3D Pose Estimation and Mapping with Time-of-Flight Cameras

Stefan May, David Droeschel, Dirk Holz and
Christoph Wiesen
Fraunhofer Institute for Intelligent Analysis and
Information Systems (IAIS)
Schloss Birlinghoven
53754 Sankt Augustin, Germany
stefan.may@iais.fraunhofer.de

Stefan Fuchs
German Aerospace Center (DLR)
Institute of Robotics and Mechatronics
82234 Wessling, Germany
stefan.fuchs@dlr.de

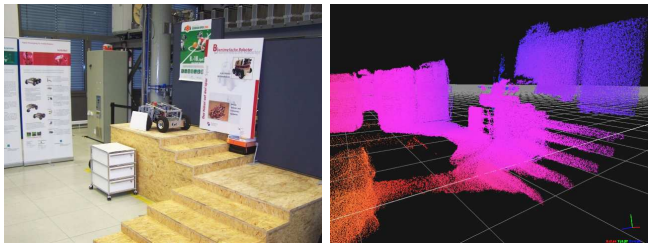
***Abstract*— This paper presents a method for precise 3D environment mapping. It employs only a 3D Time-of-Flight (ToF) camera and no additional sensors. The camera pose is estimated using visual odometry. Imprecision of depth measurements caused by external interfering factors, e.g. sunlight or reflectivities are properly handled by several filters. Pose tracking and mapping is performed on-the-fly during exploration and allows even for hand-guided operation. The final refinement step, comprising error distribution after loop-closure and surface smoothing, further increases the precision of the resulting 3D map[1].**

I. INTRODUCTION

Since their invention nearly a decade ago, Time-of-Flight (ToF) cameras have attracted attention in many fields, e.g. automotive engineering, industrial engineering, mobile robotics and surveillance. So far, 3D laser scanners and stereo camera systems are mostly used for these tasks due to their high measurement range and precision. Stereo vision requires the matching of corresponding points from two images to obtain depth information, which is directly provided by laser scanners but with the drawback of a lower frame rate. In contrast to laser scanners, ToF cameras allow for higher frame rates and thus enable the consideration of motion. However, the high frame rate has to be balanced with measurement precision. Depending on external interfering factors (e.g. sunlight) and scene configurations, e.g. distances, orientations and reflectivities, the same scene entails large fluctuations in distance measurements from different perspectives. These influences cause systematic errors besides noise and have to be handled by the application. As a result laser scanners are mostly used for mapping purposes, e.g. [14], [20], [3], [11], [19].

In this paper we present a mapping approach, which deals with large variations in precision of distance measurements. Mapping is performed on-the-fly with no additional sensory information about the sensor's ego-motion. The approach comprises: feature based ego-motion estimation, filtering of imprecise data and registration of newly acquired data for a consistent 3D environment map. After loop-closure, a refinement step distributes the error and smoothes the measurements yielding in a precise 3D map.

[1]A video, showing the performance of the approach, is available at http://www.iais.fraunhofer.de/3325.html



(a) (b)

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

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. 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

Surface reconstruction is a basic task for object detection, manipulation and environment modelling. Generally, the object's surface is reconstructed by merging measurements from different views. This approach requires depth data and sensor pose data. When both, pose and depth, are unknown, *structure from motion* is a solution. Corresponding features in consecutive images are used to estimate the ego-motion of the sensor. Based on this ego-motion information the depth without absolute scale is estimated. If only the depth information but no pose is given, i.e. by using a stereo camera or a laser scanner system without inertial sensors, the *Iterative Closest Point* (ICP) algorithm can be used to register point clouds acquired from different perspectives [2]. Finally, if pose and depth are known, the registration procedure is dispensable and the data can simply be merged. In any case, the quality of surface reconstruction depends on the precision of sensor pose estimation and depth measurement.

A 3D mapping approach tackling large environments was presented by Nuechter et al. [14] using a 3D laser scanner mounted on a mobile robot. Imprecision of inertial sensors was handled with 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, either use multiple 2D laser range finders facing different orientations, e.g. [19], or a single continuously rotating laser scanner [3], [20], [11].

