Robust 3D-Mapping with Time-of-Flight Cameras

Stefan May, David Dröschel
and Dirk Holz
Fraunhofer IAIS
stefan_may@arcor.de

Stefan Fuchs
German Aerospace Center (DLR)
Inst. of Robotics and Mechatronics
stefan.fuchs@dlr.de

Andreas Nüchter
Jacobs University Bremen
andreas@nuechti.de

Abstract— Time-of-Flight cameras constitute a smart and fast technology for 3D perception but lack in measurement precision and robustness. We present a comprehensive approach for 3D environment mapping based on this technology. Imprecision of depth measurements are properly handled by calibration and application of several filters. Robust registration is performed by a novel extension to the Iterative Closest Point algorithm. Remaining registration errors are reduced by global relaxation after loop-closure and surface smoothing. A laboratory ground truth evaluation is provided as well as 3D mapping experiments in a larger indoor environment.

I. INTRODUCTION

The mapping task poses a subset of the well-known *simultaneous localization and mapping (SLAM)* problem. Thrun et. al. declare mapping as to be one of the "core competencies of truly autonomous robots" [19]. The purpose of SLAM is to locate the robot, targets and obstacles, which is fundamental for path planning methods. Generating maps is a so-called "chicken-and-egg" problem. If the location of the robot is known, mapping is a solvable task. In reverse, localization is straightforward, if a perfect map is available. But combining both turns out to be a challenging problem referred to as the SLAM or concurrent mapping and localization problem.

Because of their high measurement range and precision, laser scanners and stereo camera systems are mostly used for SLAM so far. But there are some restrictions: Stereo vision requires the matching of corresponding points from two images, and laser scanners only measure sequentially line by line. In contrast, ToF cameras can bridge the gap by providing $2\frac{1}{2}$D images irrespective of textures or illumination. ToF cameras also allow for higher frame rates and thus enable the consideration of motion. With their small manufactured size and little maintenance requirements ToF cameras are serious competitors with laser scanners in the area of 3D mapping.

Anyhow, up to now ToF cameras did not really find their way into 3D mapping especially due to their complex error characteristics and high measurement noise. Depending on external interfering factors (e.g., sunlight) and scene configurations, i.e., distances, orientations and reflectivities, the same scene entails large fluctuations in distance measurements from different perspectives.

This paper presents a 3D mapping approach that handles these problems by calibration and appropriate filtering. It relies only on ToF camera data. No additional sensory information about the sensor's motion is needed. The approach shows promising results and highlights essential impacts that
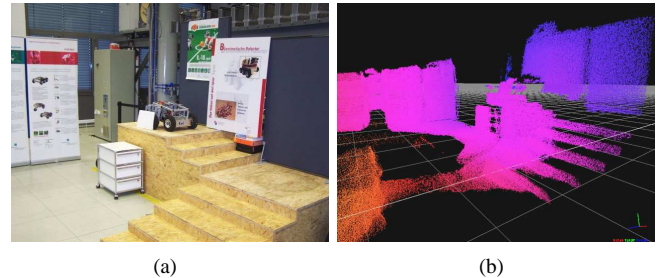

Fig. 1. a) Scenario used for mapping. b) 3D point cloud registered with data taken from a Swissranger SR-3k device (false color code relates distance to origin of coordinate system).

have to be considered in ToF camera mapping. In order to motivate further investigations in 3D mapping with ToF cameras, the underlying data is provided by the authors.

The approach comprises: Depth correction by employing an improved calibration, filtering of remaining inaccuracies, registration w.r.t. to a common coordinate system by a novel extension to the Iterative Closest Point (ICP) algorithm and map refinement including global relaxation - all combined yielding a precise and consistent 3D map.

The remainder of this paper is organized as follows: Section II elaborates 3D mapping approaches and applications related to ToF cameras. Section III describes ToF camera errors, caused by external interfering factors, and the employed depth correction method. In Section IV our mapping approach including 3D pose estimation, error handling and mapping is represented. Section V illustrates experimental results that support our accentuation of employing real-time capable ToF sensors to pose estimation and mapping tasks. Finally, section VI concludes with an outlook on future work.

II. RELATED WORK

A 3D mapping approach tackling large environments was presented by Nüchter et al. [13] using a 3D laser scanner mounted on a mobile robot. Imprecision of inertial sensors was handled by an ICP approach, both for registering consecutive scans and for *closing the loop*. Here, 3D scans were acquired in a stop-scan-go manner by a mobile robot, yielding locally consistent 3D point clouds. Other approaches in localization and mapping, based on 3D data acquired during movement, use either multiple 2D laser range finders facing different orientations, e.g., [18], or a single continuously rotating laser scanner [3], [21].

