Semantic Scene Analysis of Scanned 3D Indoor Environments

Andreas Nüchter, Hartmut Surmann,
Kai Lingemann, Joachim Hertzberg

Fraunhofer Institute for
Autonomous Intelligent Systems (AIS)
Schloss Birlinghoven
53754 Sankt Augustin
Germany

Overview

1. Introduction
2. The AIS 3D Laser Range Finder
3. Autonomous Mobile Robots
4. Range Image Registration and Feature Detection
5. Semantic Scene Interpretation
6. Model Refinement
7. Conclusion and Future Work

Introduction

Precise digital 3D models of indoor environments are needed in

- facility management
- architecture
- construction or maintenance of tunnels and mines
- rescue and inspection robotics
- . . .

The increasing need for rapid characterization and quantification of complex environments has created challenges for data analysis.

Autonomous mobile robots equipped with a 3D laser range finder are well suited for gaging the 3D data.

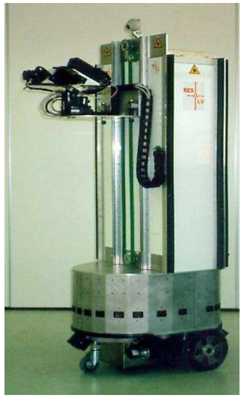
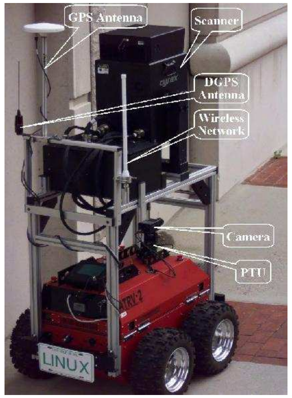
Introduction — State of the Art



In robotics digitalization and mapping of environments is done in 2D or with multiple 2D laser scanners.

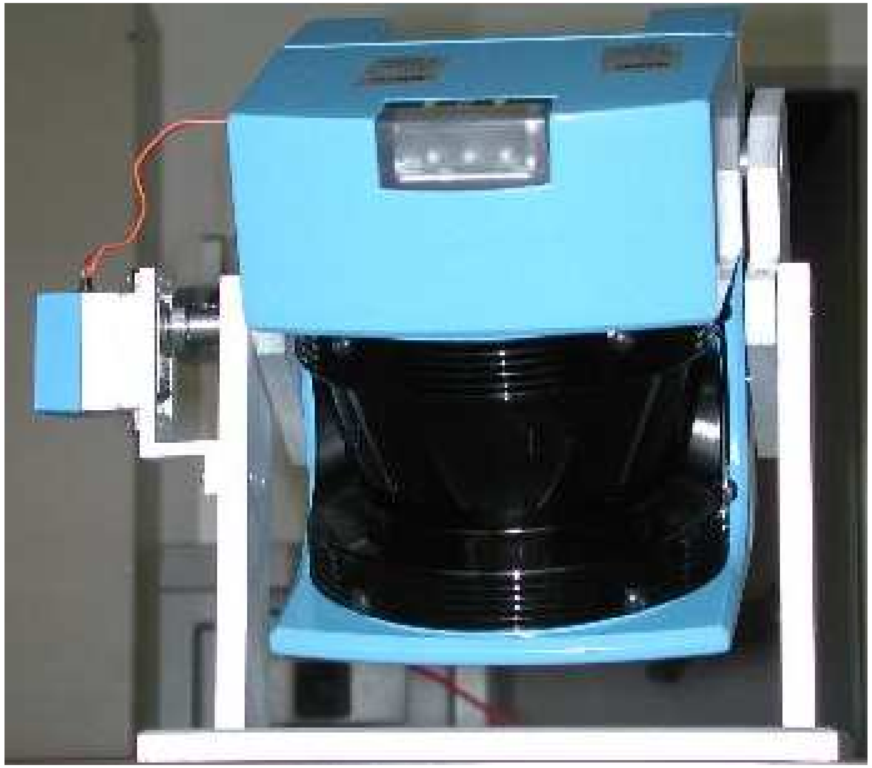
3D laser scanners are an emerging technology to gage objects and environments.



