Using an attributed 2D-grid for next-best-view planning on 3D environment data for an autonomous robot

Marcus Strand and Rüdiger Dillmann
Institute for Computer Science and Engineering (CSE)
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
{*strand, dillmann*}@*ira.uka.de*

***Abstract*—The usage of geometrical 3D-models enables humans to plan and design different aspects of indoor environments. This article describes the next-best-view planner for an autonomous 3D modelling robot using a 3D laser scanner. Besides the human user, the constructed 3D-model can also be used by other autonomous robots for navigation and localization tasks. The process of building a 3D-model without any geomatical pre-knowledge leads to a planning scheme which uses the 3D data. This article copes with that problem by using and attributed 2D grid combined with an intelligent action planning scheme for indoor environments.**

***Index Terms*—3D Data Processing, Path Planning, SLAM, autonomous systems**

I. Introduction

Knowledge of the 3D environment of a mobile robot is a keystone to achieve real autonomy. Since these models are not always available in any working scenario, the robot has to build these models at least partially by itself. In recent years several 3D-models have been introduced to make reliable robot navigation possible. While 2D-models still lack in expressiveness for robust navigation, 3D-models cover the full workspace of the robot and hence deliver the complete geometrical information of an environment. With a 3D model, for example localization also in dynamic environments is possible if the mobile robots relies on upper areas of the 3D-model (e.g. above doors) since they mostly do not change.

The 3D-model is built autnomously by capturing successive depth images. The origin of these depth images can be estimated by using the robot's odometry and a registration algorithm. Through the integration of vision data from a monucular camera, texture and color data is attached to the depth image. Using this color information for the registration process leads to better registration results and hence to models, which are more consistent. In order to plan the next best views, the full 3D information should be considered.

Different variations of moving 2D scanners have already been proposed for the modelling task. In [1], [2], [3], single scan lines are taken with a horizontal movement. This approach can generate a good model of the environment, but cannot avoid collisions. Depth image based approaches can be found in [4],[7], [5]. Here first steps towards autonomy or semi-autonomy are taken. The taken depth images show a high resolution in (for the robot) uninteresting areas like the ceiling [4] or the area laterally [7] of the robot. As stated above it is important for us to have the highest resolution in the heading direction of the robot, especially for collision avoidance and model building with a single sensor device. The current commercially avaible 3D scanners have a very small viewing cone, which is bad for registration and do not support foveal vision. To overcome these problems we developed a 3D sensor called RoSi (Rotating Sick) which is a common Sick LMS200 scanner rotating continuously around its optical axis.

The problem of finding the optimal position and orientation for the next measurement under different constraints is known as next-best-view problem. In [10] the position was determined with the help of a virtual sphere around the object. A similar problem is the *Art-Gallery-Problem* which optimizes the positions of security guards in an art gallery to cover the complete environment [11]. Due to their compactness and robustness occupancy grids are widely used for 2D environments [12].

II. Data Capture and Integration

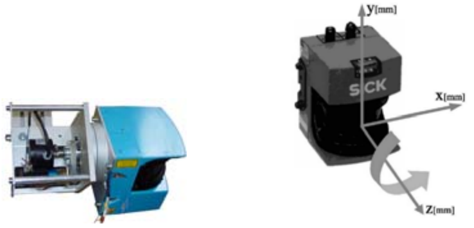
A. Detection of a point cloud

To comply with the requirements stated above, a 3D laser scanner was developed. This laser scanner consists of a commercial Sick LMS200 attached on a rotation axis for continious rotation. For data and power transmission a rotary feedthrough with sliprings is used. Figure 1 shows the developed rotating sick (Rosi) sensor and the corresponding coordinate frame. According to [8] the cartesian coordinates of a point $\vec{x} = (x, y, z)^T$ are given as

$$\begin{aligned} x_i &= \cos\beta \cdot \cos\alpha_i \cdot d_i \\ y_i &= \sin\beta \cdot \cos\alpha_i \cdot d_i \\ z_i &= \sin\alpha_i \cdot d_i. \end{aligned}$$

Here d_i is the measured distance, α_i the measurement angle in the scanning plane and β the rotation angle around the axis.