Matching two complete point clouds is often trapped in a local minimum, especially if the sensor's apex angle is small, such as for ToF cameras (compared with laser scanners). The registration result lacks in precision. In 2006 Ohno et al. used a ToF camera for estimating a robot's trajectory and reconstructing the surface of the environment [14]. The registration procedure for 3D captures was similar to the scan registration approach used by Nüchter et al.. The calculated trajectory was compared with precise reference data in order to demonstrate the algorithm's precision. The estimation error for the robot pose was up to 15 percent in translation and up to 17 percent in rotation, respectively. Also in 2006, Sheh et al. presented an application of ToF cameras in rescue robotics [17]. Their mapping approach was assisted by a human operator. In 2007 Prusak et al. presented a joint approach for robot navigation with collision avoidance, pose estimation and map building employing a ToF camera combined with a high-resolution spherical camera. Structure from motion was used to estimate an initial guess for the pose change [16]. By the use of the trimmed ICP (TrICP) approach, pose changes were finally determined. The TrICP approach extends the ICP algorithm by employing the estimated degree of overlap, which was available from the initial guess. Since ToF cameras also provide amplitude data, 2D image processing algorithms are usable. Thus, the point feature mapping in monochromatic images can be improved by the associated depth data obtaining a 3D feature tracking as performed in [16].

Results of above mentioned approaches are difficult to compare due to different underlying data sets. In contrast to the availability of Computer Vision benchmarks, standard data sets for ToF camera based registrations are not established since yet. We provide the data sets of our experiments in order to motivate benchmarking of registration methods for the ToF camera technology.

III. DESCRIPTION OF TOF CAMERA ERRORS

The performance of distance measurements with ToF cameras is limited by a number of errors. Some of them are inherent in the measurement principle and cannot be corrected. Remaining other errors are predictable and correctable by calibration due to their systematic occurrence. The following explanations relate to them as *non-systematic errors* and *systematic errors*, respectively.

A. NON-SYSTEMATIC ERRORS

There are three significant non-systematic errors. First, a bad signal-to-*noise* ratio distorts the measurement and cannot be suppressed. A solution is either carefully increasing the exposure time and amplifying the illumination or intelligent amplitude filtering. Second, due to interreflections in the scene the remitted near infrared (NIR) signal is a superposition of NIR light that has travelled different distances. This so-called *multiple ways reflection* lets hollows and corners appear rounded off and occluding shapes with a smooth transition. Third, *light scattering* occurs in the lenses of the ToF camera. Thus, near bright objects may superpose

the measurements from the background objects, which for that appear closer. The latter two effects are unpredictable because the topology of the observed scene is unknown a priori.

B. SYSTEMATIC ERRORS

Furthermore, there are three systematic errors. First, there is a *distance-related* error. The measurement principle is based on the assumption that the emitted light is sinusoidal, which is only approximately the case. Second, a so-called *amplitude-related* error is caused by non-linearities of the pixel's electronic components. As a result, the measured distance varies with object reflectivity. Third, there is a *fixed pattern phase noise*. Since the pixels on the sensor chip are connected in series, the triggering of each pixel depends on the position on chip. The farther the pixel is located with respect to a signal generator, the higher is its measurement offset. These three errors are managable by calibration. In [5] and [6] Fuchs et al. described an appropriate calibration method, which estimates above mentioned errors. It is applied for the mapping experiments in this paper.

IV. 3D MAPPING

The here presented approach acquires a metric map in form of a 3D point cloud, whereby the sensor is simultaneously localized relative to this point cloud without any external positioning system. The 3D mapping is a four-stage process. First, invalid data points are discarded by filtering. Second, the map is generated by registering consecutive 3D captures. Third, accumulated errors are relaxed among all captures. Finally, the map is enhanced with refinement filters.

1) Filtering: Errors caused by low illumination or occlusion are treated by filtering. A high confidence is related to a high amplitude (to be precise: this statement is only a compromise to what the camera provides; see [7] for a description of error influences). Thresholding the amplitude discards primarily data resulting from objects with lower infrared reflectivity, higher distance or from objects which are located at the peripheral area of the measurement volume (due to inhomogeneous scene illumination). Mismeasurements also occur on jump edges, i.e., when the transition from one to another shape appears disconnected due to occlusions. The true distance changes suddenly for the transition from one shape to the other whereas ToF cameras measure a smooth transition. In order to remove these invalid points, jump edge filtering has been applied [5]. It is important to mention that the proposed filter is sensitive to noise, i.e., besides jump edges, valid points are removed, if noise reduction filters are not being applied first. The subsequent application of median and jump edge filtering achieved the best results in our experiments.

2) Map Generation: Pose changes are assumed to be small due to the high frame rate. For that, no additional sensory data as initial guess is needed (e.g., inertial sensors). The estimation is performed relying completely on the registration of 3D data by the use of the well-known ICP algorithm [1]. It aims at finding a rigid transformation

between a model point set M and a scene point set D by performing an iterative least square minimization scheme. In each iteration step corresponding closest points are determined. Denoting corresponding point pairs as a set of N tuples $\{(\mathbf{m}_i, \mathbf{d}_i)|i = 1 \ldots N\}$ and the transformation as composition of a rotation matrix and a translation vector $(\mathbf{R}, \mathbf{t})$ the error function reads:

$$E(\mathbf{R}, \mathbf{t}) = \sum_{i=1}^{N} ||\mathbf{m}_i - (\mathbf{R}\mathbf{d}_i + \mathbf{t})||^2 . \quad (1)$$

The solution can be determined by several closed-form algorithms. A detailed description and evaluation of these algorithms is summarized in [11].