Matching two complete point clouds is often trapped in a local minimum, especially if the sensor's apex angle is small like in the case of ToF cameras (compared with laser scanners). Besides, it is computationally too expensive for real-time applications dealing with large data sets. Sappa et al. have shown that range image registration is more robust and faster using only an edge-based representation [16]. It is based on the scan line approximation technique of Jiang et al. [10], which represents a good approach in detecting distinguishable edges in range images.

In 2006 Ohno et al. used a ToF camera for estimating a robot's trajectory and reconstructing the surface of the environment [15]. The registration procedure for 3D captures was similar to the scan registration approach used by Nuechter 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.

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. Most common methods for matching 2D image features are based on the KLT (Kanade-Lucas-Tomasi) [18], SIFT (Scale-Invariant Feature Transform) [9] or SURF (Speeded Up Robust Features) [1] approaches. Calculating changes in the 3D pose based on these methods have been performed by a number of works, e.g. [4], [5], [6].

Within this paper range data registration is used to give a quantity of achievable precision for 3D mapping tasks with ToF cameras. This novel sensor technology enables the merging of 2D and 3D image processing, but suffers in measurement precision. We show how the precision and robustness of a 3D mapping process can be enhanced. Involved methods are, amongst others, a proper ToF camera calibration, amplitude and range based filtering, robust pose estimation and 3D map refinement.

III. DESCRIPTION OF TOF CAMERA ERRORS

Although a lot of effort has been investigated, depth measurements with ToF cameras are still erroneous. Within this section primarily errors are mentioned that arise from the special measurement principle of a ToF camera, i.e. modulation interferometry. We have to distinguish between systematic and non-systematic errors.

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. The answer is either to carefully increase the exposure time and to amplify the illumination respectively or to filter the measured values that are marked with low amplitudes. Second, the so-called *multiple ways reflection* occurs in concave objects, e.g. corners or hollows. Here, the remitted near infrared (NIR) signal is superposed by NIR light that has covered a longer distance. Third, *light scattering* of close objects influences the depth measurement of pixels that belong to far objects. The latter two effects are unpredictable, since the topology of the observed scene is unknown a priori.

B. SYSTEMATIC ERRORS

Furthermore, there are four systematic errors. First, there is a *distance-related*[2] error steming from the asymmetric response of a NIR-LED-signal. This asymmetric response causes a non-harmonic sinusoidal illumination of the scene. Since a harmonic sinusoidal illumination is the basic assumption in the principle of modulation interferometry, the computed phase-delay and distance respectively are inaccurate. Second, due to non-linearities of the semiconductor and its imperfect separation properties a different number of photons at a constant distance cause different distance measurements. This so-called *amplitude-related* error also depends on the actual distance. Third, there is a *latency-related* error. Since the emitted and remitted signals are correlated on the sensor array directly, different latencies for every pixel have to be considered. Finally, there is a *fixed pattern noise* (FPN) that comes along with the latency-related error. FPN is related to latencies and different material properties in each CMOS-gate. Thus, a constant pixel offset for each pixel can be identified. These four errors are manageable by calibration. In [7] and [8] Fuchs et al. described an appropriate calibration method that estimates above mentioned errors. As a result an overall precision of $1\,\mathrm{mm}$ is achievable.

IV. SIMULTANEOUS LOCALIZATION AND MAPPING

Due to large data sets, the process of creating detailed 3D maps is time consuming. Current ToF cameras provide range and amplitude images with a resolution up to 176×144 pixels and 30 frames per second. This data rate has to be processed online for applications that interact with the real world. While localization is feasible on a reduced data set, the creation of detailed 3D maps incorporates a larger portion of it, not seldomly the whole data volume. The latter process can partially be done offline and comprises error distributing and refining algorithms. In the following two subsections camera pose estimation and mapping are addressed in detail. The degree of data aggregation for the

[2]This distance-related error is also called *wiggling error* or *circular error*.

