

Highly Accurate 3D Surface Models by Sparse Surface Adjustment

Michael Ruhnke

Rainer Kümmerle

Giorgio Grisetti

Wolfram Burgard

Abstract—In this paper, we propose an approach to obtain highly accurate 3D models from range data. The key idea of our method is to jointly optimize the poses of the sensor and the positions of the surface points measured with a range scanning device. Our approach applies a physical model of the underlying range sensor. To solve the optimization task it employs a state-of-the-art graph-based optimizer and iteratively refines the structure of the error function by recomputing the data associations after each optimization. We present our approach and evaluate it on data recorded in different real world environments with a RGBD camera and a laser range scanner. The experimental results demonstrate that our method is able to substantially improve the accuracy of SLAM results and that it 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 [16], [11], [10], [7], [12], [9], [17]. 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 [4], [20]). 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 scan as *samples* generated by the surfaces of the environment. We iteratively refine the graph structure by recomputing the data association between each individual distance measurement and the local



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 highly accurate, given a mean error of 1.4 mm.

surface points. In this way, we can relax the assumptions that each range scan is a rigid body so that we obtain more accurate maps. Compared to our previous work [18] we present a novel sensor model (the original one was designed for 2D and lasers only) for 3D and RGBD sensors like the Microsoft Kinect. Additionally, we extended the accuracy and robustness with respect to errors in the data association by combining two strategies with different search ranges. Our approach is a post-processing procedure to be applied to the output of graph-based SLAM systems and is designed to provide locally more consistent models.

We evaluate the accuracy of the models computed by our method by means of ground truth measurements. We furthermore characterize the error of the Microsoft Kinect both theoretically and experimentally. In particular RGBD cameras are affected by a quantization error in the disparity measurements. This error leads to noise in the range data that grows with the distance. By specifically modeling this error, our approach is able to use long range and noisy measurements, to capture wide regions, and to assess the structure of the scene. At the same time it can refine the local structures given the more accurate short range measurements.

Figure 1 shows an example illustrating the accuracy that can be achieved with our approach for the reconstruction of a spherical object with 75 mm radius. The average error in the measured radius is below 1.5 mm.

II. RELATED WORK

The standard technique for the registration of 3D range scan pairs is the Iterative Closest Point Algorithm (ICP) proposed by Besl *et al.* [4]. ICP iteratively minimizes the distance between corresponding points, while updating the data association. Instead of minimizing the point-to-point error Chen *et al.* [6] proposed to minimize the re-projection error based on a point-to-plane metric. This idea has been

extended by Segal *et al.* [20] which minimize the error based on a plane-to-plane metric. Since range sensors observe surface structures this leads to substantially better alignments. These methods focus on the task of accurately registering two scans, but they treat them as rigid bodies.

Chang *et al.* [5] proposed model reconstruction approach for articulated objects, which is able to automatically determine the joints of an object and deform scan parts according to the configuration of the object. This method also estimates the stiffness between surface points and deforms several scan parts to get a consistent model, but does neither consider the sensor uncertainties nor deform each scan point individually.

The Non-Rigid ICP variant proposed by Li *et al.* [15] is able to deform all points on a surface graph (data) onto a second surface graph (model) by introducing a deformation model and applying least square optimization to find the optimum for deformation and registration. Since this method uses a weighted least square approximation for the model surface it is not easily extendable to the case of multiple scans and does not account for systematic sensor errors, while our method does not consider deformations.

Fleishman *et al.* [8] presented Robust Moving Least Squares (MLS), a smoothing technique based on robust regression. They assign the points to piecewise smooth surfaces and are able to obtain accurate models. In a similar context Andersen *et al.* [3] propose a Markov Random Field formulation that optimizes the parallelism between neighboring surface elements (surfels) and their overlap. This smooths out noise while maintaining sharp features. Both methods can be used to obtain accurate surface models, but ignore the model of the sensor that has been used to generate the point cloud and do not optimize over the sensor poses. Compared to our method, MLS is less robust to noisy initial configurations.