The original formulation of the ICP approach assumes that the scene point set is completely covered by the model point set [1]. If the scene includes points which are not part of the model, wrong correspondences are assigned which skew the result [4]. The simplest solution to discard scene points from a non-overlapping area is the employment of a distance threshold. Corresponding tupels are rejected, if their Euclidean distance exceeds this value. Several strategies are possible to determine suitable thresholds, e.g., a gradual decreasing threshold with respect to the iteration step. In general, these thresholds increase the registration performance significantly on only partially overlapping point clouds. For convenience, the original formulation of the ICP approach including a distance threshold is called *Vanilla ICP* approach in the following.

A plain threshold has limitations with respect to robustness and accuracy. Several approaches have been proposed to improve registration results for an unknown degree of overlap. Fusiello et al. employed the X84 rejection rule, which uses robust estimates for location and scale of a corrupted Gaussian distribution [4]. It estimates a suitable rejection threshold concerning the distance distribution between corresponding points. Niemann et al. proposed a rejection rule that considers multiple point assignments *(Picky ICP algorithm)* [12] . If more than one point from the scene point set is assigned to a corresponding model point, only the point with the nearest distance is accepted. Pajdla and Van Gool proposed the inclusion of a reciprocal rejection rule *(Iterative Closest Reciprocal Point algorithm - ICRP)* [15]. For a corresponding point pair $(\mathbf{m}_i, \mathbf{d}_i)$, which has been determined by searching the nearest neighbor of $\mathbf{d}_i$ in M, the search is reversed subsequently, i.e., for $\mathbf{m}_i$ the nearest neighbor in D is determined. This needs not to be the same scene point and is denoted with $\mathbf{d}'_i$. The point correspondence is rejected if $\mathbf{d}_i$ and $\mathbf{d}'_i$ have a distance being larger than a certain threshold.

The method proposed here to overcome an unknown degree of overlap addresses the problem formulation more precisely and intuitively. It stems from 3D computer graphics and is called *frustum culling* [8]. A frustum defines the volume that has been in the range of vision while acquiring the model point set (cf. Fig. 2). Luck et al. used frustum culling for pre-filtering based on an initial pose estimate [9]. The iterative registration process was then performed on the
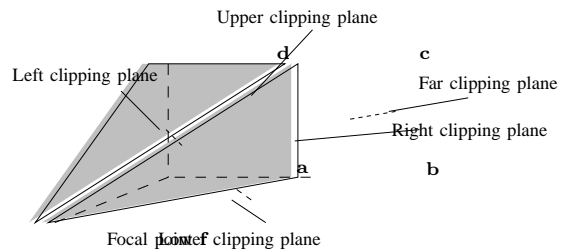

Fig. 2. Definition of frustum clipping planes.

reduced data set. Contrary, we employ no pose estimation and thus cannot pre-filter data in order to remove the non-overlapping area. Frustum culling is therefore embedded in the iteration process by employing the pose estimate of the previous iteration step. Scene points outside of the model frustum are filtered by testing against clipping planes before performing nearest neighbor searching.

Let $\mathbf{f} = \mathbf{0}$ be the focal point and $\{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}\}$ vectors from $\mathbf{f}$ to the four edge points of a far clipping plane. Lateral clipping planes are spanned by each two of adjacent vectors (cf. Fig. 2). The normal vectors $\mathbf{n} = (n_x, n_y, n_z)^\mathsf{T}$ of the four lateral clipping planes are then used to check if a point $\mathbf{x} = (x, y, z)^\mathsf{T}$ is inside the frustum by

$$xn_x + zn_z < 0 \quad (2)$$

for the left and right clipping planes and

$$yn_y + zn_z < 0 \quad (3)$$

for the upper and lower clipping planes. The approach is non-parametric and removes iteratively scene points from non-overlapping areas by evaluating their visibility from the model's viewpoint. During the iteration process scene points are clipped as soon as they leave the visibility frustum, which addresses exactly the problem formulation. This extension is called *Frustum ICP* approach in the following.

3) Error relaxation: Errors sum up due to the limited sensor precision and accumulation of registration errors from pairwise ICP matching of consecutive frames. However, when the field of view is overlapping with an earlier one, a loop can be closed. This enables SLAM algorithms to bound the error and to compute a consistent map.