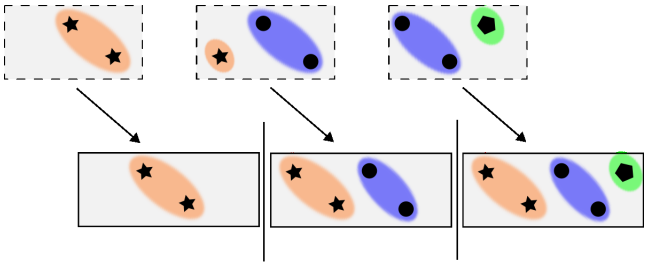
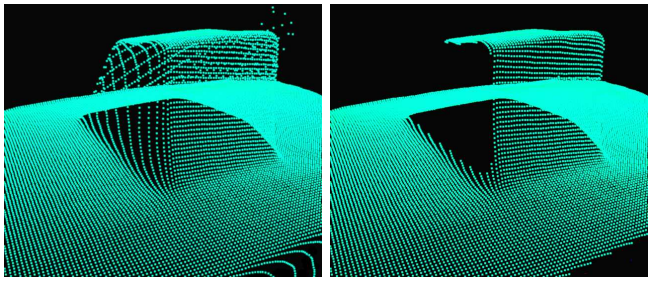


Fig. 2. The feature map is augmented with features from each frame. Upper row: Features detected in three consecutive frames. Lower row: Aggregated feature map at the first three time steps.



(a) (b)

Fig. 3. a) Unfiltered image of a cube. b) Application of Gaussian blurring and jump edge filtering.

pose estimation process effects runtime and precision. Since ToF cameras provide range and amplitude images we can apply algorithms that base on monochromatic images, e.g. KLT and SIFT, as well as the ICP algorithm which uses 3D data. We compare different setups for this task and propose the most suitable aggregation method or combination of these methods.

A. POSE ESTIMATION

There are two ways for obtaining a good and fast estimation of the performed egomotion in a first attempt. Transformations can be calculated from corresponding features in consecutive captures or current captures and an *aggregated feature map*. The latter includes distinctive features from all previous captures. Figure 2 shows how an aggregated feature map is created and subsequently augmented. The upper row (dashed rectangles) depicts the features of three consecutive frames. The bottom row shows the aggregated feature map at three different time steps augmented with new features from the current frame which could not be found in any of the previous frames. We use four different approaches to select the set of points for registration: scan line approximation extracting edge points from the range image, KLT and SIFT feature extraction to select features from the amplitude image, and no data reduction, i.e. using all points for registration. The registration process uses one of the following matching algorithms:

1) ICP on all possible feature sets (SIFT, KLT, Scanline, all): The 3D coordinates from the depth image of feature points found in the amplitude image are used to perform an ICP algorithm against the aggregated feature map until convergence.
2) SIFT descriptor matching: SIFT features are matched between consecutive frames using the SIFT feature vector and matching criterion. Related 3D coordinates of tracked features are taken to calculate pose changes (using singular value decomposition).
3) KLT image gradient based matching: KLT features are matched between consecutive frames using the gradient based KLT tracking in consecutive frames. Like for the SIFT descriptor matching, related 3D coordinates of tracked features are taken to calculate pose changes.

The latter two approaches constitute incremental approaches. A pose change estimation error is therefore expected to accumulate even if no ego-motion is performed.

B. 3D MAPPING

The 3D mapping of a scene is a three-stage process. First invalid data points are discarded by filtering. Second, the accumulated error is distributed among all captures. And finally, the map is enhanced with refinement filters. The first step is applied online, whereas the two following steps constitute post-processing.

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 [12] 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. This effect can be seen in Fig. 3. The left side of the cube was not in the range of vision from the camera's perspective and constitute a mismeasurement, which needs to be removed. Confining on jump edges, sufficient results are achieved with the following approach. From a set of 3D points $P = \{p_i \mid p_i \in \mathbb{R}^3, i = 1, \cdots, N_p\}$, jump edges can be selected by comparing the opposing angles $\xi_{i,n}$ of the triangle spanned by the focal point $\mathbf{f} = \mathbf{0}$, point $\mathbf{p}_i$ and its eight neighbors $P_n = \{\mathbf{p}_{i,n} \mid n = 1, \cdots, 8\}$ with a threshold ξ_{th} (cf. Fig. 4):

$$\xi_i = \max \arcsin \left(\frac{||\mathbf{p}_{i,n}||}{||\mathbf{p}_{i,n} - \mathbf{p}_i||} \sin \varphi \right), \tag{1}$$

$$J = \{p_i \mid \xi_i > \xi_{th}\}, \tag{2}$$

where φ is the apex angle between two neighboring pixels. The application of this filter can be seen in Fig. 3(b).