In this paper we present an approach to determine the optimal position of both the points in the scans and the robot poses. Our method is similar in spirit to traditional bundle adjustment (BA) problems in computer vision [22]. A BA algorithm seeks to find the configuration of a set of distinguishable features and camera poses that minimizes the re-projection error over the measured sequence of images. In contrast to BA, our method does not require point features and it considers the local structure of the model while performing the optimization. We globally minimize a local plane-to-plane error metric similar to Segal *et al.* [20]. Furthermore, we refine the data associations after every optimization run. This yields models with highly accurate surfaces and drastically reduces the impact of accumulated sensor noise.

III. SPARSE SURFACE ADJUSTMENT

The goal of our approach is to construct a maximally consistent 3D model of the environment from a set of roughly aligned 3D range scans. The input of our approach can be obtained by 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 range

measurements. Typical man-made environments consist of regular surfaces and a range reading can be understood as 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 surfels 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. We extract locally planar surfels from each range scan, as described in Section III-A. Each measured distance is then connected to the corresponding surfel by a constraint that depends on the sensor used (3D laser or RGBD III-B). Each surfel is parametrized by a state variable.

Nearby surfels arising from different scans are then connected by “virtual” measurements that “pull” the two surfels onto each other, while allowing them to slide along the tangential directions. These virtual measurements are iteratively refined, as explained in Section III-C. Finally, in Section III-D we will explain how to construct a sparse optimization problem that incorporates all the above constraints. Despite the high number of variables these problems can be solved in a relatively fast manner by using modern optimization techniques.

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 scan’s endpoints. We represent 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 within a global coordinate frame. 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 surfel that is measured by the k^{th} measured endpoint \mathbf{r}_{nk} that originates from the n^{th} sensor pose \mathbf{x}_n . We initialize μ_{nk} with

$$\mu_{nk} = \mathbf{r}_{nk} \oplus \mathbf{x}_n \quad (1)$$

and compute the covariance based on the endpoints in the local neighborhood of μ_{nk} within the same scan. 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. Since we update the poses of the surfels during the optimization procedure we need to recompute the local characteristics after each optimization run. We can not

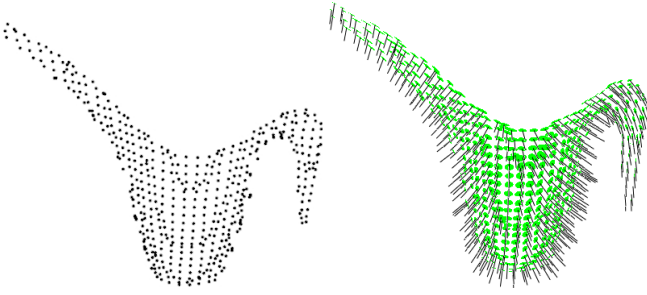


Fig. 2. This figure illustrates our surface model. The left image shows the input point cloud. In the image on the right we show the surfels extracted in correspondence of the measured points, together with the normal computed from the neighbors.

avoid the re-computation by choosing a relative coordinate frame because the surfel optimization changes the relative configuration of the neighborhood.

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 [18] to 3D. In case of RGBD and stereo cameras, the depth value of a point in the scene is obtained from at least two projections of that point measured in the image planes of two different observers. In case of a stereo camera these two observers are in fact two cameras, and the correspondences between the two images are found by a stereo matching algorithm. The Kinect sensor approaches the problem by coupling an infrared projector and an infrared cameras, thus making the correspondences more robust and easier to find. In this case the projector can be seen as an additional observation point. The displacement between two projections of the same point in the two image planes is called disparity and is quantized in sub-pixels, thus suffers of 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.