We use a GraphSLAM approach [20], [19] that extends Eq. (1). The graph contains links of all overlapping 3D scans and therefore the loop closing. Instead of minimizing the Euclidean distance between two point pairs, the Euclidean distance between 3D point clouds connected by a graph edge (j, k) is minimized:

$$E(\mathbf{R}, \mathbf{t}) = \sum_{j \to k} \sum_{i=1}^{N} ||(\mathbf{R}_j \mathbf{p}_{j,i} + \mathbf{t}_j) - (\mathbf{R}_k \mathbf{p}_{k,i} + \mathbf{t}_k)||^2 \quad (4)$$

To solve the equation above, we have to fix the first acquired 3D point cloud, i.e., set $\mathbf{R}_1 = \mathbf{I}$ and $\mathbf{t}_1 = (0, 0, 0)$. Unfortunately Eq. (4) cannot be solved in closed form. Using the small angle assumption the linearization $\sin x \approx x$ and
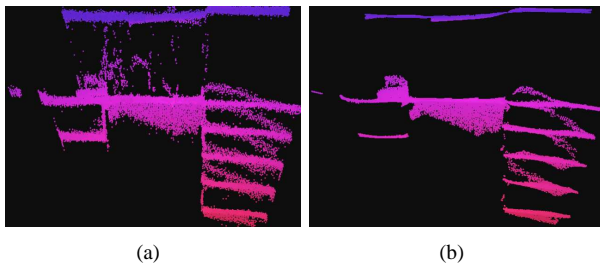
Fig. 3. a) 3D map before refinement (bird's view of scene in Fig. 1). b) 3D map after refinement. Sparse points are removed and surfaces are smoothed.

$\cos x \approx 1$ yields an approximation of a rotation matrix:

$$\mathbf{R} = \begin{pmatrix} 1 & -\theta_z & \theta_y \\ \theta_y\theta_x + \theta_z & 1 - \theta_y\theta_x\theta_z & -\theta_x \\ -\theta_y + \theta_x\theta_z & \theta_y\theta_z + \theta_x & 1 \end{pmatrix}. \qquad (5)$$

Furthermore, we assume that the multiplication of small values results in even smaller values which thus can be omitted. The rotation matrix $\mathbf{R}$ then simplifies to

$$\mathbf{R} = \begin{pmatrix} 1 & -\theta_z & \theta_y \\ \theta_z & 1 & -\theta_x \\ -\theta_y & \theta_x & 1 \end{pmatrix}. \qquad (6)$$

Using these approximations, Eq. (4) is transformed into a quadratic form and solved. Note: This approximation is valid, if we apply the centroid trick, i.e., subtracting the center of mass calculated from corresponding points in Eq. (1) and (4). This is commonly used for solving the ICP error function [1]. A detailed derivation of this global relaxation formula and a solution of the resulting minimization problem is given in [11].

4) Refinement: The refinement of a 3D map comprises filtering of sparse points and approximation of plain patches to the neighborhood of each point. Removing the set of sparse points S is done by determining the mean distance of k-nearest neighbors as a density measure,

$$d(\mathbf{p}_i) = \frac{1}{k}\sum_{n=1}^{k} ||(\mathbf{p}_i - \mathbf{p}_{i,n})||, \qquad (7)$$

$$S = \{\mathbf{p}_i \in P \mid d(\mathbf{p}_i) < d_{th}\}, \qquad (8)$$

where d_{th} is a constant threshold. The set of point candidates Q for the resulting 3D map includes all remaining points after filtering:

$$Q = P\backslash(J \cup S). \qquad (9)$$

Compared with laser scanners, the noise level of ToF cameras is much higher. Smooth surfaces appear with a certain thickness of some centimeters. This effect can be seen in Fig. 3(a). For the refinement of a composed 3D point cloud a principle component analysis (PCA) is performed to detect surface normals. Related pixels are shifted along these vectors towards the detected surfaces. The resulting 3D map after applying the proposed refinement is contrasted in Fig. 3(b).

V. EXPERIMENTS AND RESULTS

The following experiments demonstrate the accuracy and robustness of the proposed 3D mapping approach. Therefore, a SR-3k ToF camera with a resolution of 176×144 pixels is used. The standard deviation of this camera varies from 20 mm to 50 mm, the horizontal and vertical apex angles are $45\,^\circ$ and $35\,^\circ$ respectively. At first, laboratory experiments demonstrate the reachable accuracy and influences of the light scattering effect. For this purpose, a ground truth evaluation is provided by an industrial robot arm (KUKA KR 16) to which the ToF camera has been attached. The robot's positioning system has an accuracy of $1\,\text{mm}$ and $0.1\,^\circ$. At second, the robustness of the 3D mapping approach is shown in a larger environment.

A. EVALUATION MEASURES

The quality of an 3D mapping approach can be rated either by comparing the estimated path (ego motion) of the sensor with a ground truth or by comparing the created 3D map with a ground truth. The absolute comparison of poses against this ground truth measure provides only a weak objectivity since registration errors can be compensated by other registration errors. Therefore, incremental measures are more suitable in terms of validity, i.e., an integration of angular and translational error increments induced by registration of subsequent frames. The translational error measure is defined by

$$e_{inc,\Delta\mathbf{t}} = \sum ||\Delta\mathbf{t}_{inc,i}||, \qquad (10)$$

where $\Delta\mathbf{t}_{inc,i}$ is the translational registration error of frame i. In order to make rotational errors comparable, the rotation axis has to be considered:

$$e_{inc,\Delta\theta} = \sum || \,|\Delta\theta_{r,i}|\mathbf{a}_{r,i} - |\Delta\theta_{e,i}|\mathbf{a}_{e,i}||, \qquad (11)$$

where $\Delta\theta_{r,i}$ is the ground truth angle around the ground truth rotation axis $\mathbf{a}_{r,i}$ of frame i, whereas $\Delta\theta_{e,i}$ and $\mathbf{a}_{e,i}$ constitute the estimated counterparts of ICP registration. These definitions provide an uniform measure for the accumulation of translational and rotational errors, Eq. (11) is similar to Eq. (10).

