Three-Dimensional Mapping with Time-of-Flight Cameras

Stefan May, David Droeschel, and Dirk Holz
Fraunhofer IAIS
Schloss Birlinghoven
53754 Sankt Augustin, Germany
e-mail: stefan_may@arcor.de

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

Ezio Malis
INRIA
2004, route des Lucioles
06902 Sophia-Antipolis, France
e-mail: ezio.malis@sohpia.inria.fr

Andreas Nüchter
Jacobs University Bremen
Campus Ring 1
28759 Bremen, Germany
e-mail: andreas@nuechti.de

Joachim Hertzberg
Knowledge-Based Systems Research Group
University of Osnabrück
49069 Osnabrück, Germany
e-mail: joachim.hertzberg@uos.de

Received 3 January 2009; accepted 10 August 2009

This article investigates the use of time-of-flight (ToF) cameras in mapping tasks for autonomous mobile robots, in particular in simultaneous localization and mapping

Journal of Field Robotics 26(11–12), 934–965 (2009)
Published online in Wiley InterScience (www.interscience.wiley.com). • DOI: 10.1002/rob.20321

(SLAM) tasks. Although ToF cameras are in principle an attractive type of sensor for three-dimensional (3D) mapping owing to their high rate of frames of 3D data, two features make them difficult as mapping sensors, namely, their restricted field of view and influences on the quality of range measurements by high dynamics in object reflectivity; in addition, currently available models suffer from poor data quality in a number of aspects. The paper first summarizes calibration and filtering approaches for improving the accuracy, precision, and robustness of ToF cameras independent of their intended usage. Then, several ego motion estimation approaches are applied or adapted, respectively, in order to provide a performance benchmark for registering ToF camera data. As a part of this, an extension to the iterative closest point algorithm has been developed that increases the robustness under restricted field of view and under larger displacements. Using an indoor environment, the paper provides results from SLAM experiments using these approaches in comparison. It turns out that the application of ToF cameras is feasible to SLAM tasks, although this type of sensor has a complex error characteristic.

1. 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, three-dimensional (3D) laser scanners and stereo camera systems are used mostly for these tasks due to their high measurement range and accuracy. Stereo vision requires the matching of corresponding points from two images to obtain depth information, which is directly provided by active sensors, such as laser scanners or ToF cameras. ToF cameras provide high frame rates while preserving a compact size. This feature has to be balanced with measurement accuracy and precision. Depending on external interfering factors (e.g., sunlight) and scene configurations, i.e., distances, surface orientations, and reflectivities, distance measurements from different perspectives of the same scene entail large fluctuations in accuracy and precision. These influences cause as well systematic errors as noise, both needing to be handled by the application. As a result, laser scanners are still the most common sensors used for 3D mapping purposes, e.g., Cole and Newman (2006), Holz, Lörken, and Surmann (2008), Nüchter, Lingemann, Hertzberg, and Surmann (2007), and Thrun et al. (2006).

In this article we present a mapping approach that deals with large variations in precision of distance measurements. We provide the underlying data in order to motivate further investigations in 3D mapping with ToF cameras. The proposed mapping approach is performed robustly with no additional sensory information about the ToF camera's ego motion. The approach comprises depth correction by employing an improved calibration, filtering of remaining inaccuracies, registration with respect to a common coordinate system, and map refinement including global relaxation—all combined yielding a precise and consistent 3D map.

The article is organized as follows: Section 2 elaborates 3D mapping approaches and applications related to ToF cameras. Section 3 describes ToF camera errors caused by external interfering factors and the employed depth correction method. In Section 4 our mapping approach including 3D pose estimation, error handling, and mapping is represented. Section 5 illustrates experimental results that support our accentuation of employing real-time-capable ToF sensors to pose estimation and mapping tasks. Finally, Section 6 concludes with an outlook on future work.

2. RELATED WORK

One of the first applications in robotics considering ToF cameras as an alternative to laser scanning, stereo or monocular vision, was presented in 2004. Weingarten, Grüner, and Siegwart (2004) evaluated a SwissRanger SR-2 device in terms of basic obstacle avoidance and local path planning capabilities. The ToF camera was calibrated photogrammetrically to determine parameters for the perspective projection to the image plane. Additionally, an empirically determined depth correction method employing a unique distance scaling and offset value was proposed. Navigation and path planning could be performed robustly based on the provided data.

In 2005, Sheh et al. presented an application based on the same ToF camera at the RoboCup in

Osaka (Sheh, Kadous, & Sammut, 2006). The data take involved the rotation by a pan-tilt unit in order to obtain an almost circumferential view. Generally, this involved taking 10 range images at intervals of 36 deg, while stopping at each location long enough to avoid motion blurring. The whole process took approximately 20 s for each data take. Their mapping procedure was assisted by a human operator, who had to identify landmarks.