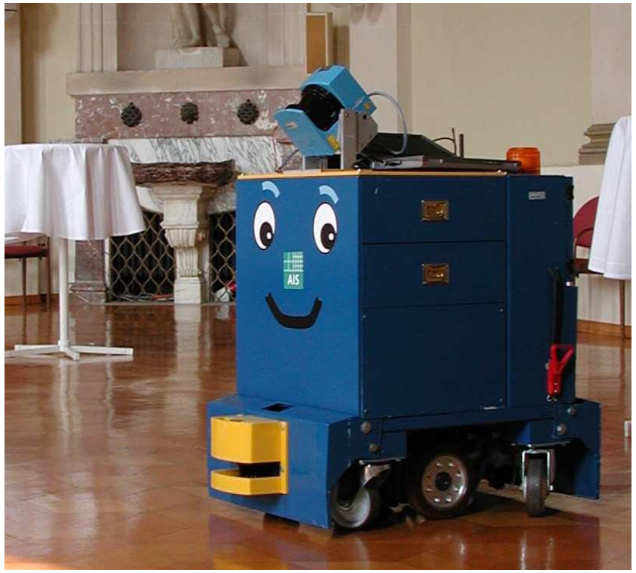
A few groups aim to gage environments with 3D scanners **and** autonomous mobile robots.

The AIS 3D Laser Range Finder



- A standard 2D laser range finder is extended by a mount and a small servo motor.
- Up to 302820 range and reflectance points per 3D scan,
  apex angle $180^\circ \times 120^\circ$,
  accuracy $\approx$1cm.
- The low cost 3D scanner is commercially available for 8500 Euro.

The Autonomous Mobile Robots



The robot Ariadne is an industrial platform (weight: 250kg, payload 200kg). Usually it navigates with a 2D laser range finder. 2D-Scan Video

Result: Equipped with a 3D scanner the robot navigates safely. Video Animation

The robot Kurt3D is a lightweight (weight: $\sim$22.5 kg). It's the fastest (reliably controlled) robot of the world. Video

Overview

Range Image Registration

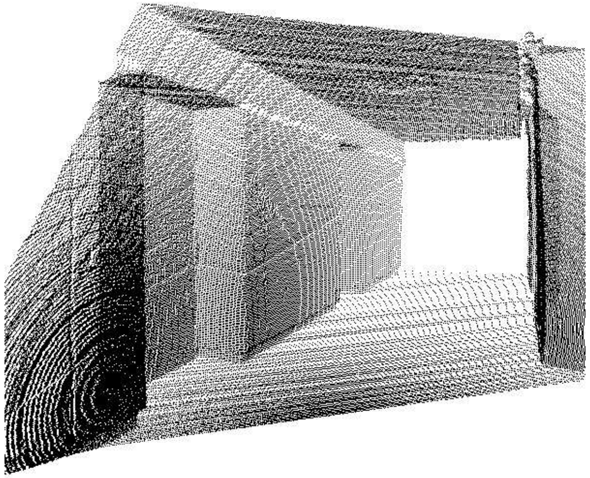
- Iterative Closest Points (ICP) algorithm for rough alignment of 3D scans:
  1. select the closest points as correspondences $w_{i,j}$, and
  2. find the transformation (rotation $\mathbf{R}$, translation $\mathbf{t}$), that minimizes

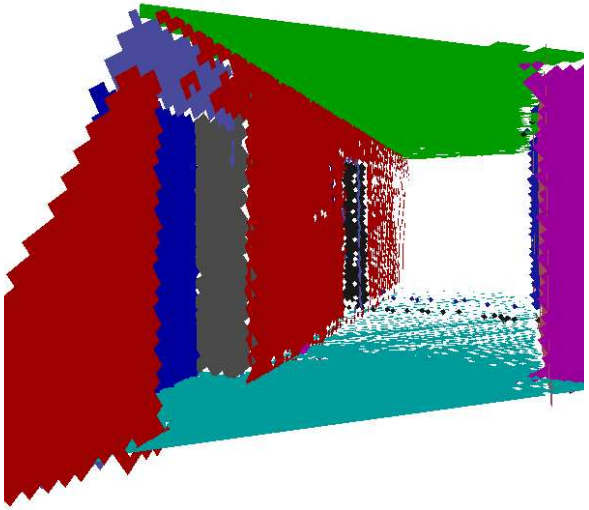
$$E(\mathbf{R}, \mathbf{t}) = \sum_{i=1}^{N_m} \sum_{j=1}^{N_d} w_{i,j} \, ||\mathbf{m}_i - (\mathbf{R}\mathbf{d}_j + \mathbf{t})||^2 \, .$$