The resolution of the point cloud depends on the scanner configuration and the rotation velocity $\dot{\beta}$. The viewing cone



(a) Rosi Sensor (b) Coordinate Frame

Fig. 1.

is $100°$ or $180°$ (which corresponds to one hemisphere). Figure 2 shows a real scene and the detected point cloud from different views with viewing cone of $100°$.

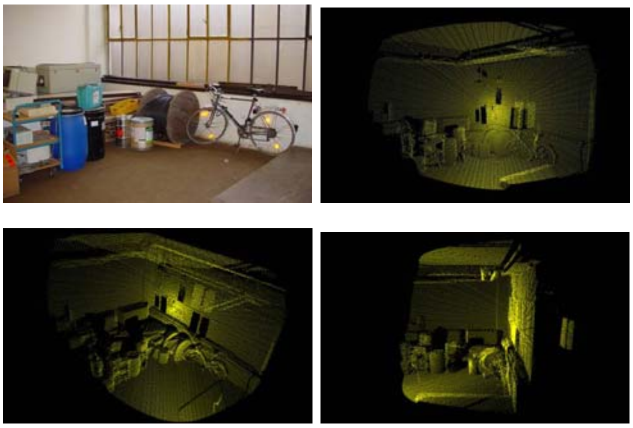
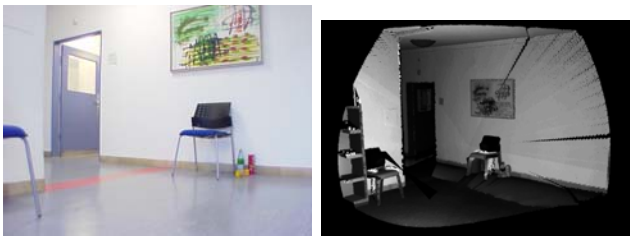


Fig. 2. Real scene and detected point cloud

Additionally the remission value of each point is read out and attached to the point. This value corresponds to the amaount of reflected light of the object in the laser wave length of $\lambda = 905nm$. Figure 3 shows a real scene and the detected point cloud with the remission value as grey value of each point.



(a) (b)

Fig. 3. Real scene (a) and detected point cloud with remission values (b)

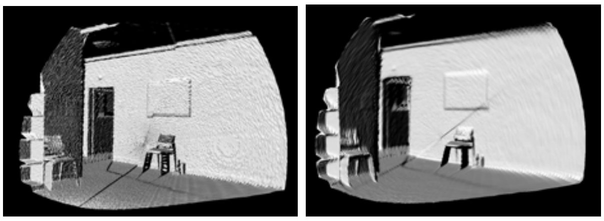
B. Preparation of point cloud

Invalid points have to be filterd out of the point cloud. A point is invalid if it fulfills at least one of these condidtions:

- The point is out of range
- The laser beam is reflececed by two objects (one in foreground and one in backgroud)
- The point is part of the mobile robot

To reduce the gaussian noice in the point cloud, a mean value filter is applied to the points. The operation is carried out on the cartesian coordinates along the scanlines i and scanrows j with filter width of b. For each point $\vec{x}_{ij}$ a smoothed point $\vec{x}_{ij}^{s}$ is calculated. An iterative application of the filter leads to a smoother image. In order to retain the basic structures, the final point $\vec{x}_{ij}^{*}$ is only assigned to its smoothed value if the euclidian distance to the original point is below the threshold δ.

In a further step the point cloud is triangulated and hence surface patches built. The triangulation algorithm uses known neighborship relations between the points, due to the specific architecture of the sensor system. The triangles are built by connecting two neighboring points in the same scan row ($\vec{x}_{i,j}$ and $\vec{x}_{i+1,j}$) with the corresponding point in the next row ($\vec{x}_{i+1,j+1}$) (respectively the former row, $\vec{x}_{i,j-1}$. A variable range threshold avoids the system to triangulate areas which do not belong to a reflected object. Figure 4 shows a raw triangulated depth image (a) and a smoothed triangulated depth image (b).



