

Range Sensor Based Model Construction by Sparse Surface Adjustment

Michael Ruhnke

Rainer Kümmerle

Giorgio Grisetti

Wolfram Burgard

Abstract—In this paper, we propose an approach to construct highly accurate 3D object models from range data. The main advantage of sensor based model acquisition compared to manual CAD model construction is the short time needed per object. The usual drawbacks of sensor based model reconstruction are sensor noise and errors in the sensor positions which typically lead to less accurate models. Our method drastically reduces this problem by applying a physical model of the underlying range sensor and utilizing a graph-based optimization technique. We present our approach and evaluate it on data recorded in different real world environments with an RGBD camera and a laser range scanner. The experimental results demonstrate that our method provides more accurate maps than standard SLAM methods and that it additionally compares favorable over the moving least squares method.

I. INTRODUCTION

Accurate 3D models are envisioned to be essential for the next generation of robotic applications. To accomplish their tasks in the real world, robots require such models to perform navigation, reasoning and manipulation. The problem of learning models of the environment with a mobile robot is known as simultaneous localization and mapping (SLAM). There exists a variety of approaches that can be used to solve the SLAM problem based on 3D range data [8], [5], [6], [4]. Most of the approaches split the SLAM problem into two separate tasks. The first one consists of the estimation of the relative transformation between pairwise observations (often carried out using ICP [2], [10]). In the second step the maximum likelihood configuration of the robot poses is estimated based on these pairwise constraints.

Whereas the modern SLAM techniques produce highly accurate maps, the majority of approaches treats the individual scans as rigid bodies. This leads to an increased uncertainty about the exact position of the surface and also introduces errors in the robot pose estimates. In maps generated by such approaches one commonly observes artifacts such as blurred walls for example.

In this paper we propose a graph-based approach for refining 3D SLAM solutions by considering it as a joint optimization task that simultaneously estimates the robot poses and the surfaces in the environment. Our approach applies a physical model of the underlying range sensor and considers the endpoints of a range measurement as *samples* generated by the surfaces of the environment. We iteratively refine the graph structure by recomputing the data associations between each individual distance measurement

All authors are with the Department of Computer Science, University of Freiburg. G. Grisetti is also with Sapienza, University of Rome. This work has partly been supported by the EC under contract numbers FP7-231888-EUROPA and FP7-248873-RADHAR.



Fig. 1. The left picture shows an accumulated model of a metallic sphere used as input to our method. The right picture shows the result of our method after 20 iterations. As can be seen, the resulting model has a smooth surface. Furthermore our evaluation demonstrates that the model is also very accurate, given a mean error of 1.4 mm.

and the local surface parameters. In this way, we can relax the assumptions that each range measurement is a rigid body so that we obtain more accurate maps.

Figure 1 shows an example illustrating the accuracy that can be achieved with our approach for the reconstruction of a spherical object. The average error in the measured radius between the model and the real object is always below 1.5 mm, while its diameter is 75 mm.

II. SPARSE SURFACE ADJUSTMENT

The goal of our approach is to construct a maximally consistent 3D model of the environment from a set of 3D range measurements of which we approximately know the pose of the sensor. The input of our approach can be the output of a traditional SLAM algorithm. Our approach is able to compensate for small errors in the sensor position and it takes into account the noise affecting the measurement itself. Typical man-made environments consist of regular surfaces and a range reading can be understood as a sample generated by the underlying observed surface. We exploit the regularity assumption by approximating a surface by a set of small locally planar patches, characterized by their normals in direction to the sensor. In the remainder of this paper we will refer to these as *surfels*.

The main idea is to construct an optimization problem that tries to adjust the poses of the sensor and of the surfels to find a maximally consistent configuration. To achieve this task, we assess the surfel structure from each individual measurement endpoint. We then minimize the distance between nearby surfels acquired from different robot poses. This is done by taking into account the uncertainties of the sensor measurements via appropriate sensor models.