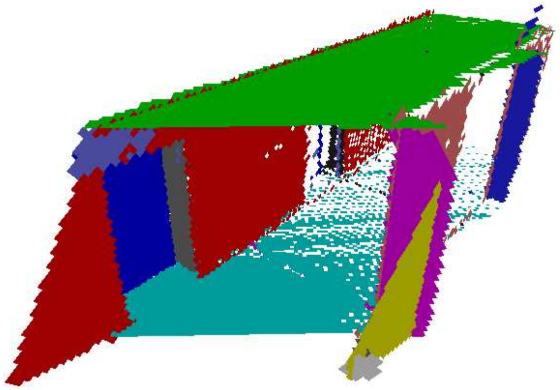
- ICP searches for local minima.
- As starting guess the odometry based pose estimate is used. Animation
- An extension to multiple 3D scans is required to build an overall consistent map of the environment (*simultaneous localization and mapping problem – SLAM*).

Feature Extraction

- Extraction of 3D planes from registered unmeshed range data.
- Mixture of the RANSAC and the ICP algorithm.
  - Randomly select a point and estimate a plane through two neighbored data points.
  - Mark all points that are in an ε-area of the estimated plane.
  - Start the ICP with a point-to-plane metric and adjust the estimated plane.



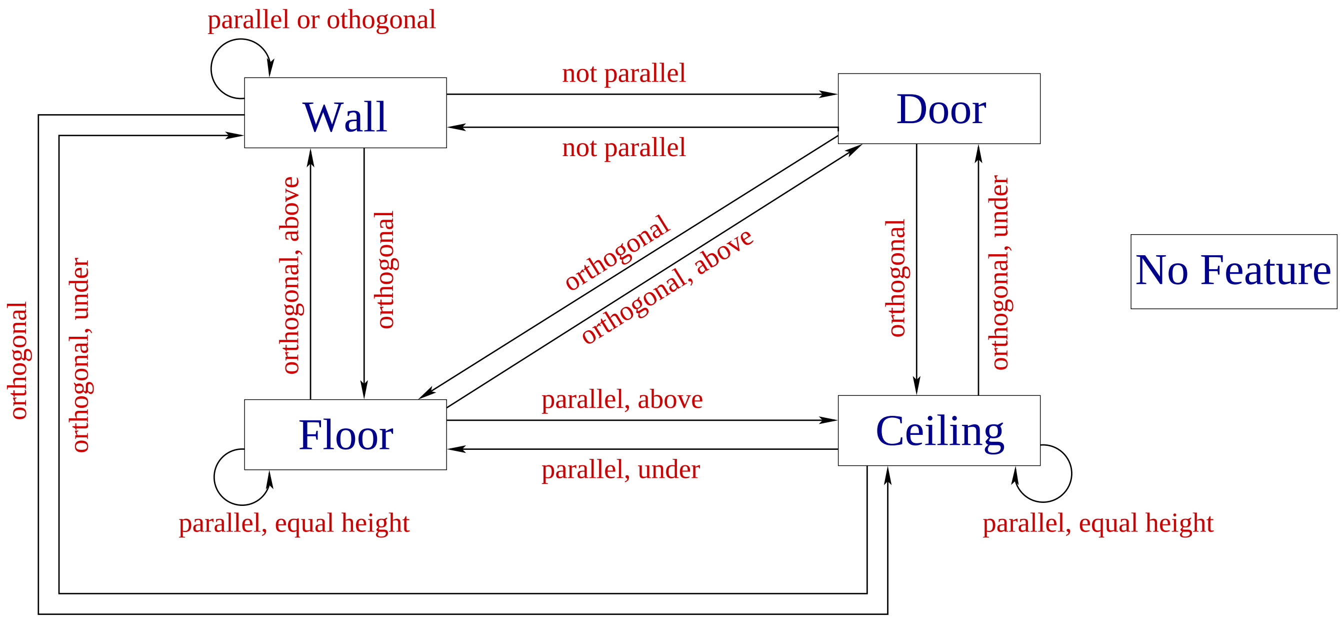




Semantic Scene Interpretation

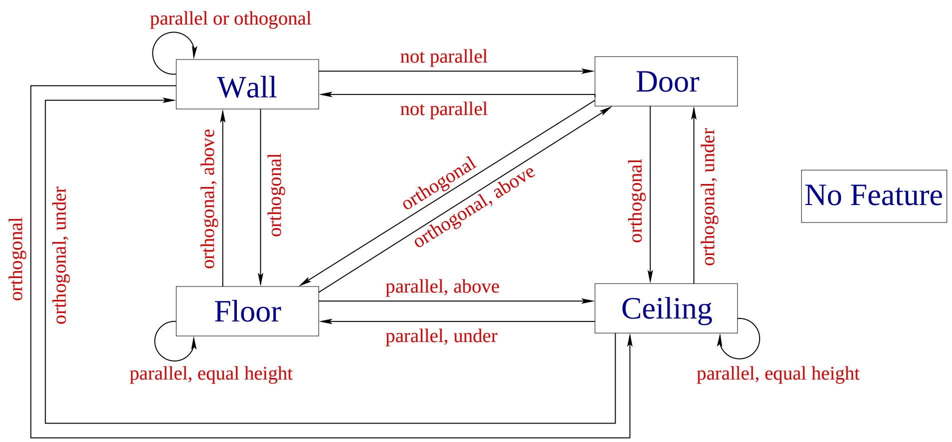
The background for interpretation comprises generic architectural knowledge. A model of an indoor scene is implemented as a semantic net based in analogy to Cantzler et al. [DAGM 2002].

- Nodes of a semantic net represent entities of the world / model, i.e., $L = \{\texttt{Wall}, \texttt{Floor}, \texttt{Ceiling}, \texttt{Door}, \texttt{NoFeature}\}$.
- The relationship between the entities is encoded using different connections, i.e., $R = \{\texttt{parallel}, \texttt{orthogonal}, \texttt{above}, \texttt{under}, \texttt{equal height}\}$.



A depth first search (backtracking) is implemented to assign the labels to the set of planes P according to the constraints in the semantic net.

Externalized Knowledge Representation in Prolog




```
parallel(floor,floor).
parallel(ceiling,floor).
parallel(ceiling,ceiling).
parallel(floor,ceiling).
parallel(wall,wall).
parallel(X,_) :- X == nofeature.
parallel(_,X) :- X == nofeature.
orthogonal(ceiling,door).
orthogonal(ceiling,wall).
...
equalheight(floor,floor).
equalheight(ceiling,ceiling).
equalheight(door,_).
...
```

Externalized knowledge representation in Prolog with facts for every arc of the net and a condition for the `NoFeature` label.

- Use Prolog's unification and backtracking algorithm for the labeling task.

Plane labeling with Prolog

In addition to the representation of the semantic net, a clause of the following form is compiled from the analysis of the planes. The planes are represented by variables `P0`, `P1`, etc.:

```
labeling(P0,P1,P2,P3,P4) :- parallel(P0,P1), under(P0,P1),
                            orthogonal(P0,P2), under(P0,P2),
                            orthogonal(P0,P3), under(P0,P3),
                            parallel(P0,P4), above(P0,P4), ···
```

To start the unifi cation execute:

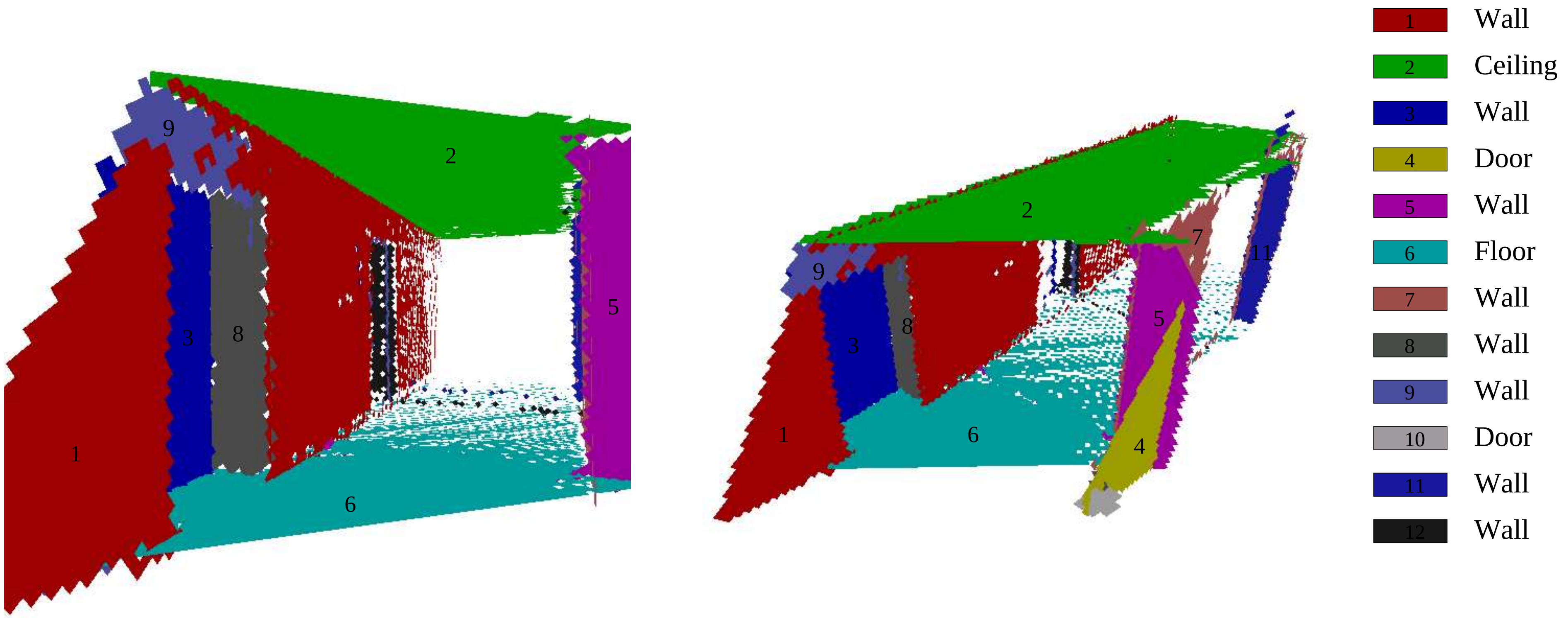
```
consistent_labeling(P0,P1,P2,P3,P4) :- labeling(P0,P1,P2,P3,P4).
```

The label `nofeature` is considered, iff the unifi cation fails. Compute combinations to unify the variable:

```
consistent_labeling(P0,P1,P2,P3,P4) :- combination([P0,P1,P2,P3,P4],[nofeature]),
                                       labeling(P0,P1,P2,P3,P4).
```

The process is continued with assigning two variables the label to `nofeature`, and so on until a Prolog’s unification succeeds.

Semantic Scene Interpretation — Results



The 3D planes get labels.

➡ Use this derived interpretation for automatic model refinement.

3D Model Refinement

The 3D planes are adjusted such that the planes explain the 3D data and the semantic constraints like parallelism or orthogonality are enforced.

An error function is built to enforce the parallelism or orthogonality constraints.

$$E(P) = \sum_{p \in P} \sum_{\mathbf{x} \in p} ||(\mathbf{x} - \mathbf{p}) \cdot \mathbf{n}|| + \gamma \sum_{p_i \in P} \sum_{p_j \in P} c_{i,j}, \tag{1}$$

where $c_{i,j}$ expresses the parallelism (2) or orthogonality (3) constraints according to

$$c_{i,j} = \min\{|\arccos(\mathbf{n}_i \cdot \mathbf{n}_j)|, |\pi - \arccos(\mathbf{n}_i \cdot \mathbf{n}_j)|\} \tag{2}$$