(a) (b)

Fig. 4. raw (a) and smoothed (b) depth image

Additionally vision data from an attached camera is integrated into the depth image. Therefore a calibration between the two coordinate frames is necessary. The calibration matrix is determined through a correlation between the remission values and the intesity of the color image. Figure 5 shows the resulting depth image with a textured surface.

C. Data Integration

To integrate the collected data into one common coordinate frame, the depth images are registrated with an ICP-Algortihm [6]. According to the measuring method of the RoSi scanner the center of range images has the highest point resolution which decreases at the boarders. This is a advantageous property for collision detection however, this raises difficulties for the registration of two range images. The influences of the center are enormous on the registration

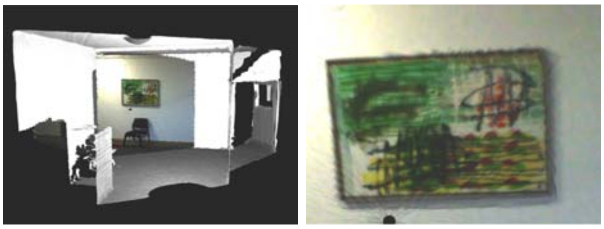

(a) (b)

Fig. 5. Depth Image with texture (a) and in detail (b)

results, due to the high amount of points. If the centers of the range images are far away from each others and the overlapping area is low, it occurs that the points in the center pull the registration results in the wrong way. To comply with this an octree based matching strategy was designed [9]. This matching strategy merges the selection and matching stages of the ICP algorithm in one stage and hence supersedes the selection stage. In our matching strategy we start with the contours of the scene and register them until it converged in a local minimum. Instead of terminating the algorithm we refine the contours with more details from the range images and register the two restructured surfaces again. This variant of the ICP algorithm repeats this procedure until it reaches the last refinement level and terminates there.

For better registration results the remission data and the vision data are integrated into the distance function. To accelerate the computation of the pairwise distances, a gridbased pairing strategy according to [9] is used. Figure 6 shows sample point clouds from above together with the registration results.

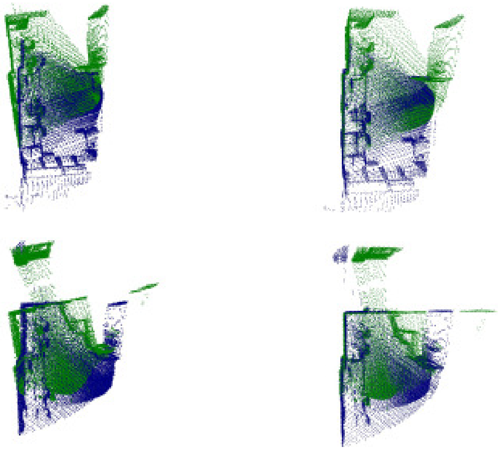


Fig. 6. Point cloud with origin estimated from odometry (left) and after registration (right)

With every registration of two range images a small registration error remains. To minimize this error in a captured loop, the error is distributed equally on the involved range images (loop closing). Therefore the homogenous transformation matrices ${}^0T_{i+1}...{}^0T_j$ of each depth image are modified so that

$$I = \prod_{k=i}^{j-1} \left({}^kM_{k+1}\right) \cdot \; {}^jM_i$$

with I as homogenous unit matrix through interpolation. To interpolate between rotations, a $SLERP$-Interpolation is used.

Figure 7 shows depth images in an open loop (left) and after the loop closing process in a closed loop (right).