According to [13] the range z of the Kinect is calculated using the following equation:

$$z = \frac{q_{\text{pix}} \cdot b \cdot f}{d}. \quad (2)$$

Here q_{pix} is the subpixel resolution of the device in the calculation of the disparity, b is the baseline, f is the focal length and d the normalized disparity. In the case of the Kinect $q_{\text{pix}} = 8$. If we transform this equation to get the disparity for a range z , we obtain

$$d = \frac{q_{\text{pix}} \cdot b \cdot f}{z}. \quad (3)$$

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:

$$e_{\text{quant}}(z) = \frac{q_{\text{pix}} \cdot b \cdot f}{2} \cdot \left[\frac{1}{\text{Rnd}\left(\frac{q_{\text{pix}} \cdot b \cdot f}{z} + 0.5\right)} - \frac{1}{\text{Rnd}\left(\frac{q_{\text{pix}} \cdot b \cdot f}{z} - 0.5\right)} \right]. \quad (4)$$

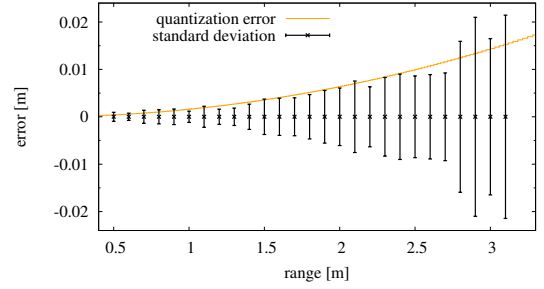


Fig. 3. This figure shows the range dependent theoretical quantization error in orange. The standard deviation calculated from experimental Kinect data is shown in black.

Figure 3 illustrates the quantization error. We evaluated the statistical error of a Kinect facing a planar surface in a range between 0.5 m and 3.1 m. Every 0.1 m we acquired 100 scans and calculated the error to a fitted plane. The resulting standard deviation is shown in Figure 3. The standard deviations behave similar to the analytic error function. Note that the Kinect data have larger error bounds as the quantization error for ranges over 2.5 m. This depends on the plane fitting routine, whose performance decreases in presence of the high noise that occurs in this range leading to additional errors.

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

$$\sigma_{11} = (e_{\text{quant}}(\|\mathbf{r}_k\|))^2. \quad (5)$$

An RGBD sensor observes neighboring measurements at the same time and there is no additional dynamic involved. Therefore, we assume a range dependent error between neighbors in X-direction of the image, modeled in σ_{22} , and the Y-direction of the image, modeled in 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

$$\sigma_{22} = \tan\left(\frac{\beta_X}{2}\right) \cdot \|\mathbf{r}_k\| \quad (6)$$

$$\sigma_{33} = \tan\left(\frac{\beta_Y}{2}\right) \cdot \|\mathbf{r}_k\|. \quad (7)$$

Here, 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 scan 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.

Given a surfel we search for potential correspondences between this surfel and all other surfels observed from different sensor poses within a search radius. Since our method

is able to deform scans, we have to carefully select the search range. A too large range might result in stretched or fragmented surfaces and erroneous sensor poses introduced by false correspondences and a too small range might result in unconnected surfaces, which might lead to an inconsistent model.

Instead of searching a trade off between accuracy and robustness we combine two different strategies. To ensure the local connectivity we apply nearest neighbor (NN) in close range and to be more robust against initial alignment errors from missing or suboptimal loop closures we apply “normal-shooting” (NS) proposed by Chen *et al.* [6] with a larger search range. Given an initial configuration of a surfel $\langle \mu_{ni}, \Sigma_{ni} \rangle$, whose normal is well defined, we search along its normal direction and in the nearby neighborhood to seek for the closest surfel belonging to a *different* observation. Let this surfel be $\langle \mu_{mj}, \Sigma_{mj} \rangle$. If the normals of the two surfels 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. As maximum search distance in the object model experiments we used 2 cm for normal shooting and 5 mm for nearest neighbor. In the environment model experiments we used 15 cm (NS) and 5 cm (NN).

D. Least Squares Optimization

The overall goal of our approach is to determine the model and the configuration of robot poses that are maximally consistent with the observations. In the remainder of this section we propose error functions and error weights for the previously introduced constraints, with the goal to construct a least squares optimization problem.

The model \mathcal{M} is represented as a set of surfels. Each surfel is described by a Gaussian $\langle \mu_{nk}, \Sigma_{nk} \rangle$, where μ_{nk} denotes the k^{th} measurement of the n^{th} robot pose \mathbf{x}_n , as described in Section III-A.

