method described by Budroni und Böhm [2], so that a distinction between doors and windows is possible and the opening direction of the doors can be determined.

In order to do this a 50 cm wide horizontal stripe in about 1 m height on the detected wall surfaces is selected and checked for gaps. To distinguish between doors and windows the areas with detected gaps are examined on the entire height. The classification is based on the dimension of the gap.

To detect the opening direction of doors, further demands on the measurement environment are required. The basic idea is to provide a rotational sweep around the door frames. If the door is opened and visible from the recording location, it generates an additional peak, which does not correspond to the direction of the wall.

Floor Plan. The final result is shown in figure 1 (b). The presented method is generally suitable to map unfurnished and geometrically simple interiors. In order to map slightly furnished or more sophisticated rooms an extension of the proposed algorithm is necessary.

2.3 3D Reconstruction

The 3D reconstruction allows the detection of block-like objects in the room. Furthermore a virtual 3D model of the room can be created. The colored images captured by the Kinect sensor can be used to texture the model.

The basic algorithm substantially remains unchanged for the 3D reconstruction. Additionally vertical planes are determined, so that the cells conform to cuboids. These cuboids are used for the iterative interior detection. Analogous to the 4-connected neighborhood in the 2D reconstruction, the 6-connected neighborhood of the cuboid is considered here.

2.4 Merging of Multiple Scans

For more sophisticated rooms or room sizes above the recommended range for the Kinect sensor, mapping from a single recording location can result in erroneous reconstructions. Therefore, methods to merge multiple scans were evaluated.

A histogram-based approach that utilizes the Manhattan world assumption was chosen to determine a globally optimal alignment. The basic idea is that the distribution of points in two partially overlapping similarly oriented scans is comparable. Corresponding methods are presented in [13, 11]. Advantages of this approach compared with the ICP algorithm is the efficiency and the determination of a global optimum.

3 Results

3.1 Experiments with Synthetic Data

The evaluation is performed first on synthetic data. Point clouds with random size between 4 m×3 m×2.5 m and 5 m×4 m×3 m were generated. To simulate the noise of the Kinect sensor, different levels of Gaussian noise are added. With a low noise level ($\sigma \leq 20$ mm), the deviations for the plane sweep and the edge-based plane detection are less than 10 mm. The deviation for the rotation determination is less than 0.1°.

3.2 Experiments with Real Data

Distance Experiments. To evaluate the influence of the distance to the sensor, planar surfaces of walls were measured at various distances from 0.3 m to 6.2 m. The depth resolution decreases quadratically with increasing distance from the sensor. Thus, the detection of an unambiguous wall plane is more difficult with an increasing distance. In the study, the relative deviation between measured and real wall distance is usually less than 1%. The ground truth was manually measured with a laser distance meter. Truly orthogonal walls are assumed.

Determination of Floor Plans. As a first example a reconstruction of an empty room with a size of about 3 m×4 m is considered. Fig. 3 shows the projected point cloud (left) and the determined floor plan (right). The door and the two windows were correctly detected. The comparison between real and determined room sizes for the main planes, i.e. length, width and height, results in only small deviations of less than 1%. These accuracies were confirmed also in the evaluation of other rooms.

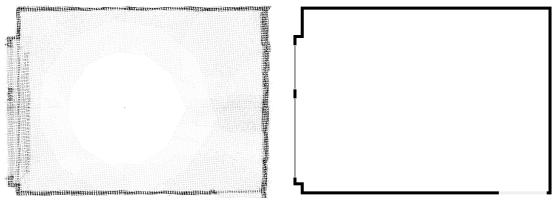
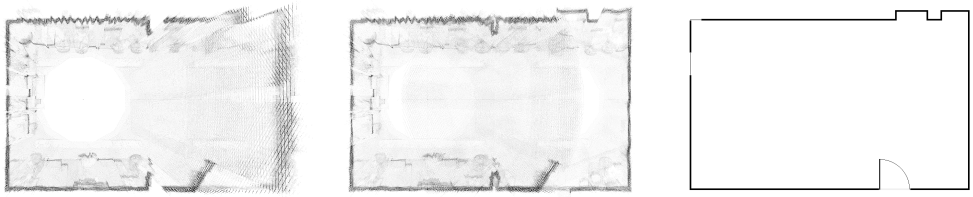


Fig. 3. Point cloud and determined floor plan of an empty room. Solid walls: thick lines. Door and windows: thin lines

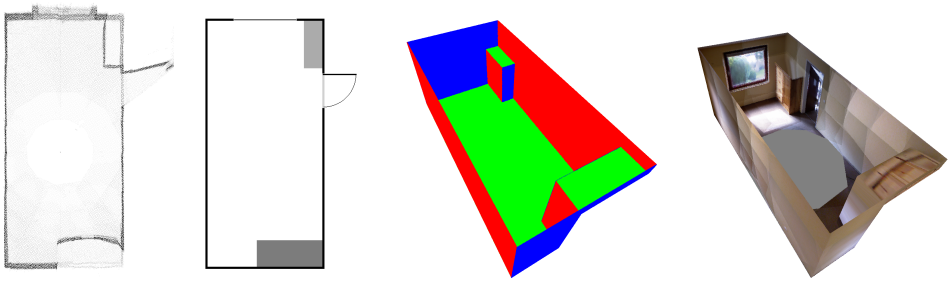
Merging of Multiple Scans. For large rooms merging of multiple scans can be useful to achieve a higher accuracy. Fig. 4 shows a room with a size of about 8.5 m×5 m. Clearly visible is the high noise level at the farther wall, i. e. the right-hand wall shown in Fig. 4 (a). By merging two scans (Fig. 4 (b)) the deviation of the room length was reduced from 1.1% to 0.8%.