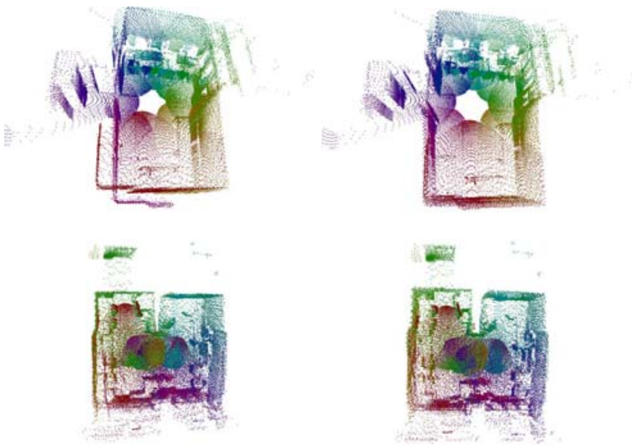


Fig. 7. open loop (right) and closed loop (left)

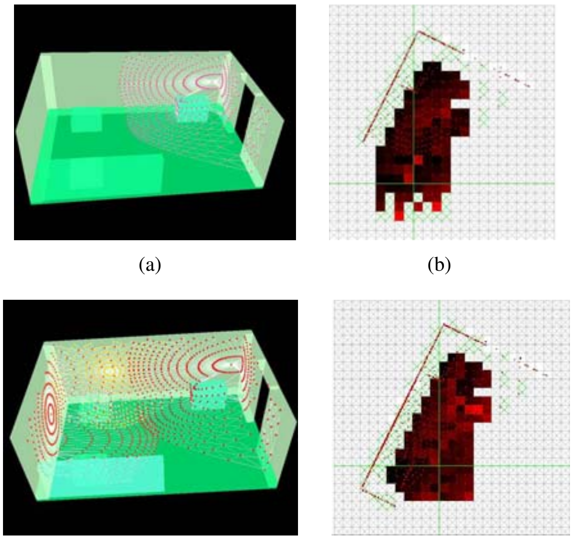
III. Action Planning

A. Next best view

For the determination of the next best view the sensed data has to be segmented into different areas (like unknown, occupied, passable). To do this efficiently the available large amount of data is reduced to a planning model. With the planning model, next views can be calculated involving different quality criteria (like length of travelling distance, amount of overlap). The planning model is a 2D-grid which conserves the available navigation-specific information of the 3D world in cell attributes through projection onto the grid. By means of that, the advantages of a 3D representation (geometrical information in every height) and 2D representation (easy to process, high data reduction) can be combined. Figure 8 shows a 3D model and the corresponding 2D grid in a simulation environment.

The cell attributes can be divied into direct attributes and indirect attributes. Direct attributes determine their value directly from the depth images. This group contains

- **isExplored:** is $true$ if the cell contains already a scan point
- **min,max:** define the minimal an maximal height values of a cell



(a) (b) (c) (d)

Fig. 8. Point clouds in a simulation environment (a),(c) and the corresponding 2D planning model (b),(d)

- **hasLowObstacle:** is $true$ if the cell contains scan points which leads to a collision with the exploration platform
- **hasHighObstacle:** is $true$ if the cell contains scan points which are higher than the exploration platfrom

The group of indirect attributes contains

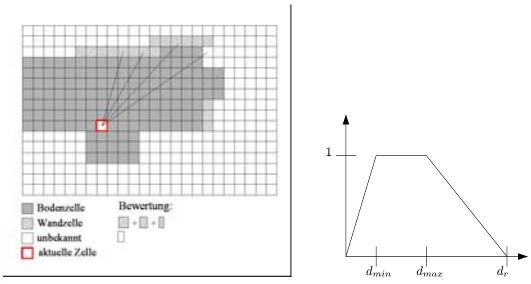
- **isWall:** is $true$ if the cell contains a high and a low obstacle
- **isGround:** is $true$ if the cell contains no low obstacles
- **isNearObstacle:** is $true$ if the cell contains no obstacles but is close to cells with low obstacles

B. Rating of cells

For every cell with the $isGround$ attribute set, horizontal scans with variable step width are simulated. For the determination of the rating G of a simulated scan, several simulated laser beams are sent, and the attributes of every cell it traverses is considered. The beam is followed until is reacheds a cell with attribute $isWall$ (Hit: t_h) or $!isExplored$ (Hit: t_u). Both attributes lead to a termination of the beam. Figure 9a shows the process weighting with simulated scans.