Each surfel $\langle \mu_{nk}, \Sigma_{nk} \rangle$ is connected to the robot pose by a measurement \mathbf{r}_{nk} . This error is distributed according to the covariance Σ^{meas} computed either from the corresponding laser sensor model [18] or Section III-B:

$$e_{nk}^{\text{me}} = ((\mu_{nk} \ominus \mathbf{x}_n) - \mathbf{r}_{nk})^\top (\Sigma_{nk}^{\text{meas}})^{-1} ((\mu_{nk} \ominus \mathbf{x}_n) - \mathbf{r}_{nk}). \quad (8)$$

The error of corresponding surfels ni and mj belonging to *different* scans is modeled with the following error vector:

$$\mathbf{v}_{nimj}^{\text{cor}}(\mu_{ni}, \mu_{mj}) = \mu_{ni} - \mu_{mj} = \Delta \mu_{nimj}. \quad (9)$$

Accordingly, the error is minimal if both surfels are in the same location. We intend to allow surfels to “slide” onto each other along their tangent directions, while we want them to be more constrained along their normals. This can easily be modeled by the sum of the inverses of the covariance matrices of the surfel ellipsoids:

$$\Omega_{nimj}^{\text{cor}} = \Sigma_{ni}^{-1} + \Sigma_{mj}^{-1}. \quad (10)$$

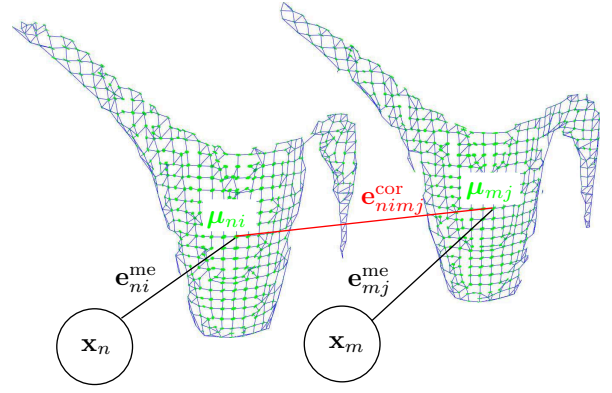


Fig. 4. This figure illustrates the graph structure of our optimization problem. \mathbf{x}_n and \mathbf{x}_m are the sensor poses from which two 3D scans have been acquired. Two corresponding surfels μ_{ni} , and μ_{mj} , extracted from the different scans are connected by a “virtual” measurement e_{nimj}^{cor} . A measurement generated by a surfel depends on the surfel 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} .

The quadratic error function then is

$$e_{nimj}^{\text{cor}} = \Delta \mu_{nimj}^\top \Omega_{nimj}^{\text{cor}} \Delta \mu_{nimj}. \quad (11)$$

Having introduced these quadratic error functions, we can formulate a least squares minimization problem to find the configuration of sensor poses $\mathbf{x}_{1:n}^*$ and surfels \mathcal{M}^* that minimize the following function:

$$\langle \mathbf{x}_{1:n}^*, \mathcal{M}^* \rangle = \underset{\mathbf{x}_{1:n}, \mathcal{M}}{\operatorname{argmin}} \sum_{\langle n,k \rangle} e_{nk}^{\text{od}} + \sum_{\langle n,m,i,j \rangle} e_{nimj}^{\text{cor}} + \sum_{\langle n,k \rangle} e_{nk}^{\text{me}} \quad (12)$$

where e_{nk}^{od} is the error of the odometry measurements between two successive frame acquisitions as described in our previous work [18].

Relevant 3D datasets typically have millions of measurements and robot poses, which results in an optimization problem with millions of variables. Nonetheless, the obtained optimization problems are usually relatively sparse. The objective function is the sum of factors involving only pairs of state variables. Thus, the approximated Hessian contains a number of non-zero entries that is proportional to the number of constraints. The number of introduced constraints inside an observation is linear in the number of measurements. Since the applied sensors have a limited field of view and a limited range, the constraints between observations are in general only local and result in a sparse approximated Hessian. Therefore, we can efficiently compute this minimum by utilizing the g²o framework [14], 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 surfel properties for the updated system and construct a new optimization problem. We perform this procedure until the changes in the objective function fall below a threshold or the maximum number of iterations has been reached.