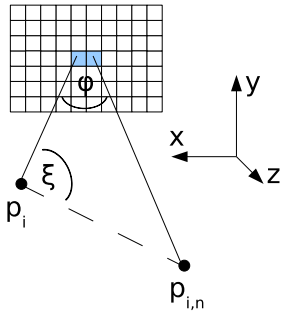


Fig. 4. Illustration of jump edge filter.

2) Error distribution: Due to the limited sensor precision a registration error accumulates from pairwise matching of consecutive frames. This error can be detected when the field of view is overlapping with an earlier one, i.e. a loop can be closed. Matching related captures determines the accumulated error, represented by a transformation matrix $\mathbf{T}_{(\mathbf{3\times 4})} = \begin{pmatrix}\mathbf{R}_{(\mathbf{3\times 3})} & \mathbf{t}^{\mathsf{T}}\end{pmatrix}$. From $\mathbf{R}$, the rotation axis $\mathbf{a}$ and rotation angle θ are determined. Rotation around axis $\mathbf{a}$ and translation along vector $\mathbf{t}$ are then equally distributed with a constant factor. See [14] for more details.

3) 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,

$$d(\mathbf{p}_i) = \frac{1}{k}\sum_{n=1}^{k} ||(\mathbf{p}_i - \mathbf{p}_{i,n})||, \tag{3}$$

$$S = \{\mathbf{p}_i \in P \mid d(\mathbf{p}_i) < d_{th}\}, \tag{4}$$

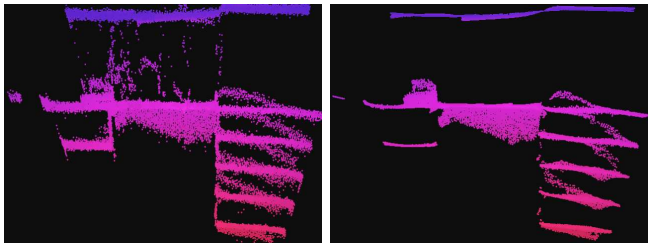
where d_{th} is a constant threshold. The set of point candidates Q for the resulting 3D map are all remaining points after filtering:

$$Q = \{\mathbf{q}_i | \mathbf{q}_i \in P \backslash (J \cup S)\}. \tag{5}$$

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. 5(a). A median filter can be applied to single captures. 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 this vector towards the detected surface. The result of this smoothing step can be seen in Fig. 5(b).

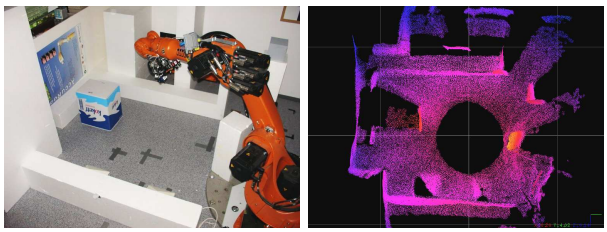
V. EXPERIMENTS AND RESULTS

The experiments demonstrate the robustness and the accuracy of our proposed registration procedure. We used the ToF-camera SR-3k with a resolution of 176×144 pixels. The standard deviation of distance measurements varies from 20 mm to 50 mm.



(a) (b)

Fig. 5. 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.