Through a distance dependend weighting, cells in undesired areas can be weighed less. These undesired areas are for example cells where the sensor is very close to obstacles or far away from any explored cell. Figure 9b shows a possible weighting function, a piecewise linear function $g(d_t)$ with

$$g(d_t) = \begin{cases} \frac{1}{d_{min}} \cdot d_t & 0 \leq d_t \leq d_{min} \\ 1 & d_{min} < \ d_t \leq d_{max} \\ -\frac{1}{d_r - d_{max}} \cdot d_t & d_{max} < d_t \leq d_r \end{cases}$$



(a) (b)

Fig. 9. Simulated scans of a cell (a) and weighting function (b)

with d_r as scanner range.
For both termination types the hits t_u and t_h are summerized with the weight w_h resp. w_u. The rate of the current cell and orientation is calculated with

$$G = \sum w_h \cdot g_h \cdot t_h + \sum w_u \cdot g_u \cdot t_u.$$

The weights w_h resp. w_u depend on the current exploration strategy and the resulting constraints (for example size of overlapping area).

In order to calculate the simulated scans only once for all orientations, the values for one cell are determined in one step. Therefore the simulated laser beams are sent in 360° and the termination type determined. When calculating the rating function, only the laser beams, which lie within the scan area are summerized.

C. Exploration strategy

The indoor environment, which is explored is supposed to consist of rooms and floors. Since the maximal scanner range is at 8 m the explored rooms should have an extent below 8 m. If there is a high amount of unreflected laser beams (and hence a high amount of points at the 8 m range) the system will detect a floor. If most of the laser beams are reflected the system detects a room. Besides rooms and floors a collision-free navigation through doors has to be managed separately.
To take advantage of the loop closing operation without handling complex registration graphs the exploration strategy is split into two stages. In the first stage a loop with several overlapping depth images is created and closed through the loop closing algorithm. Missing areas of a room or floor (e.g. through occlusion) are sensed in the second stage an registered against the loop-closed depth images. In a last step open doors are searched to explore a new room or floor. Figure 10 depicts this strategy.

To apply the different stages several modi have been implemented:

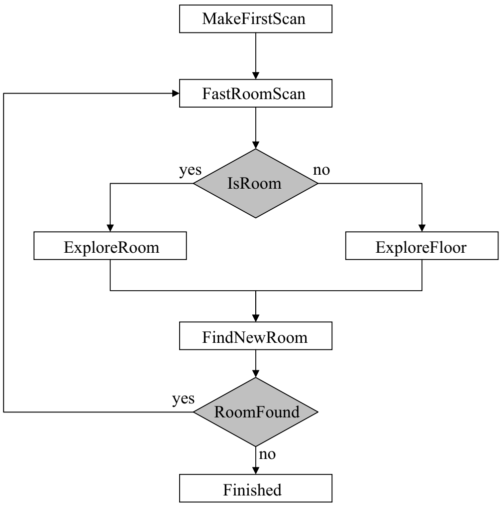


Fig. 10. Diagram of the exploration strategy

- **MakeFirstScan:** This is the inital state of the system. One depth image is taken from the current position.
- **FastRoomScan:** This is the first stage. n depth images are taken with an angular distance of $\frac{360}{n}$ at one position. After that the loop is closed and the decision between room and floor is made.
- **ExploreRoom:** With help of the room quality function and the 2D-grid the rest of the room is explored. The corresponding depth image for registration is determined on the 2D-grid.
- **ExploreFloor:** With help of the floor quality function and the 2D-grid the rest of the floor is explored. The corresponding depth image for registration is determined on the 2D-grid.
- **FindNewRoom:** New doors a found on the 2D-grid and a new room is approached.
- **Finished:** If no open door is found the exploration is finished.

D. Modelling doors

Due to the extent of the exploration plattform, the transition of doors has to be considered seperately to avoid collisions. Doors can easily be extracted from the attributed 2D grid. A cell z_t which lies within a door on the ground contains no low obstacle (attribute *hasLowObstacle*) but contains a high obstacle (attribute *hasHighObstacle*), because above doors there is a wall or a fanlight. Depending on the cell size this has to be valid also for the neighbouring cells. The left and right border cells of a door (z_l,z_r) are required to have a cell with the *isWall*-attribute set in their neighbourship. If the border cells lie within a given bound and the *isWall*-attribute is set at different sides, then a door is found.