Accurate ego motion does not inevitably result in a perfect 3D map. Even with the exact sensor localization, inaccuracy can be induced by corrupted depth measurements. Even if the sensor is localized exactly, corrupted depth data will result in an inaccurate map. On this account, the second measure is constituted by the *isometry* of the resulting map. Characteristic opposing walls within the scene were measured with a tape (cf. Fig. 4).

B. CALIBRATION AND LIGHT SCATTERING

First, the achievable accuracy in ego motion estimation was investigated in applying two different trajectories. Fig. 5 and Fig. 6 depict scenes and performed paths. In the first scene (*SI*, cf. Fig. 5) the camera moves around a basic geometric Styrofoam object. Contrary, the object was moved around the camera in the second scene (*SII*, cf. Fig. 6). Data takes for both trajectories were performed twice: Primary,

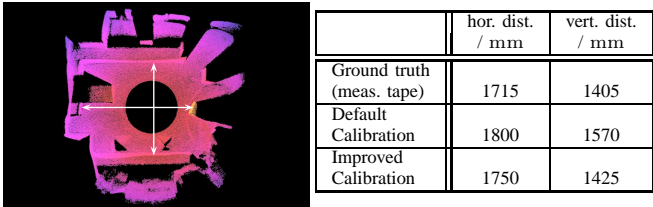|	hor. dist. / mm	vert. dist. / mm
Ground truth (meas. tape)	1715	1405
Default Calibration	1800	1570
Improved Calibration	1750	1425

Fig. 4. Isometry of resulting 3D map of the laboratory scene (*SIII*). The distances of opposing walls were manually measured and assumed to be the ground truth. The improved calibration reduces deviations from the ground truth and provide a more accurate mapping result.
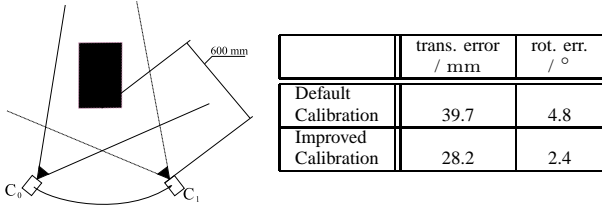

	trans. error / mm	rot. err. / °
Default Calibration	39.7	4.8
Improved Calibration	28.2	2.4

Fig. 5. Experimental setup (*SI*) for identifying the accuracy and demonstrating the impact of calibration. The camera is rotated by $90\,\deg$ around an object in a distance of $600\,\mathrm{mm}$ while keeping it in the center of the field of view. Thus, a distance of $950\,\mathrm{mm}$ is covered. Both, the translational error and the rotational error decrease due to the proper calibration.

the scenes were captured with the default (manufacturer's) calibration. This calibration considers fixed pattern phase noise. Then, the scenes were captured with an improved calibration. The camera was calibrated according to Fuchs et. al. [6] in order to additionally consider distance-related errors. This approach employs a spline for estimating a pixel-wise depth measurement correction value. Fig. 8 depicts the computed correction spline.

In both scenes the calibration reduces the error in ego motion estimation by $\approx 25\,\%$. But, compared to the length of both trajectories, the translational error in *SII* is significantly larger ($20\,\%$) compared to the error resulting from *SI* ($3\,\%$). We conclude that the type of movement is crucial for the result. Especially in *SII*, the observed object is moving at the margin of the field of view where the low resolution and small apex angle handicap the ego motion estimation.

In addition, the light scattering has to be considered, as the second investigation shows. Here, we induced scattering effects by adding an Styrofoam object to *SII* at several distances. Fig. 6 demonstrates the results. The nearer the disturbing object moves to the camera the more the ego motion estimation results degrade. Obviously, the light scattering affects only the translation. Strong influences are noticable when high reflective objects come into the field of view or when distances to objects in the scene are heterogeneous. Hence, we considered this fact in designing the next experiment.