IV. 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 both visually and quantitatively by measuring the entropy of the generated models. To compute the entropy, we project the range measurements into a 3D grid, where we calculate the reflection probabilities, and the times a cell has been visited. Based on these two quantities we obtain the entropy of the surface models by using the model proposed by Stachniss *et al.* [21].

A. Environment Models

In our first experiment we evaluated our method on three different real world datasets. The first dataset was recorded with a Kinect sensor on a mobile robot in the corridor of building 79 on the Freiburg campus. The second dataset was recorded in the same corridor using a tilting laser range finder. As third dataset we choose the publicly available AASS-loop¹ [1] dataset which was recorded with a spinning laser range finder. We computed SLAM solutions for both corridor datasets and used the given SLAM solution of the AASS dataset as baseline for our consistency comparison. Afterwards we applied our method on all three datasets and computed the entropy on a 3D grid with cell size 5 cm. The first row of Figure 5 shows the used datasets. The second row shows a detailed view of one example region before optimization and the third row shows the same region after the optimization procedure. Table I gives detailed statistics of the involved datasets.

Our approach substantially reduced the entropy in all three datasets. The entropy of the Kinect dataset was reduced from 174,092 down to 143,087, which corresponds to a reduction of 18 %. The entropy of the Building 079 corridor dataset was reduced from 143,569 down to 92,003.6, which is a reduction of 35.9 %. The entropy reduction of 7.6 % on the AASS dataset is lower than on the other datasets, because of the very specific sampling strategy of the rotating laser. The performances of our approach depend on the characteristics of the environment, on the accuracy of the initial solution computed by the SLAM algorithm, and on the characteristics of the sensor. 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 more dense is the dataset, the better our approach can perform the alignment. For these reasons, the magnitude of the entropy reduction differs between the datasets, but it is always positive.

There are two effects, that enable our approach to improve the results of graph-based SLAM algorithms. The optimization of a combined problem, using all available sensor information, instead of optimizing over fixed constraints, introduced by pairwise scan-matching and secondly the possibility to refine observations. Especially in the case of

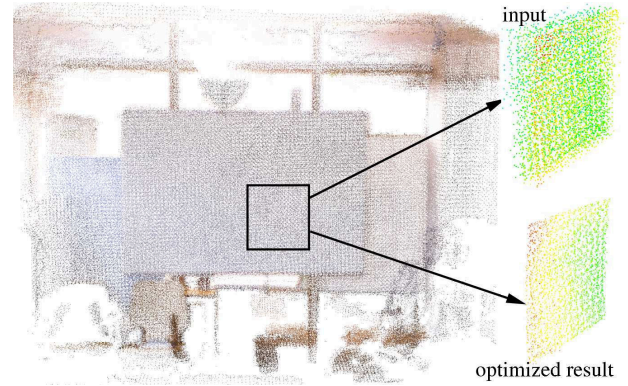


Fig. 6. The left picture shows the point cloud of a model recorded in an office environment, which consists of 20 Kinect scans. The two magnified views on the right show a part of the surface from a different viewpoint. The top right image shows the input data and the bottom right image shows the same region after optimization. The accumulated errors are significantly reduced.

range dependent sensor errors, like the quantization errors of the Kinect, the measurement refinement is essential to build consistent models. As already stated in the introduction, we performed an experiment to evaluate the impact of this effect. To this end, we recorded a set of colored point clouds, while approaching a planar surface. The upper part of Figure 6 shows the complete model and the magnified view of an example region before and after optimization. The accumulated quantization effects of the model are substantially reduced after optimization. In this example, the entropy was reduced by 44 %.