(a) (b)

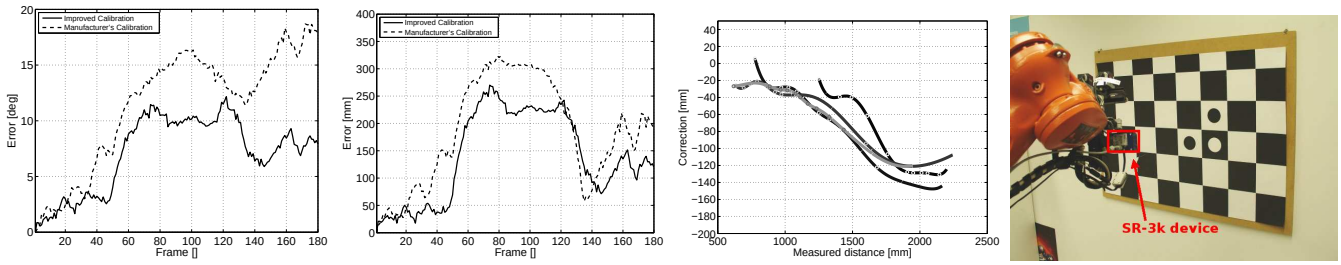
Fig. 6. a) Laboratory environment used for ground truth evaluation. The ToF camera is mounted on an industrial robot arm (KUKA KR 16). b) Bird's view of 3D map created in this environment (false color code relates distance to origin of coordinate system).

A. ACCURACY EVALUATION

First, the accuracy of ego-motion estimation and 3D map registration was investigated by applying the approach to a laboratory scene (room size $\approx 1800\,\mathrm{mm} \times 1800\,\mathrm{mm}$). It contained basic geometric styrofoam objects, e.g. cuboids. The ToF-camera was attached to an industrial robot (model KUKA KR 16). The scene (cf. Fig. 6) was circumferentially captured by moving the camera on a circular path with a diameter of 300 mm. In total, 180 captures were taken. For every capture the camera pose was computed from the extrinsic camera calibration and the robot's positioning system providing an accuracy of $1\,\mathrm{mm}$ and $0.1°$. This camera pose was assumed to be the ground truth (GT).

Since registration of measured point clouds in consecutive frames provides an increment of the performed camera motion, an accumulation of it relates to the camera pose w.r.t. to the starting frame. The deviation between estimated pose for each capture and computed GT is defined as the *Accumulated Error*.

Two test series were performed. In order to reduce the measurement noise the camera's automatic integration time controller was activated. It adjusted the integration time between $10000\,\mu\mathrm{s}$ and $14000\,\mu\mathrm{s}$. For the first test series we used the manufacturer's calibration. This calibration considers fixed pattern noise and latency-related errors. For the second test-series the camera was calibrated according to Fuchs et. al. [8] in order to additionally consider the distance- and amplitude-related error. Fig. 7(c) depicts the computed correction splines for an integration time of $12000\,\mu\mathrm{s}$. This parameter is a good approximation of the above mentioned working range of the camera in terms of integration time.



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

Fig. 7. Comparison of manufacturer's and improved calibration. Both calibrations were applied to test series which were captured with automatic adjustment of exposure time. In diagram (a) the accumulated rotational error and in diagram (b) the accumulated translational error is shown. Both diagrams demonstrate that the improved calibration provides better results. c) The resulting splines of the improved depth calibration. Every spline represents the depth correction for a special amplitude and distance interval. An overall sinusoidal curve of the distance-related error is apparent. The curves indicate that the distance correction increases with the amplitude interval (the higher the amplitude, the brighter the spline). d) The calibration was performed against a checkerboard pattern. The ToF camera is mounted on a KUKA robot arm for the calibration process.

In Fig. 7 the registration results of both test series are contrasted. We observed better results in test series based on the improved calibration. Here, the rotational and translational error accumulates to 7° and $120\,\mathrm{mm}$ contrary to 18° and $190\,\mathrm{mm}$. We suppose, that the manufacturer's calibration was performed for a different working range than the needed working range in our test series. Obviously, the improved calibration covers the working range in a better way.

Furthermore, Fig. 9 demonstrates the performance of all methods in estimating ego-motion, i.e. KLT, SIFT, ICP and their combination. With regard to rotational errors the SIFT method provided the best results. In terms of translational errors the rating is less clear. At a first glance, the final accumulated translational error using SIFT is lower (120 mm) than for the combined KLT-ICP method (214 mm). However, the average value for both methods is nearly identical with approximately 140 mm.