C. 3D MAPPING OF LABORATORY SCENE

Second, the laboratory scene was enlarged. The Styrofoam objects were assembled in a square, which measured approximately $1800\,\mathrm{mm}$. This scene (cf. Fig. 7) was circumferentially captured by moving the camera on a circular path with a diameter of $300\,\mathrm{mm}$. In total, 180 captures were

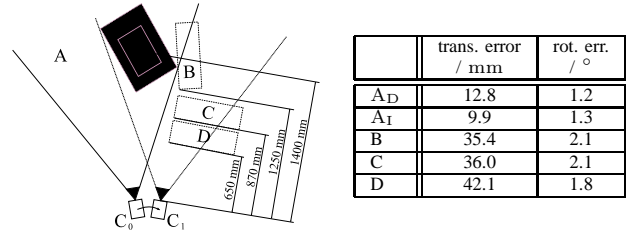|	trans. error / mm	rot. err. / °
A_D	12.8	1.2
A_I	9.9	1.3
B	35.4	2.1
C	36.0	2.1
D	42.1	1.8

Fig. 6. Experimental setup (*SII*) for identifying the accuracy and the impact of light scattering. The camera is stepwise moving ($50\,\mathrm{mm}$) and rotating ($22\,\deg$) from C_0 to C_1. Initially the scene consists of two Styrofoam cuboids standing on top of each other (case A). The improved calibration (A_I) shows to reduce the translational error from $12.8\,\mathrm{mm}$ to $9.9\,\mathrm{mm}$. Then (cases B, C, D), an additionally Styrofoam cuboid was put into the scene. The ego motion estimation results degrade to $42.1\,\mathrm{mm}$. The rotational error nearly stays constant.
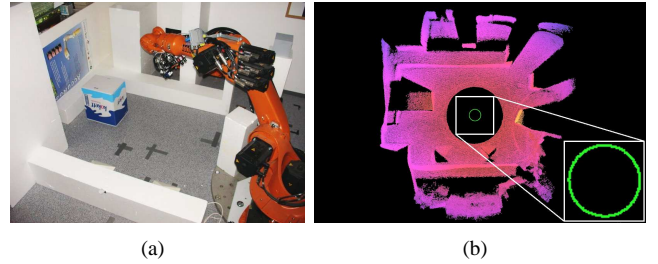

(a)   (b)

Fig. 7. a) Laboratory environment used for ground truth evaluation. The ToF camera is mounted on an industrial robot arm (KUKA KR 16). b) Bird's view of the 3D map created in this environment. The camera was stepwise moved on a circular path with a diameter of $300\,\mathrm{mm}$ (green line).

taken. In order to reduce the measurement noise, the camera's automatic integration time controller was activated, which adjusted the integration time between $8000\,\mu\mathrm{s}$ and $12000\,\mu\mathrm{s}$.

Again, two test series with the default and improved calibration were performed. Here, the improved calibration reduces the pose-estimation error only slightly. We assume the light scattering effect completely adumbrates the results. However, the major benefit concerns the isometry of the resulting map. The deviation from the ground truth (measuring tape) reduces from $85\,\mathrm{mm}$ for the horizontal and $165\,\mathrm{mm}$ for the vertical distance measure to $35\,\mathrm{mm}$ and $20\,\mathrm{mm}$ (cf. Fig. 4).
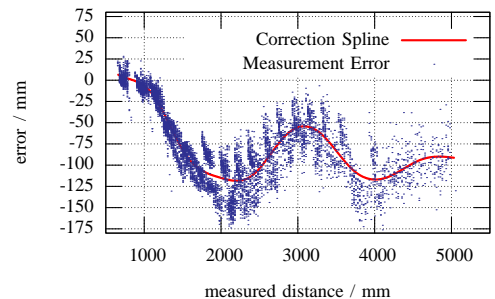

Fig. 8. Distance calibration result: The identified distance-related error is plotted by subtracting the measured distance from the real distance. An overall sinusoidal curve of the distance-related error is apparent. Altogether, the camera measures a greater distance. Thus a spline is fitted into the error for correction of distance measurements.
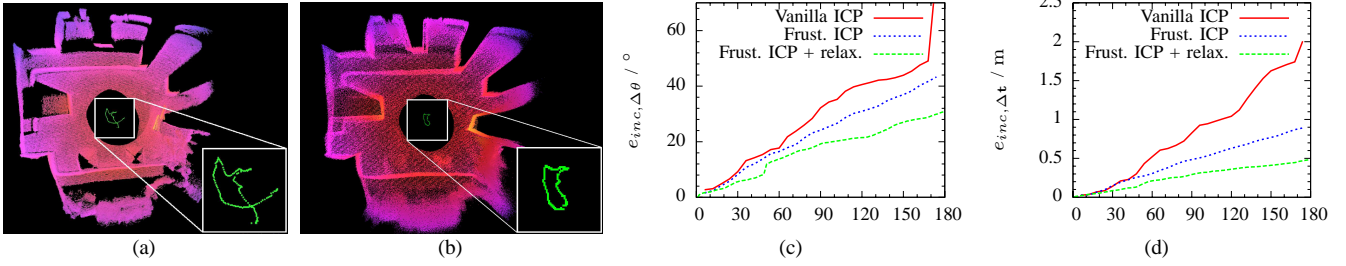
Fig. 9. a,b) 3D Map reconstruction: The estimated trajectories are drawn in green color. a) Frustum ICP approach. b) Frustum ICP approach with error relaxation. Remaining trajectory distortions are caused by non-systematic errors. c,d) Ground truth comparison of ICP registration. The proposed Frustum ICP extension is benchmarked against the Vanilla ICP approach. Global relaxation further improves the accuracy. c) Incremental angular error measure. Please note the large error induced for the Vanilla ICP approach between frame 160 and 170 in (c), which were caused by convergence to wrong local minima. The Frustum ICP approach provided more robustness. d) Incremental translational error measure.