A. Surface Model

In our approach, we model the surface of a 3D range observation as a set of surfels. We assume the sensed surface to be piecewise smooth and that we are able to extract local normals around the measurement’s endpoints. We represent

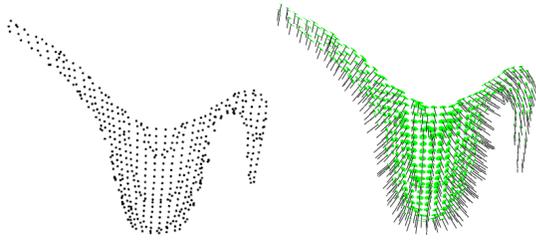


Fig. 2. This figure illustrates our surface model. The left image shows the input point cloud. In the image on the right we show the surface patches extracted in correspondence of the measured points, together with the normal computed from the neighbors.

every single measurement endpoint by a small planar region, which gives a local description of the surface. The local characteristics of a surfel are modeled by Gaussian distributions and are calculated from the neighboring endpoints of the same observation. These Gaussians encode the orientation of the normal and how well the local surface can be represented by a plane. The mean μ_{nk} of a Gaussian represents the center of the surface patch that is measured by the k^{th} measured endpoint \mathbf{r}_k that originates from the n^{th} sensor pose \mathbf{x}_n . Its covariance is given by considering the endpoints within a given region around μ_{nk} within the same measurement.

Once the Gaussian is computed, the estimated normal of the surface $\hat{\mathbf{n}}_k$ is the eigenvector of the smallest eigenvalue of the covariance matrix oriented towards the sensor. The right picture of Figure 2 gives an example of the computed local characteristics.

B. Sensor Model for RGBD Cameras

A detailed description of a laser sensor model can be obtained by an appropriate extension of the model proposed in our previous work [9] to 3D. In the case of an RGBD camera, the depth values are calculated from correspondences between at least two sources. In the case of stereo the two sources are images and in case of an active sensor like the Microsoft Kinect, one source is an infrared camera and the other source is a projector. In both cases the distance for a correspondence can be computed by considering the disparity, the focal length, and the known offset between the two sources. The measured disparity between corresponding points is quantized in sub-pixels, which introduces a systematic error. The depth is inversely proportional to the disparity. Therefore, the impact of the quantization error in the disparity on the depth estimate increases with the range.

The quantization error can be assumed to be uniformly distributed, and the width e_{quant} of the distribution depends on the sensed range, according to the following equation:

$$e_{\text{quant}}(z) = \frac{q_{\text{pix}} \cdot b \cdot f}{2} \cdot \left[\frac{1}{\text{Rnd}\left(\frac{q_{\text{pix}} \cdot b \cdot f}{z} + 0.5\right)} - \frac{1}{\text{Rnd}\left(\frac{q_{\text{pix}} \cdot b \cdot f}{z} - 0.5\right)} \right]. \quad (1)$$

Figure 3 illustrates the quantization error. We evaluated the statistical error of a Kinect facing a planar surface in a range between 0.5m and 3.1m. Every 0.1m we acquired 100 observations and calculated the error to a fitted plane.

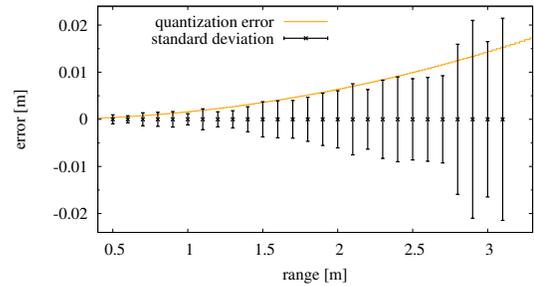


Fig. 3. This figure shows the range dependent theoretical quantization error in orange. The standard deviation calculated from experimental Kinect data is shown in black.

The resulting standard deviation is shown in Figure 3. The standard deviations behave similar to the analytic error function.

To construct the least squares problem we approximate the uniform distribution along the direction of an endpoint with a Gaussian, whose covariance is proportional to the variance of the uniform distribution of width e_{quant} . To be more detailed, we represent the uncertainty of a measured endpoint with a covariance Σ^{meas} . The covariance σ_{11} along the direction of k^{th} endpoint $\|\mathbf{r}_k\|$ is modeled as

$$\sigma_{11} = k_{11} \cdot (e_{\text{quant}}(\|\mathbf{r}_k\|))^2, \quad (2)$$

where k_{11} is a weighting factor.