B. Object Models

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

Figure 7 shows the input data (a) and the resulting model for a black cup (b), computed on point clouds with a resolution of 1 mm. The resulting model looks more consistent. The entropy was reduced by 17.1 %. Additionally, we compared the outcome of our method with the results of the MLS implementation provided in the Point Cloud Library (PCL) [19]. Figure 7 (c) shows the resulting point cloud for MLS applied on the input point cloud with default parameters and a search radius of 1 mm. Since MLS does not consider errors in robot poses, the resulting model is not consistent.

C. Ground Truth Data

Obtaining a ground truth for models of arbitrary objects is rather difficult, thus we approached the problem with an object having a simple geometry. We decided to evaluate the accuracy achievable with our method on a geometrically simple object like a sphere with a known radius of 7.5 cm

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

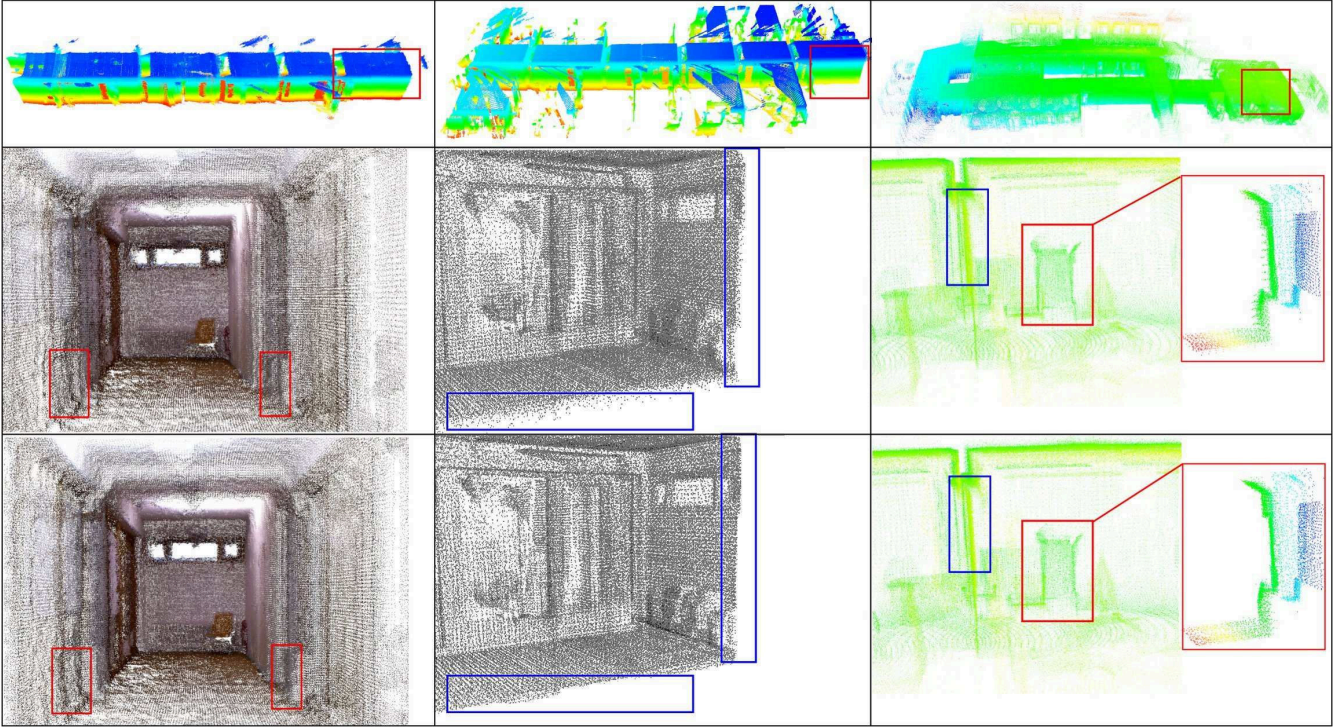


Fig. 5. The first row shows the three datasets (Building 079 - Kinect, Building 079 - Laser and AASS Building), which were used in our experiment IV-A. The second row shows particulars of the datasets before optimization and the bottom row are the same datasets after optimization. We highlighted with rectangles regions where the effect of the optimization is particularly evident.