and

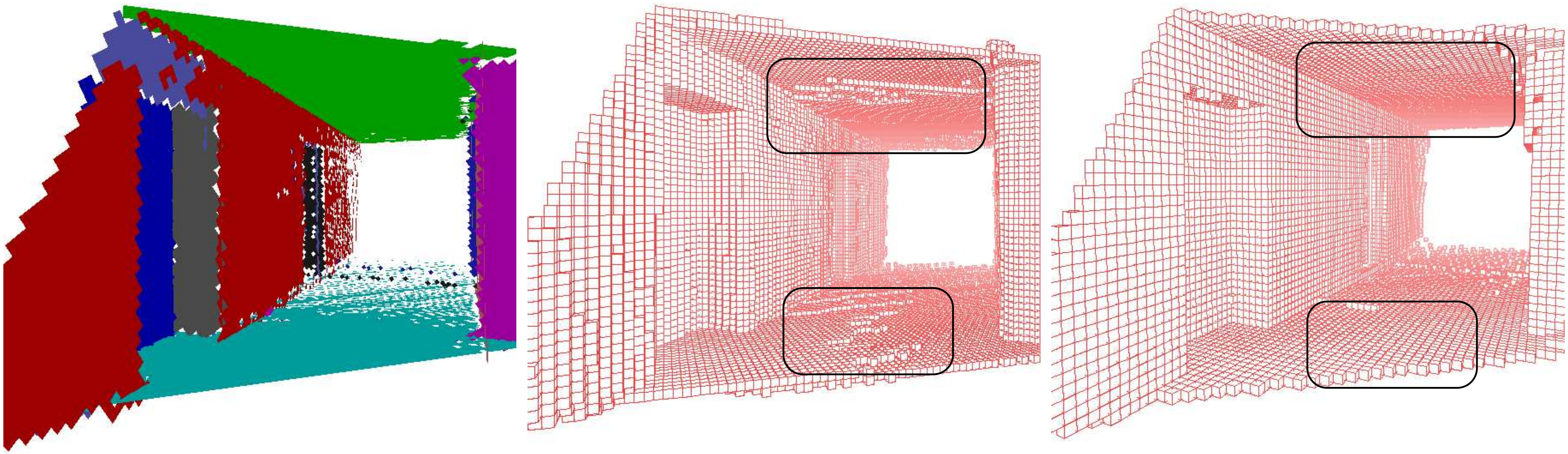
$$c_{i,j} = |\frac{\pi}{2} - \arccos(\mathbf{n}_i \cdot \mathbf{n}_j)|. \tag{3}$$

Minimization of E(P) (eq. 1) is a nonlinear optimization process. For minimization use a heuristic based Powell's method or the downhill simplex method.

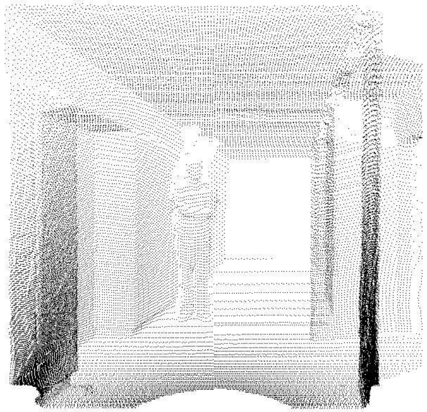
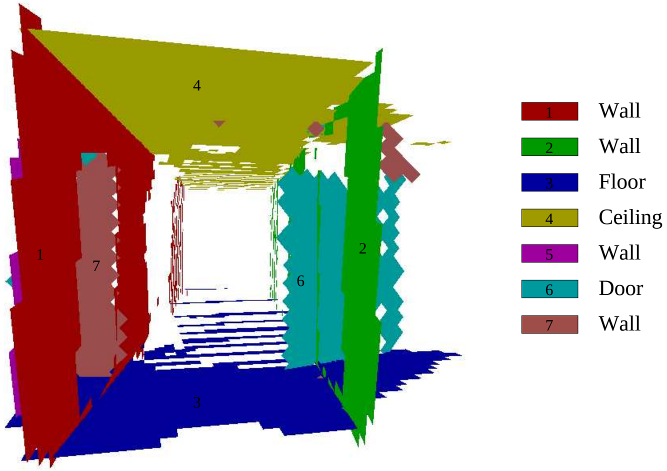
3D Model Refinement — Results