We assume a range dependent error between neighbors in X-direction and Y-direction of the image, modeled in σ_{22}, σ_{33} . This error depends on the size of the surface “covered” by a pixel. Given the angular resolutions β_X, β_Y we can model the error as

$$\sigma_{22} = k_{22} \cdot \tan\left(\frac{\beta_X}{2}\right) \cdot \|\mathbf{r}_k\| \quad (3)$$

$$\sigma_{33} = k_{33} \cdot \tan\left(\frac{\beta_Y}{2}\right) \cdot \|\mathbf{r}_k\|. \quad (4)$$

Here, k_{11}, k_{22} , and k_{33} are constant and sensor dependent factors and σ_{11}, σ_{22} , and σ_{33} are the respective entries in the covariance matrix Σ^{meas} .

C. Surface Correspondences

In the previous sections we described a model for the individual endpoints of a single range measurement by assuming that the sensed surface is locally regular. In this section, we will explain how we determine potential correspondences between regions of surfaces sensed from different sensor positions.

To find correspondences in a relatively large distance we utilize the “normal-shooting” heuristic proposed by Chen *et al.* [3]. Given an initial configuration of a patch $\langle \mu_{ni}, \Sigma_{ni} \rangle$, whose normal is well defined, we search along its normal direction to seek for the closest patch belonging to a *different* observation. Let this patch be $\langle \mu_{mj}, \Sigma_{mj} \rangle$. If the normals of the two patches have a similar orientation, we add a constraint between them, since it is very likely that they belong to the same surface. If the normals differ more than a given threshold (20 degrees in our current implementation), we assume that we hit another surface and we reject the match.

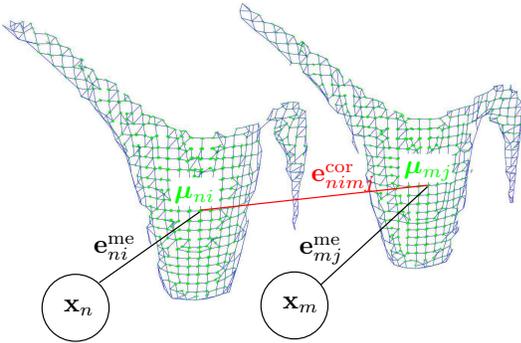


Fig. 4. This figure illustrates the graph structure of our optimization problem. \mathbf{x}_n and \mathbf{x}_m are the sensor poses from which two 3D measurements have been acquired. Two corresponding surface patches μ_{ni} and μ_{mj} , extracted from the different measurements are connected by a “virtual” measurement e_{nim}^{cor} . A measurement generated by a patch depends on the patch characteristics and on the position of the sensor. In this example a beam’s measurement of the patch μ_{ni} sensed from \mathbf{x}_i is captured by the error function e_{ni}^{me} . Similarly, a measurement of μ_{mj} sensed from \mathbf{x}_j is captured by the error function e_{mj}^{me} .

D. Least Squares Optimization

We formulate a least squares minimization problem to find the configuration of sensor poses $\mathbf{x}_{1:n}^*$ and surface patches \mathcal{M}^* that minimize the following function:

$$\langle \mathbf{x}_{1:n}^*, \mathcal{M}^* \rangle = \operatorname{argmin}_{\mathbf{x}_{1:n}, \mathcal{M}} \sum_{\langle n, m, i, j \rangle} e_{nim}^{\text{cor}} + \sum_{\langle n, k \rangle} e_{nk}^{\text{me}}. \quad (5)$$

Relevant 3D datasets typically have millions of measurement and robot poses, which results in an optimization problem with millions of variables. Nonetheless, the obtained optimizations problems are usually relatively sparse. Therefore, we can efficiently compute this minimum by utilizing the g^2o framework [7], which applies sparse linear algebra libraries.

Since we do not know the true correspondences, we have to iteratively refine the correspondences after every optimization run. Additionally, we recompute the surface patch properties for the updated system and construct a new optimization problem. We perform this procedure until the maximum number of iterations has been reached.