In 2006, Ohno et al. used a SwissRanger SR-2 camera for estimating a robot's trajectory and reconstructing the surface of the environment (Ohno, Nomura, & Tadokoro, 2006). 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% in translation and up to 17% in rotation, respectively.

The aforementioned approaches show that applications with ToF cameras have to face two problems. First, registration of range images provided by ToF cameras is more difficult than registration of laser range finder data due to the lower measurement accuracy. That is why some groups investigated the error modeling and calibration of these devices. Because ToF cameras provide monochromatic reflectance images and are based on a pinhole camera model, photogrammetic calibration is feasible in order to determine intrinsic and extrinsic parameters. Additionally, a calibration of the provided range data has to be performed. Currently, only a few authors have proposed a calibration method considering jointly reflectance and range data. Lindner and Kolb as well as Kahlmann et al. estimated intrinsic parameters of a ToF camera using the reflectance image of a checkerboard (Lindner & Kolb, 2006) and a planar test field with near-infrared (NIR) LEDs, respectively (Kahlmann, Remondino, & Ingensand, 2006). A per-pixel precision of at least 10 mm was achieved.

The second problem concerns the smaller field of view of a ToF camera compared to 3D laser scanners. It has to be ensured that the depth image is geometrically unambiguous in terms of registration to a previous depth image. Sheh et al. (2006) handled this problem by employing a pan-tilt unit for circumferential data take while stationary, which entails a slow data acquisition rate.

In contrast to a pure range image–based registration, features obtained from the reflectance image can contribute to the stability as long as the scene comprises a certain degree of texturedness. The first approaches employing the monochromatic reflectance image for feature tracking were proposed after 2007. The main problem is the low resolution of ToF cameras, providing only a few features for certain scenes. Additionally, the measurable depth value at the feature's location might be of low accuracy and precision (May, Pervölz, & Surmann, 2007). Swadzba, Liu, Penne, Jesorsky, and Kompe (2007) performed only coarse registration on depth data related to features in the reflectance image, whereas the fine registration has been calculated by the use of the whole depth image. Therefore, a few researchers proposed fusion with other sensors. Prusak, Melnychuk, Roth, Schiller, and Koch (2007) 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. A rough registration was performed on the circumferential view of the spherical camera. The registration was then refined based on the range image while employing the initial guess from the coarse registration. Huhle, Jenke, and Straßer (2007) also used a joint approach of feature tracking and range image registration. They fused data from a high-resolution camera with a ToF device to obtain colored 3D point clouds. Feature determination performed much better on the high-resolution color images than on the low-resolution monochromatic reflectance images of the ToF device.

A comparison of tracking based on data acquired from a ToF sensor with data provided by a high-resolution color camera has been provided by Sabeti, Parvizi, and Wu (2008). They concluded that high-resolution color sensors are more suitable for outdoor applications and purposes that require the detection of fine details, whereas ToF sensors are more appropriate for object tracking that requires information about the distance between object and camera. The fusion of both sensors is therefore a profitable approach. Additionally, it can be expected that the resolution of ToF cameras will increase in the future.

The approach described in this article relies exclusively on ToF camera images. We present an investigation of how precise scene reconstruction is currently possible based on ToF camera data, comprising an evaluation of calibration and filtering as well as the application of different registration approaches. To our knowledge, some of the registration methods are applied to these special sensor types for the first time.

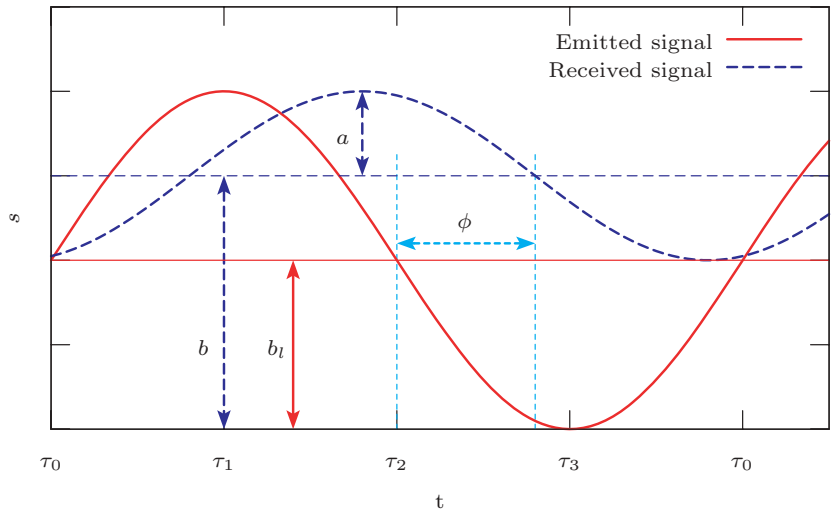


Figure 1. Received sinusoidally modulated input signal, sampled with four sampling points per modulation period T.