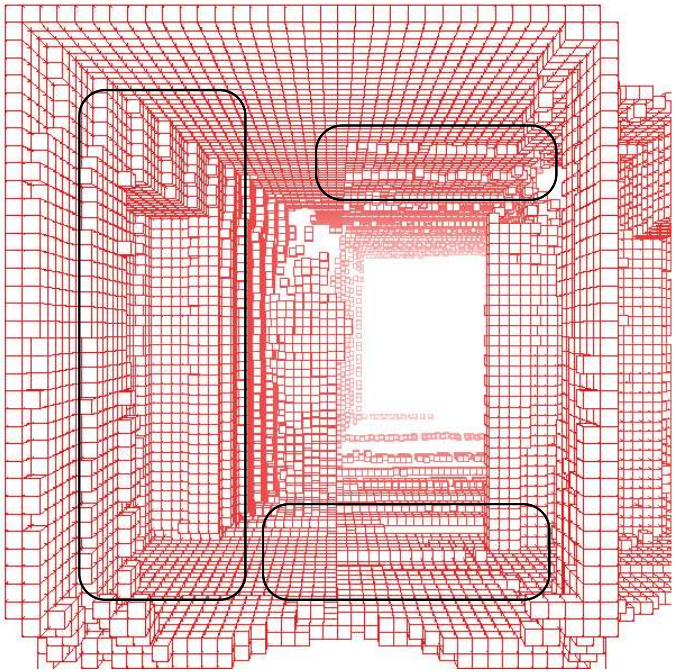
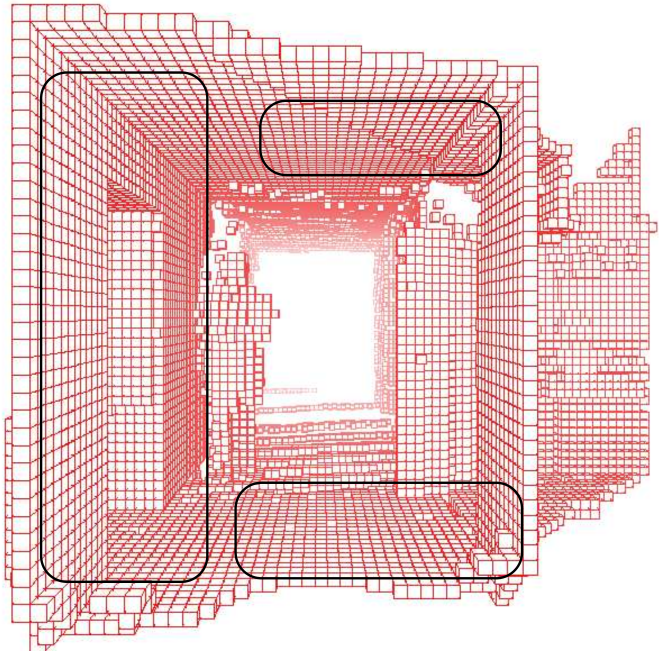
An octree-based algorithm generates the mesh (cube width: 5cm) to visualize the differences between the images.

The semantic description, i.e., the ceiling and walls, enable to transform the orientation of the model along the coordinate axis.



Model Refinement Results (cont.)

Overview

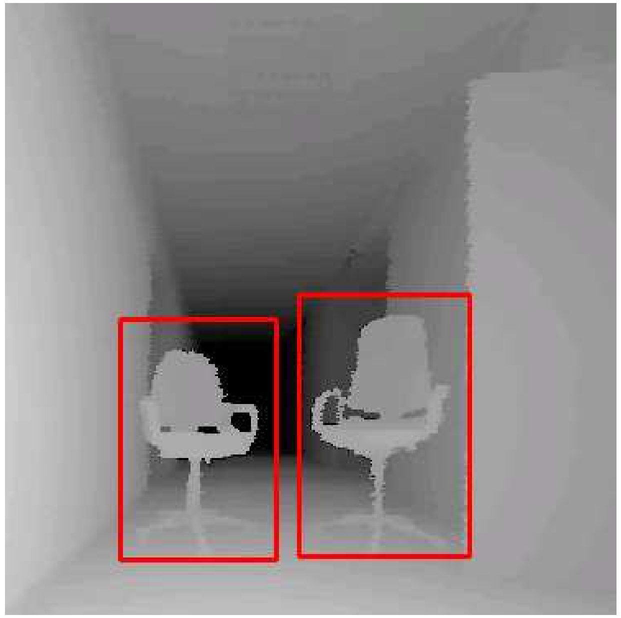
1. Introduction
2. The AIS 3D Laser Range Finder
3. Autonomous Mobile Robots
4. Range Image Registration and Feature Detection
5. Semantic Scene Interpretation
6. Model Refinement
7. Conclusion and Future Work