Every time the systems chagnes into the *FindNewRoom*-mode, the extraction of doors has to be carried out. A new door is stored in a set of doors, if it is not already registered in this set. It is supposed that neigbouring doors have a distance of at least one door width. If a new door is found, it is stored with the posistion, orientation and a flag which shows that the door has not been used yet.

IV. Experimental Results

Different experiments were carried out in the laboratory environment and in for the robot unknown environments. The expreriments were carried out with the $Rosete$-platform, which is shown in figure 11. As main computer a empedded pc with 1GHz clock rate is used. The mobile platform uses a pentium III processor with 1,4GHz. Communication between the two system is done via LAN. Each depth image consists of 64800 single points which leads to a angular resolution of $0.5°$. The maximal velocity of the mobile platform is $v = 1.5\frac{m}{s}$.

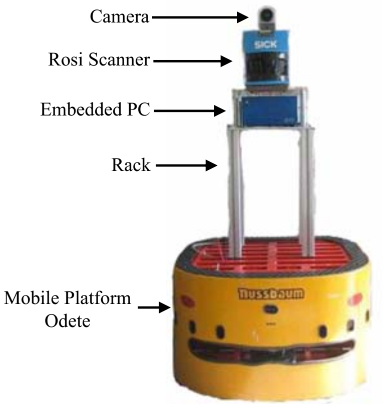


Fig. 11. Exploration system $Rosete$

Among others, a experiment with 3 rooms and one connecting floor was carried out. The exploration started in one of the rooms. After the completion of the first room with five scans a door to the floor was detected and approached. After the **FastRoomScan**-mode a floor was detected an the quality function adapted. After completion of the floor with 10 additional scans the second room was approached and scanned. The last room was approached via the floor and explored. After no open doors were left the system terminated. The exploration lasted about 33 minutes with overall 36 scans. Figure 12 shows the incremental construction of the 3D-model and the 2D-grid. Figure 13 shows the resulting 3D-model with overlapping depth images.

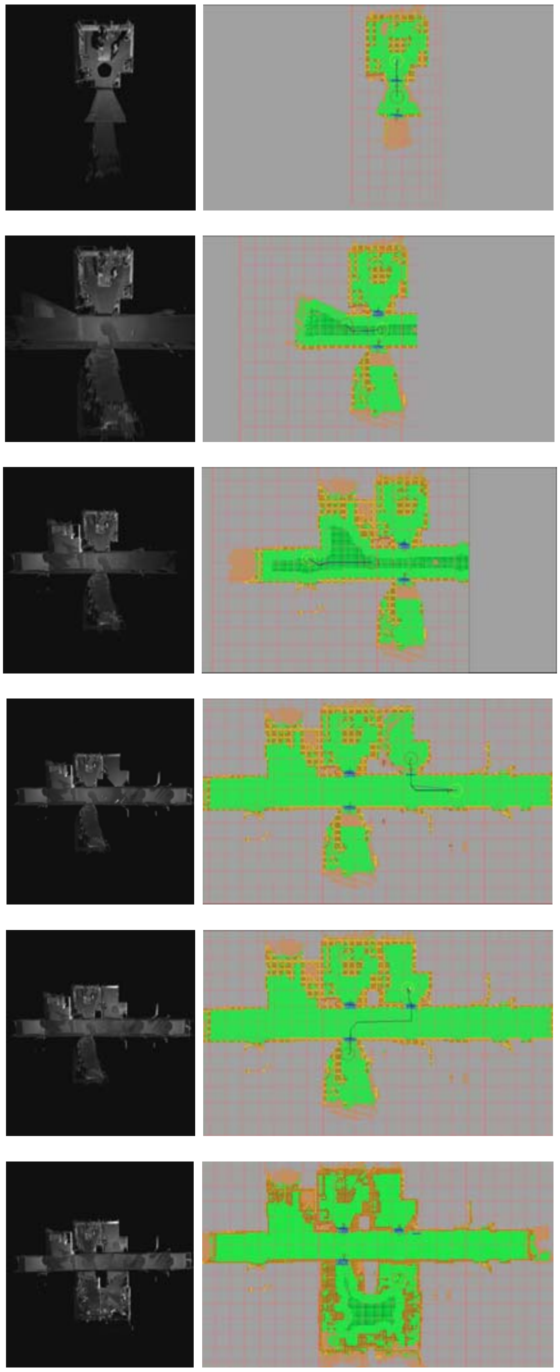


Fig. 12. Incremental Exploration of three rooms and a floor with 3D-model (left) and 2D-grid (right)