(a) 1. Scan (b) Merging (2 Scans) (c) Floor plan

Fig. 4. Merging from two scans to determine the floor plan

3D Reconstruction. Figure 5 shows the point cloud of a slightly furnished room, the determined floor plan and the created 3D models. The size of the room is about 6 m×3 m. Door and window as well as the objects in the room were correctly recognized. A comparison of the real dimensions of the objects with the determined measures results in only minor deviations of up to 25 mm.



(a) Point cloud (b) Floor plan (c) 3D model (d) Textured 3D model

Fig. 5. 3D reconstruction of a slightly furnished room

4 Conclusion

The developed algorithm provides an automatic acquisition and generation of floor plans that reaches an accuracy which is already sufficient in many fields. The relative deviations between measured and real room sizes are usually less than 1%.

The analysis shows that normally only one scan per room is required for the mapping of living areas. The total duration for the generation of a floor plan from one recording location is usually less than 3 minutes. In larger rooms, such as living rooms, with a length of more than 7 meters, it is useful to merge multiple scans from different positions in order to get more accurate results.

References

1. Adan, A., Huber, D.: 3D reconstruction of interior wall surfaces under occlusion and clutter. In: International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT). pp. 275–281 (2011)
2. Budroni, A., Böhm, J.: Automatic 3D modelling of indoor manhattan-world scenes from laser data. In: Proceedings of the ISPRS Commission V Mid-Term Symposium 'Close Range Image Measurement Techniques'. vol. XXXVIII, pp. 115–120. Newcastle upon Tyne, UK (2010)
3. Coughlan, J.M., Yuille, A.L.: Manhattan world: Compass direction from a single image by bayesian inference. In: Proceedings of the Seventh IEEE International Conference on Computer Vision. vol. 2, pp. 941–947 (1999)
4. Du, H., Henry, P., Ren, X., Cheng, M., Goldman, D.B., Seitz, S.M., Fox, D.: Interactive 3D modeling of indoor environments with a consumer depth camera. In: Proceedings of the 13th International Conference on Ubiquitous Computing. pp. 75–84 (2011)
5. Gallup, D., Frahm, J.M., Mordohai, P., Yang, Q., Pollefeys, M.: Real-time plane-sweeping stereo with multiple sweeping directions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR '07). pp. 1–8. Minneapolis, USA (Jun 2007)
6. Hähnel, D., Burgard, W., Thrun, S.: Learning compact 3D models of indoor and outdoor environments with a mobile robot. Robotics and Autonomous Systems 44(1), 15–27 (2003)
7. Johnston, M., Zakhor, A.: Estimating building floor-plans from exterior using laser scanners. In: SPIE Electronic Imaging Conference, 3D Image Capture and Applications. vol. 3 (2008)
8. Neverova, N., Muselet, D., Trémeau, A.: 2 1/2D scene reconstruction of indoor scenes from single RGB-D images. In: Tominaga, S., Schettini, R., Trémeau, A. (eds.) Computational Color Imaging, pp. 281–295. No. 7786 in Lecture Notes in Computer Science, Springer Berlin Heidelberg (2013)
9. Newcombe, R.A., Davison, A.J., Izadi, S., Kohli, P., Hilliges, O., Shotton, J., Molyneaux, D., Hodges, S., Kim, D., Fitzgibbon, A.: KinectFusion: real-time dense surface mapping and tracking. In: 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR). p. 127–136 (2011)
10. Okorn, B., Xiong, X., Akinci, B., Huber, D.: Toward automated modeling of floor plans. In: Proceedings of the Symposium on 3D Data Processing, Visualization and Transmission. vol. 2 (2010)
11. Rofer, T.: Using histogram correlation to create consistent laser scan maps. In: Proceedings of the IEEE International Conference on Robotics Systems (IROS-2002). pp. 625–630. Lausanne (2002)
12. Taylor, C., Cowley, A.: Parsing indoor scenes using RGB-D imagery. In: Proceedings of Robotics: Science and Systems. Sydney (Jul 2012)
13. Weiß, G., Wetzler, C., von Puttkamer, E.: Keeping track of position and orientation of moving indoor systems by correlation of range-finder scans. In: Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems '94. 'Advanced Robotic Systems and the Real World', IROS '94. vol. 1, pp. 595–601 (1994)