We assume that the accuracy of all methods depends on scene settings and performed movements. The SIFT method provided the best results for this laboratory scene while moving the camera on a circular path, but for the next experiment a different algorithm performed best.

B. 3D MAPPING OF LARGE ENVIRONMENTS

In a second experiment the robustness of the proposed mapping approach was tested. Thus, the mapping of an arbitrary environment - the robotic pavilion at the Fraunhofer Institute IAIS (cf. Fig. 1) - has been performed. The pavilion sizes $19.4\,\mathrm{m}$ in the longest distance. 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 any case. These mismeasurements occur especially when surfaces with specular reflectivity are present, e.g. mirrors, windows and metallic surfaces. This problem can be addressed with tracking, i.e. the detection when related measurements are shifted by an offset of maximum range. But it is more suitable to solve this problem on the sensoric level, e.g. the employment of multiple frequencies. After removing a few objects with specular reflectivity, a 3D map of the whole scenario could be generated (cf. Fig. 8).

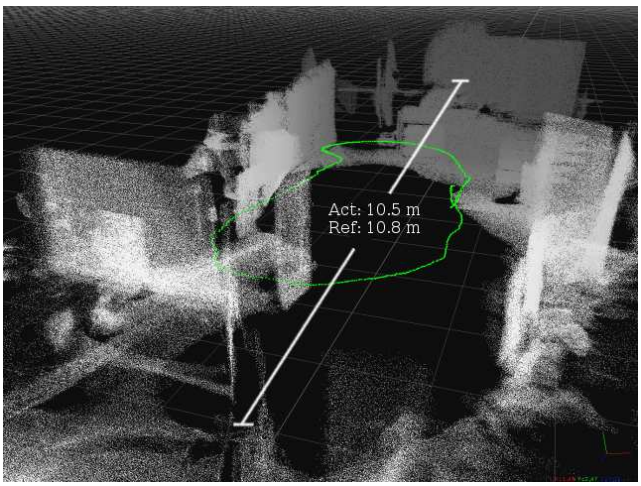
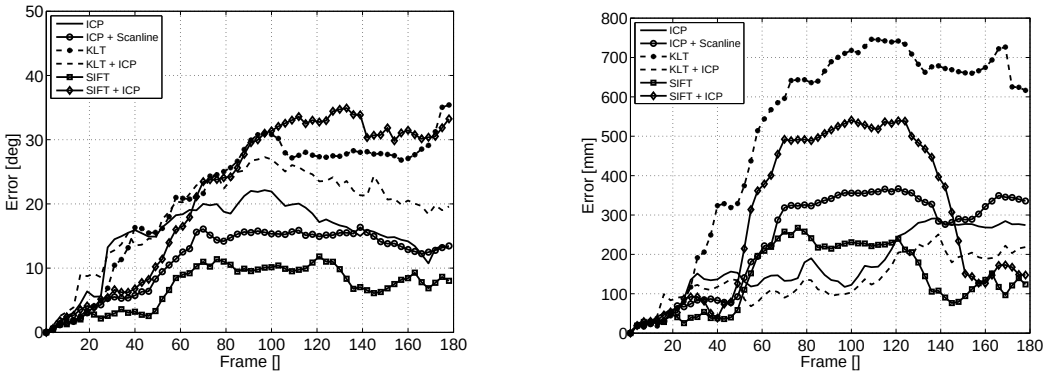


Fig. 8. 3D map of the whole robotic pavilion (perspective view of scene in Fig. 1). The trajectory (estimated poses of each frame) is drawn in green.

The above mentioned ego-motion estimation method has been applied for this test series, too. Ground truth data, in terms of an external positioning system, was not given. Therefore, two different measures evaluate the reconstruction precision of the resulting scene: the distance between two opposing walls and the pose estimation error after closing the loop. The best results were yielded applying the KLT-ICP method. Here, with start-to-end frame registration, the rotational error was determined with 2.2° and the translational error with $0.71\,\mathrm{m}$ in x-, $0.04\,\mathrm{m}$ in y- and $-0.21\,\mathrm{m}$ in z-direction. The second measurement addresses the isometry of the map. The calculated distance of two opposing walls from the 3D map compared with the distance determined with a measuring tape deviates only 0.3 m (Measuring tape vs. 3D cloud: 10.8 m / 10.5 m; cf. Fig. 8). A video showing the performance of map creation in this environment can be found in [13].