V. Conclusion

The article presented a next-best-view planner for 3D environments. The developed 3D-sensor as well as the processing and integration of the data was explained. The next best views are planned by means of a reduced 2D-grid which conserves navigation-important 3D information. A scanning strategy for indoor environments was presented. In an experiment the application and the functionality of the system has been shown.

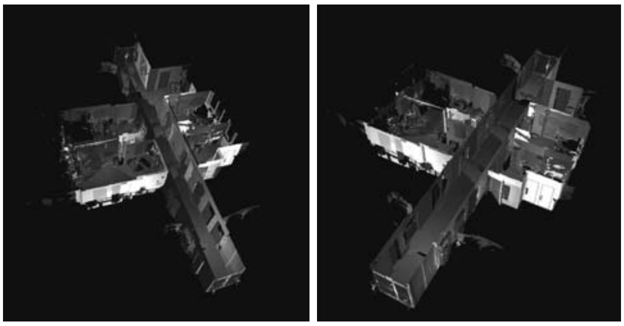


Fig. 13. 3D-model of three rooms and a floor from above

Further work will focus on the integration of inclinometers for better registration results on slopes with known initial values for the roll- and pitch angles. Furthermore the planning strategy will be extented to more than two different types of rooms so that the system is also able to explore more complex environments.

References

[1] C. Früh, ”Automated 3D Model Generation for Urban Environments”, 2002, Universität Karlsruhe, phd thesis

[2] S. Thrun, D. Hhnel, W. Burgard. ”Learning compact 3D models of indoor and outdoor environments with a mobile robot”, In IJCAI, 2001.

[3] W. Burgard et. al. ”A real-time algorithm for mobile robot mapping with applications to multi-robot and 3D mapping”. In IEEE International Conference on Robotics and Automation, San Francisco, 2000.

[4] Oliver Wulf, Bernardo Wagner, Mohamed Khalaf-Allah, ”Using 3D data for Monte Carlo localization in complex indoor environments”, European Conference on mobile Robotics, ECMR, 2005, Ancona

[5] H. Andreasson, R. Triebel and W. Burgard, ”Improving Plane Extraction from 3D Data by Fusing Laser Data and Vision”, In Proc. of the International Conference on Intelligent Robots and Systems(IROS), 2005

[6] P. J. Besl, N. D. McKay, ”A Method for Registration of 3-D Shapes”, IEEE transactions on pattern analysis and machine intelligence, vol. 14, no. 2, February 1992

[7] A. Nuechter et. al., ”Heuristic-Based Laser Scan Matching for Outdoor 6D SLAM”, University of Osnabrueck, Institute of Computer Science, Knowledge Based Systems Research Group, Osnabrueck, Germany, 2005

[8] Peter Steinhaus, ”Navigation mobiler Systeme in dynamischen Umgebungen auf der Basis verteilter Sensoren”,Phd-Thesis,Universitt Karlsruhe,2004

[9] Marcus Strand, Frank Erb, Rüdiger Dillmann, ”Range Image Registration Using an Octree based Matching Strategy”, International Conference on Mechatronics and Automation (ICMA ’07), 2007, Harbin, China

[10] Connolly, C. J., ”The determination of next best views”,Proceedings of the International Conference on Robotics and Automation, 1985

[11] Joseph O’Rourke, ”Art Gallery Theorems and Algorithms”, Oxford University Press, 1987

[12] Cyrill Stachniss, ”Exploration and Mapping with mobile Robots”, Universitt Freiburg, 2006