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

Dataset	Figure	# Scans	# Points	comp. time	grid resolution	entropy input	entropy optimized	entropy reduction
Building 079 - Kinect	5	48	1,434,298	24 min	50 mm	174,092	143,087	17.8%
Building 079 - Laser	5	32	688,977	11 min	50 mm	143,569	92,003.6	35.9%
AASS Building	5	60	2,266,519	48 min	50 mm	287,433	265,585	7.6%
Planar Structure	6	20	313,673	5 min	50 mm	12,493	6,911.43	44.7%
Black Cup	7	51	107,281	6 min	2 mm	8,184.52	6,781.88	17.1%

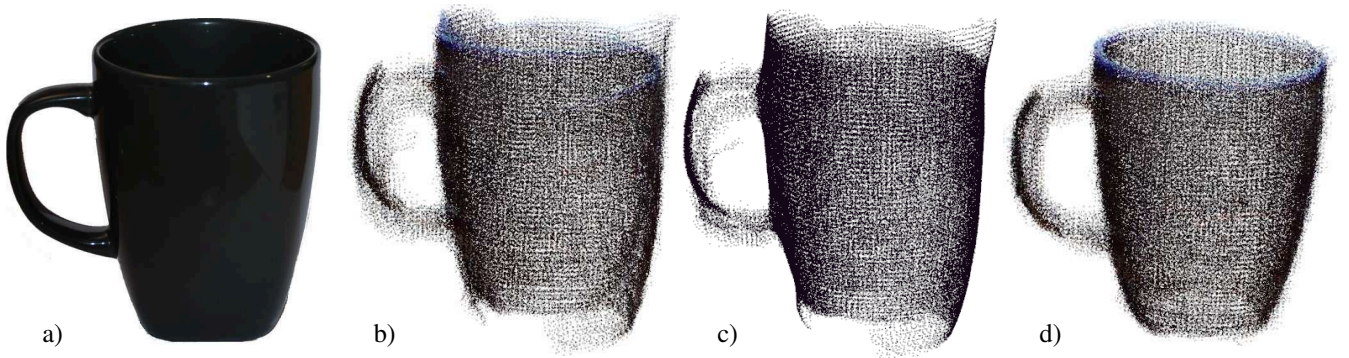


Fig. 7. This figure illustrates the differences between our method and MLS. We acquired multiple scans of a cup, shown in the image (a). In (b) we show the point clouds registered by using ICP. In (c) we show the effect of MLS on the input data. Whereas the model looks smoother than the input it shows 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.

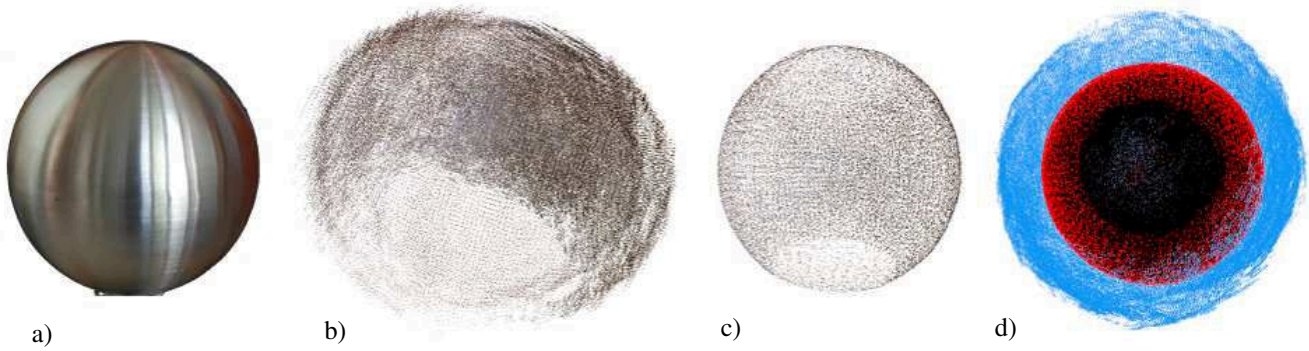


Fig. 8. The first picture (a) shows the object used to acquire the data. (b) shows the accumulated model based on the motion capturing poses. (c) shows the result of our method after 20 iterations. For the direct comparison between the input model (blue) and the resulting model (red) we put both models into (d) together with a sampled sphere as ground truth (black). Note for visibility reasons the picture shows the bottom of the spheres, where the blue and red spheres have no points.