III. EXPERIMENTS

In this section, we present experiments on real world data to evaluate the performance of our approach and to discuss its advantages. The main purpose of our approach is to improve the consistency of models. We evaluate this consistency visually and furthermore use the entropy to obtain quantitative results. To compute the entropy, we project the range measurements into a 3D grid, where we calculate for each cell the number of times a beam intercepts a cell without ending in it and the number of times a beam ends in this cell. Based on these two quantities we can calculate the probability that a beam is reflected by the corresponding area in the environment. The entropy for the map is then computed as the sum of all entropies of the individual cells.

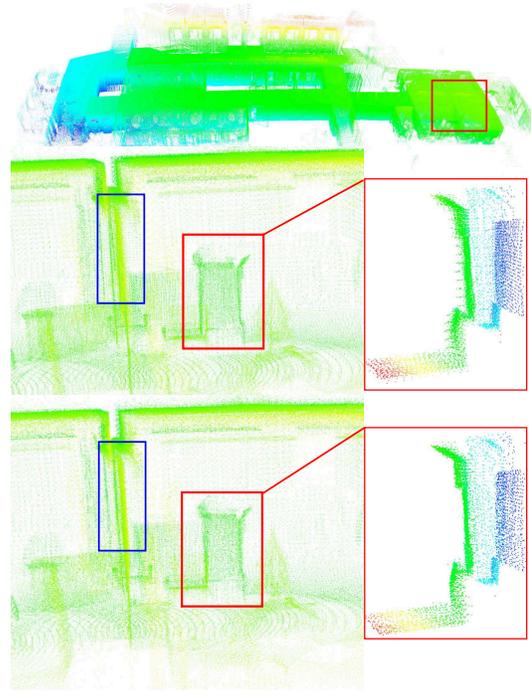


Fig. 5. The top picture shows an overview of the AASS Building dataset. The middle picture shows particulars of the dataset before optimization and the bottom picture shows the same datasets after optimization. We highlighted with rectangles regions where the effect of the optimization is particularly evident.

A. Environment Models

In our first experiment we evaluated our method on the publicly available AASS-loop¹ [1] real world dataset. We applied our method on the given SLAM solution for this dataset and computed the entropy on a 3D grid with cell size 5 cm. Figure 5 shows an overview of the dataset (top), a detailed view of one example region before optimization (middle) and the same region after the optimization procedure (bottom). Table I gives the corresponding statistics.

Our approach substantially reduced the entropy by 7.6% on the AASS dataset and was able to improve the local consistency of the resulting model. Clearly, the better the initial SLAM solution is, the more accurate the result of our model will be, since there will be less ambiguities in the data association. Finally, the denser the dataset the better our approach can perform the alignment.

B. Object Models

In our second experiment we demonstrate that our approach can also efficiently build consistent object models up to a resolution of 1 mm. Therefore, we applied our method on an object model dataset, acquired with a Kinect on a rotating turntable. For the initial registration of the point clouds we applied incremental scan matching. Afterwards we applied our method and computed the entropy before and after optimization with a grid size of 2 mm.

Figure 6 shows the input data (b) and the resulting model for a black mug (d), computed on point clouds with a resolution of 1 mm. In both cases the resulting models look

¹Courtesy of Martin Magnusson, AASS, Örebro University, Sweden

TABLE I
DETAILED OVERVIEW OF THE DATASETS USED IN OUR EXPERIMENTS.

Dataset	Figure	# Scans	# Points	comp. time	grid resolution	entropy input	entropy optimized	entropy reduction
AASS Building	5	60	2,266,519	48 min	50 mm	287,433	265,585	7.6%
Black Cup	6	51	107,281	6 min	2 mm	8,184.52	6,781.88	17.1%