(a) (b)

Fig. 9. Comparison of pose estimation error employing different methods for registration. a) Rotational error. b) Translational error.

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 removement of mismeasurements has been achieved by the development of suitable filters. Third, robust pose estimation was achieved by visual odometry. Several algorithms have been investigated to achieve a maximum in precision and robustness, namely KLT, SIFT and ICP. Finally, the consistency of a 3D map has been enhanced in a refinement step, distributing accumulated errors and smoothing surfaces. The robustness of the entire approach has been demonstrated while registering 616 single captures obtained from a large environment.

Future work will concentrate on the improvement of calibration, e.g. the consideration of fluctuation in depth measurements caused by exposure time control, and the improvement of 3D map creation, e.g. by enhancing the semantic information.

REFERENCES

[1] H. Bay, T. Tuytelaars and L. Van Gool, SURF: Speeded Up Robust Features, In: *Proceedings of the European Conference on Computer Vision*, 2006.2

[2] P. Besl and N. McKay, A method for Registration of 3–D Shapes. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 239–256, Vol. 14, No. 2, February 1992.

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

[4] L. Clemente, A. Davison, I. Reid, J. Neira and Juan Domingo Tardós, Mapping Large Loops with a Single Hand-Held Camera, In: Proceedings of the Robotics: Science and Systems Conference, June, 2007.

[5] A. J. Davison and N. D. Molton, MonoSLAM: Real-Time Single Camera SLAM, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 29, No. 6, 2007.

[6] A. Diosi, A. Remazeilles, S. Segvic and F. Chaumette, Outdoor Visual Path Following Experiments. In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4265–4270, San Diego, CA, 2007.

[7] S. Fuchs and S. May, Calibration and registration for precise surface reconstruction with ToF cameras. In: *Proceedings of the Dynamic 3D Imaging Workshop in Conjunction with DAGM (Dyn3D)*, Vol. I, Heidelberg, Germany, September 2007.

[8] S. Fuchs and G. Hirzinger, Extrinsic and Depth Calibration of ToF-cameras. In. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, USA, June 2008.

[9] D. G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision*, 60, 2, pp. 91–110, 2004.

[10] X. Jiang and H. Bunke, Edge Detection in Range Images Based on Scan Line Approximation, *Computer Vision and Image Understanding*, Vol. 73, No. 2, pp. 183–199, February 1999.

[11] D. Holz, C. Loerken and Hartmut Surmann, Continuous 3D Sensing for Navigation and SLAM in Cluttered and Dynamic Environments, *Proceedings of the International Conference on Information Fusion*, 2008.

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

[13] 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.

[14] A. Nuechter, 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, August 2007.

[15] K. Ohno, T. Nomura and S. Tadokoro, Real-Time Robot Trajectory Estimation and 3D Map Construction using 3D Camera. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Beijing, China, October 2006.

[16] A. Sappa, A. Restrepo-Specht and M. Devy, Range Image Registration by using an Edge-Based Representation. In: *Proceedings of the International Symposium of Intelligent Robotic Systems (SIRS)*, pp. 167–176, Toulouse, France, 2001.

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

[18] J. Shi and C. Tomasi, Good Features to Track. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, Jun 1994.

[19] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann, K. Lau, C. Oakley, M. Palatucci, V. Pratt, P. Stang, S. Strohband, C. Dupont, L.-E. Jendrossek, C. Koelen, C. Markey, C. Rummel, J. van Niekerk, E. Jensen, P. Alessandrini, G. Bradski, B. Davies, S. Ettinger, A. Kaehler, A. Nefian and Pamela Mahoney, Stanley: The Robot that Won the DARPA Grand Challenge, *Journal of Field Robotics*, Vol. 23, No. 9, pp. 661–692, June 2006.

[20] O. Wulf, D. Lecking and Bernardo Wagner, Robust Self-Localization in Industrial Environments based on 3D Ceiling Structures, In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006.