shown in Figure 8 (a). The spherical geometry allowed us to sample a perfect object model as ground truth and the data association is more challenging since the normals provide no distinctive information. We recorded a dataset using a hand held Kinect capturing a metallic sphere on a table and measured the poses of the camera with a motion capturing system. The dataset is publicly available here [2]. Afterwards we selected reference scans every 0.2m and manually removed all points not belonging to the sphere. The accumulated model using the motion capturing poses can be seen in Figure 8(b). We used this model as reference for comparison and as input to our method. The result obtained with our technique can be seen in Figure 8(c). We aligned both spheres to a sampled sphere with the radius of 7.5 cm. The mean error for the input model is 1.94 cm with a maximum error of 4.43 cm. For our method the mean error is 1.4 mm and the maximum error is 6.5 mm. Therefore, our model is substantially more accurate than the input data as can also be seen in Figure 8(d).

V. CONCLUSIONS

In this paper we presented a novel approach to improve the results of pre-aligned 3D point clouds by jointly optimizing the robot poses and the surface points. Our method treats range readings as samples of a regular surface and 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 and for known objects demonstrate that our approach yields substantial improvements compared to a standard SLAM technique and an alternative state-of-the-art method.

REFERENCES

- [1] "Aass-loop dataset," <http://kos.informatik.uni-osnabrueck.de/3Dscans>.
- [2] "Freiburg 2 metallic sphere 2 dataset," <http://cvpr.in.tum.de/data/datasets/rgbd-dataset/>.
- [3] V. Andersen, H. Aanæs, and J. Bærentzen, "Surfel based geometry reconstruction," *Theory and Practice of Computer Graphics* 8.
- [4] 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.
- [5] W. Chang and M. Zwicker, "Global registration of dynamic range scans for articulated model reconstruction," *ACM Trans. Graph.*, vol. 30, pp. 26:1–26:15, May 2011.
- [6] 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.
- [7] F. Dellaert, "Square Root SAM," in *Proc. of Robotics: Science and Systems (RSS)*, Cambridge, MA, USA, 2005, pp. 177–184.
- [8] S. Fleishman, D. Cohen-Or, and C. T. Silva, "Robust moving least-squares fitting with sharp features," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 544–552, 2005.
- [9] 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.
- [10] G. Grisetti, S. Grzonka, C. Stachniss, P. Pfaff, and W. Burgard, "Efficient estimation of accurate maximum likelihood maps in 3D," in *Proc. of the Int. Conf. on Intelligent Robots and Systems (IROS)*, San Diego, CA, USA, 2007.
- [11] 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.
- [12] 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.
- [13] K. Konolige and P. Mihelich, "Kinect calibration tutorial." [Online]. Available: http://www.ros.org/wiki/kinect_calibration/technical
- [14] 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.
- [15] H. Li, R. W. Sumner, and M. Pauly, "Global correspondence optimization for non-rigid registration of depth scans," *Computer Graphics Forum (Proc. SGP'08)*, vol. 27, no. 5, July 2008.
- [16] 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.
- [17] E. Olson, J. Leonard, and S. Teller, "Fast iterative optimization of pose graphs with poor initial estimates," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2006, pp. 2262–2269.
- [18] 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.
- [19] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2011.
- [20] A. V. Segal, D. Haehnel, and S. Thrun, "Generalized-ICP," in *Proc. of Robotics: Science and Systems (RSS)*, 2009.
- [21] C. Stachniss, G. Grisetti, and W. Burgard, "Information gain-based exploration using rao-blackwellized particle filters," in *Proc. of Robotics: Science and Systems (RSS)*, 2005.
- [22] B. Triggs, P. McLauchlan, H. Richard, and A. Fitzgibbon, "Bundle adjustment – a modern synthesis," 2000.