Conclusions

A new approach to sensor and knowledge based reconstruction of 3D indoor environments with autonomous mobile robots has been presented. Four main steps:

1. An autonomous mobile robot and a SLAM algorithm are used to acquire a 3D point cloud.
2. A fast feature extraction, i.e., 3D plane detection is employed. It consists of a combination of the well-known algorithms RANSAC and ICP.
3. The computed planes are labeled with a predefined semantic net. The semantic net contains and implements general knowledge of indoor scenes.
4. The model is refined with the constraints arising from the semantic labeling. A numerical algorithm based on Powell’s method is used for the 3D model improvement.

Current and Future Work

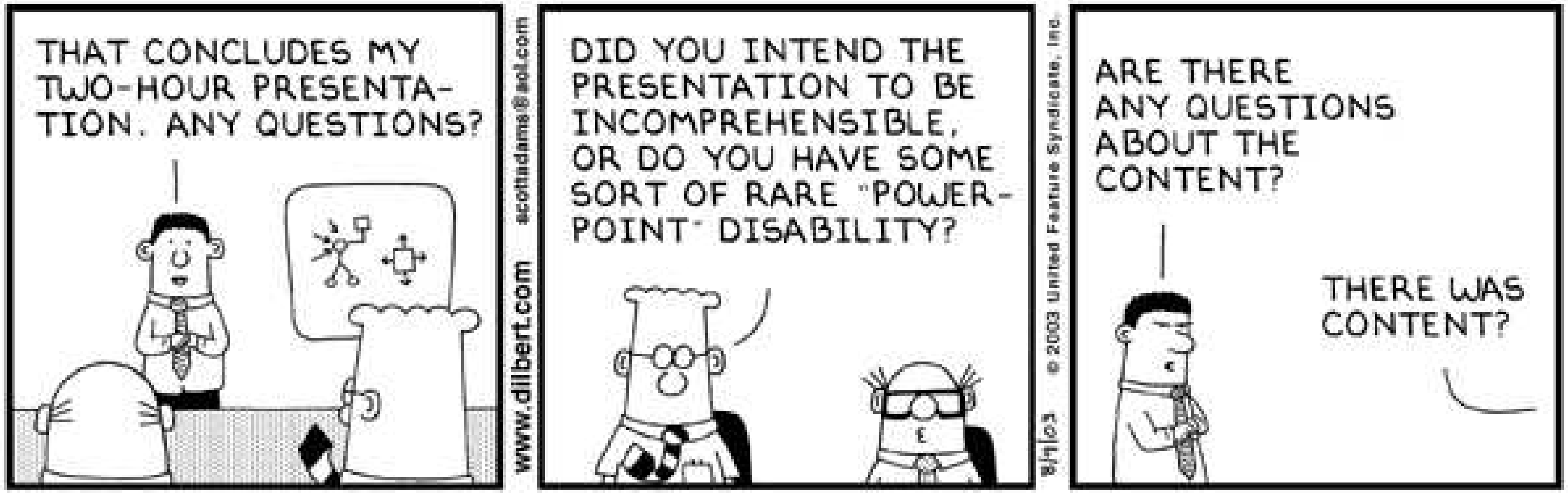


- Autonomous exploration and mapping of sewerage systems.
- Two pan and tilt cameras for texture mapping will be employed to acquire more realistic 3D models.
- 3D object classification and semantic map building.

The End

Questions? Suggestions?

The End



Questions? Suggestions?