Furthermore, we compared the performances of Vanilla ICP, Frustum ICP and Frustum ICP with error relaxation applying them to the calibrated data. In Fig. 9(a) and 9(b) the resulting maps are illustrated. The green line represents the estimated ego motion of the camera. Both paths only partially agree with the circle in Fig. 7. An experiment employing the ground truth poses as initial guess for the ICP registration provided convergence to nearly the same minima. Thus, errors in ego motion estimation primarily arise from unsystematic errors, i.e., light scattering and multiple-ways-reflection. The effect is more noticeable for the second half of the circle, where near objects appeared in the field of view. Those influences are dependent on the scene and thus difficult to model or compensate. Fig. 9(c) and 9(d) contrasts the performance of the employed ICP approaches. The Frustum ICP approach provided more robustness and higher accuracy as the Vanilla ICP approach. The registration error could be additionally reduced by error relaxation.

D. 3D MAPPING OF LARGER ENVIRONMENTS

In a further experiment the robustness of the proposed mapping approach was tested in a larger environment, the robotic pavilion at the Fraunhofer Institute IAIS (cf. Fig. 1). It sizes $19.4\,\mathrm{m}$ in the longest distance. In total, 325 captures have been taken on a closed trajectory. Since the unambiguity interval is limited to $7.5\,\mathrm{m}$, measurements appear closer (modulo $7.5\,\mathrm{m}$) than they are, if this value is exceeded. Mostly, the amplitude value can be used to discard those measurements, but not in all cases. These mismeasurements occur especially when surfaces with specular reflectivity are present, e.g., mirrors, window panes and metallic surfaces. Some objects have been covered or removed from the scene before the exploration was started. Future technologies will address this problem on the sensor level, e.g., by the employment of multiple frequencies or binary coded sequences.

The scene features more dynamic range compared to the laboratory scene, i.e., a larger working range and variability of infrared reflectivity. The automatic integration time controller adjusted the exposure time between $21000\,\mu\mathrm{s}$ and $65000\,\mu\mathrm{s}$.

Ground truth data, in terms of an external positioning system, was not available. Therefore, the evaluation of

Calib. method	3DOF		6DOF	
	$\lVert\Delta t\rVert$ / m	$\Delta\theta$ / °	$\lVert\Delta t\rVert$ / m	$\Delta\theta$ / °
Default	0.27	26.10	1.23	31.58
Improved	0.30	0.81	1.74	9.03

TABLE I
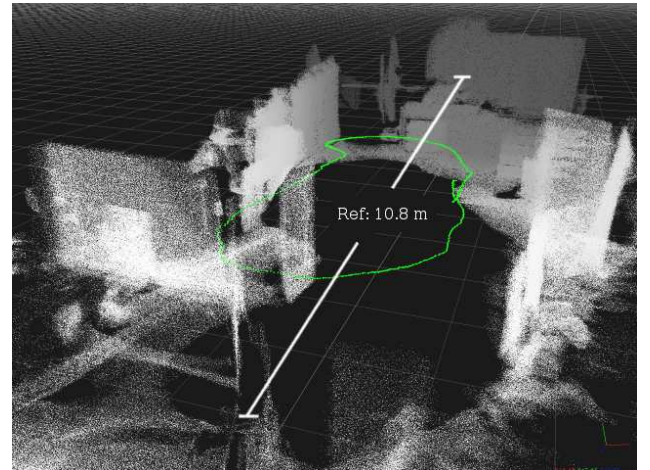COMPARISON OF POSE ESTIMATION EMPLOYING DIFFERENT CALIBRATION METHODS.

Fig. 10. 3D map of the whole robotic pavilion (perspective view of scene in Fig. 1). The trajectory (estimated poses of each frame) is drawn in green.

precision in reconstruction is specified with the distance between two opposing walls and the pose estimation error after closing the loop. The start-to-end frame registration before relaxation provided an accumulated translational error of $\lVert\Delta\mathbf{t}\rVert = 1.74\,\mathrm{m}$ and an accumulated rotational error of $\lVert\Delta\theta\rVert = 9.03°$, which is small enough to perform loop-closure. The calculated distance of two opposing walls from the 3D map compared with the distance determined with a measuring tape deviates $0.4\,\mathrm{m}$ (Measuring tape vs. 3D cloud: $10.8\,\mathrm{m}$ / $11.2\,\mathrm{m}$; cf. Fig. 10). Influences of the improved calibration approach were more noticeable in this experiment (cf. Table I). A video showing the performance of map creation in this environment can be found in [10].