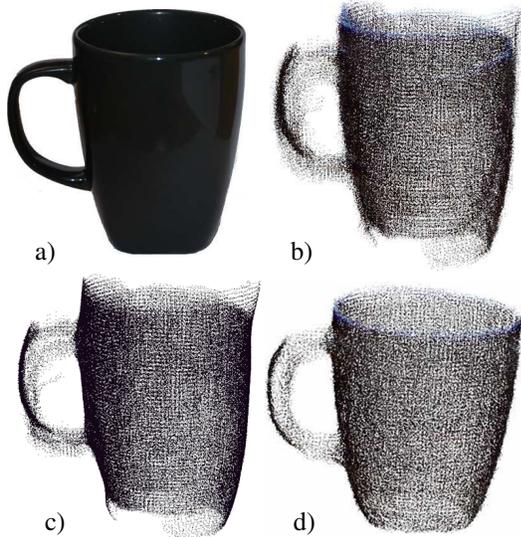


Fig. 6. This figure illustrates the differences between our method and MLS. We acquired multiple measurements of a cup, shown in the image a). In b) we show the point clouds registered by using ICP. In c) we show the effect of MLS on the input data. Whereas the model looks smoother than the input it shows inconsistencies arising from errors in the estimate of the initial positions of the sensor. In d) we show the results obtained by our approach on the same input data. The result appears to be more consistent.

more consistent. The entropy of the black mug model was reduced by 17.1%. Additionally, we compared the outcome of our method with the results of MLS. Figure 6(c) shows the resulting point cloud for MLS applied on the input point cloud. Since MLS does not consider errors in robot poses, the resulting model is not consistent.

IV. APPLICATIONS

In this section we will briefly discuss possible applications of our method. The described technique is able to reduce inconsistencies and the overall uncertainty in maps, which improves the localization accuracy and therefore the overall performance of a map based navigation system. The market for SLAM based navigation systems starts to evolve with the next generation of vacuum-cleaner and lawn-mower robots. The proposed method makes it possible to either improve the performance of the navigation system or utilize cheaper sensors at a comparable performance level.

In the context of mobile robots with manipulation abilities, our method is one key component for autonomous object model acquisition. Such models are essential for object detection, grasping and manipulation tasks. Given the amount of different objects a robot might face, it seems infeasible to construct object models manually or in an assisted fashion. Right now this scenario is more relevant in a research context since robots with such skills are still years away from being market ready.

But there are already possible application scenarios for

model reconstruction techniques. Video games and 3D movies utilize virtual models of objects and also entire environments. In most cases models are manually constructed right now and it seems reasonable to perform this task in an autonomous or assisted fashion. Given one or two RGBD cameras, a simple turntable and our method it is possible to virtualize real world objects with a high accuracy and equipment costs below 200\$.

V. CONCLUSIONS

In this paper we presented a novel approach to reconstruct environments or object models. Our method is able to improve the consistency of 3D models by efficiently solving a least squares optimization problem that is constructed based on an accurate model of the sensor. Experimental results obtained with an RGBD camera and a laser range finder in real-world settings demonstrate that our approach yields substantial improvements compared to state-of-the-art methods.

REFERENCES

- [1] "Aass-loop dataset," <http://kos.informatik.uni-osnabrueck.de/3Dscans>.
- [2] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.
- [3] Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 1991.
- [4] U. Frese, P. Larsson, and T. Duckett, "A multilevel relaxation algorithm for simultaneous localisation and mapping," *IEEE Transactions on Robotics*, vol. 21, no. 2, pp. 1–12, 2005.
- [5] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments," in *Proc. of the Int. Symposium on Experimental Robotics (ISER)*, 2010.
- [6] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Trans. on Robotics*, vol. 24, no. 6, pp. 1365–1378, Dec 2008.
- [7] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimization," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2011.
- [8] A. Nüchter, K. Lingemann, J. Hertzberg, and H. Surmann, "6d SLAM with approximate data association," in *Proc. of the 12th Int. Conference on Advanced Robotics (ICAR)*, 2005, pp. 242–249.
- [9] M. Ruhnke, R. Kümmerle, G. Grisetti, and W. Burgard, "Highly accurate maximum likelihood laser mapping by jointly optimizing laser points and robot poses," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2011.
- [10] A. V. Segal, D. Haehnel, and S. Thrun, "Generalized-ICP," in *Proc. of Robotics: Science and Systems (RSS)*, 2009.