VI. CONCLUSIONS AND FUTURE WORK

This paper presented a method for precise 3D environment mapping. It employed only a 3D time-of-flight (ToF) camera and no additional sensors. The achieved precision for a 3D map composed of 180 single captures was calculated against ground truth data provided by an industrial robot. Several measures were necessary for achieving the presented results in 3D map creation. First, a calibration method has been applied to reduce distance measurement errors. Second, the removal of mismeasurements has been achieved by the development of suitable filters. Third, robust pose estimation was achieved by an improved ICP algorithm. Finally, the consistency of a 3D map has been enhanced in a refinement step, relaxing accumulated errors and smoothing surfaces. The robustness of the entire approach has been demonstrated while registering 325 single captures obtained from a larger indoor environment. By providing the underlying data set of the laboratory scene on our website[1], we want to motivate further investigations in 3D mapping based on ToF camera data and enable comparable results and benchmarks.

Future work will concentrate on the improvement of calibration, e.g., the consideration of fluctuation in depth measurements caused by exposure time control, on the improvement of 3D map creation, e.g., by enhancing the semantic information, and on fusing ToF camera data with additional sensory information, e.g., with inertial measurement units.

REFERENCES

[1] P. Besl and N. McKay. "A method for Registration of 3–D Shapes", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, no. 2, pp. 239–256, 1992.

[2] D. Chetverikov, D. Svirko, D. Stepanov and P. Krsek. "The trimmed iterative closest point algorithm". *Proceedings of the 16th International conference on pattern recognition (ICPR)*, vol. 3, pp. 545–548, 2002.

[3] D. M. Cole and P. M. Newman. "Using Laser Range Data for 3D SLAM in Outdoor Environments", *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2006.

[4] A. Fusiello, U. Castellani, L. Ronchetti and V. Murino. "Model acquisition by registration of multiple acoustic range views", *Proceedings of the 7th European Conference on Computer Vision (ECCV)*, pp. 805–819, 2002.

[5] S. Fuchs and S. May. "Calibration and registration for precise surface reconstruction with ToF cameras". *Proceedings of the Dynamic 3D Imaging Workshop in Conjunction with DAGM (Dyn3D)*, vol. I, 2007.

[6] S. Fuchs and G. Hirzinger. "Extrinsic and Depth Calibration of ToF-cameras". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[7] R. Lange. "3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology", Dissertation, University of Siegen, 2000.

[8] E. Lengyel. Game Programming Gems, chapter 4.5: A Fast Cylinder-Frustrum Intersection Test. Charles River Media, 2000.

[9] J. Luck, C. Little and W. Hoff. "Registration of Range Data Using a Hybrid Simulated Annealing and Iterative Clostest Point Algorithm", *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2000.

[10] S. May, D. Droeschel, D. Holz, C. Wiesen and S. Fuchs. Video: "3D Pose Estimation and Mapping with Time-of-flight Cameras", http://www.iais.fraunhofer.de/3325.html, 2008.

[11] A. Nüchter. *3D Robotic Mapping*, Springer STAR series, 2008.

[12] H. Niemann, T. Zinßer and J. Schmidt. "A refined ICP algorithm for robust 3-D correspondence estimation", *Proceedings of the IEEE International Conference on Image Processing*, 2003.

[13] A. Nüchter, K. Lingemann, J. Hertzberg and H. Surmann. "6D SLAM - 3D Mapping Outdoor Environments", *Journal of Field Robotics*, vol. 24, no. 8–9, pp. 699–722, 2007.

[14] K. Ohno, T. Nomura and S. Tadokoro. "Real-Time Robot Trajectory Estimation and 3D Map Construction using 3D Camera". *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2006.

[15] T. Pajdla and L. V. Gool. "Matching of 3-D curves using semi-differential invariants", *Proceedings of the Fifth International Conference on Computer Vision (ICCV)*, pp. 390–395, 1995.

[16] A. Prusak, O. Melnychuk, H. Roth, I. Schiller and R. Koch. "Pose Estimation and Map Building with a PMD-Camera for Robot Navigation". *Proceedings of the Dynamic 3D Imaging Workshop in Conjunction with DAGM (Dyn3D)*, vol. I, 2007.

[17] R. Sheh, M. W. Kadous and C. Sammut. "On building 3D maps using a Range camera: Applications to Rescue Robotics", Technical Report UNSW-CSE-TR-0609, School of Computer Science and Engineering, The University of New South Wales, Sydney, Australia, 2006.

[18] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, et al., "Stanley: The Robot that Won the DARPA Grand Challenge", *Journal of Field Robotics*, vol. 23, no. 9, pp. 661–692, 2006.

[19] S. Thrun, W. Burgard and D. Fox, *Probabilistic Robotics*, The MIT Press, 2005

[20] S. Thrun and M. Montemerlo. "The GraphSLAM Algorithm With Applications to Large-Scale Mapping of Urban Structures", *International Journal on Robotics Research*, vol. 25, no. 5/6, pp. 403–430, 2005.

[21] O. Wulf, D. Lecking and Bernardo Wagner. "Robust Self-Localization in Industrial Environments based on 3D Ceiling Structures", *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2006.

[1]http://www.robotic.de/(242/