3. 3D TOF RANGING

This section introduces the reader to the emerging ToF camera technology. To start, the measurement principle is described. Then, the occurring error sources are explained. And finally, appropriate methods for error handling are presented. We constrain the explanations on depth measuring and calibration to the sinusoidal modulation principle. The ToF camera employed in this study, a SwissRanger SR-3k, uses this principle. The results reported here have previously been described in Fuchs and Hirzinger (2008) and Fuchs and May (2007). We are summarizing them here to keep this article self-sufficient.

3.1. Measurement Principle

ToF cameras using signal modulation are active sensors that measure distances based on the phase-shift principle. The observed scene is illuminated with modulated NIR light, whereby the modulation signal is assumed to be sinusoidal with frequencies in the order of some megahertz. The reflected light is projected onto a charge-coupled device (CCD)- or complementary metal-oxide-semiconductor (CMOS)-sensor or a combined technology. There, the phase shift, which is proportional with the covered distance, is measured in parallel within each pixel.

Let $S_i(t) = \{s_i(t_0), s_i(t_1), \ldots, s_i(t_m) | i = 1, \ldots, n\}$ be $m+1$ measurements of an optical input signal taken at each of n pixel locations in the image array. Further let $A = \{a_i | i = 1, \ldots, n\}$ be the set of amplitude data and $B = \{b_i | i = 1, \ldots, n\}$ the set of intensity (offset) data. From the reflected sinusoidal light four measurements $s_i(\tau_0)$, $s_i(\tau_1)$, $s_i(\tau_2)$, and $s_i(\tau_3)$ at 0, 90, 180, and 270 deg of the phase are taken each period $T = 1/f_m$. A pixel's phase shift ϕ_i, amplitude a_i and intensity b_i (i.e., the background light) can be calculated by (Lange, 2000) (see Figure 1)

$$\phi_i = \arctan\left[\frac{s_i(\tau_0) - s_i(\tau_2)}{s_i(\tau_1) - s_i(\tau_3)}\right], \tag{1}$$

$$a_i = \frac{\sqrt{[s_i(\tau_0) - s_i(\tau_2)] + [s_i(\tau_1) - s_i(\tau_3)]}}{2}, \tag{2}$$

$$b_i = \frac{\sum_{j=0}^{3} s_i(\tau_j)}{4}. \tag{3}$$

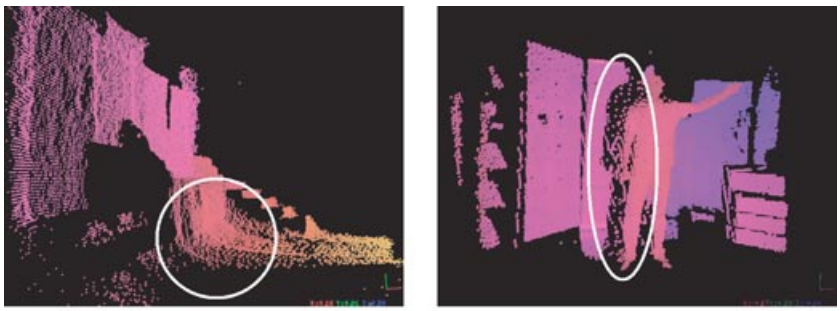
The distance measurements $D = \{d_i | i = 1, \ldots, n\}$ between image array and object is then determined by

$$d_i = \frac{\lambda_m}{2}\frac{\phi_i}{2\pi}, \tag{4}$$

where λ_m is the wavelength of the modulation signal.

3.2. Error Model

The performance of distance measurements with ToF cameras is limited by a number of error sources. In the following, the most important error sources are



(a) (b)

Figure 2. Interreflection effects due to multimodal reflections: (a) Corners or hollows appear rounded off. (b) Occluding shapes have a smooth transition.

explained. More detailed information on ToF error sources can be found in Lange (2000). Some error sources are predefined by the design of the hardware or its physical properties and cannot be corrected by calibration. The influence of these effects can be either identified and discarded or estimated while making assumptions about the configuration of the measured scene. The following explanations relate to them as *random errors*. In contrast, *systematic errors* comprise all errors that can be identified and corrected due to their systematic occurrence.

3.2.1. Random Errors

Noise limits the performance of ToF cameras and can be subdivided into three different classes: *photocharge conversion noise*, *quantization noise*, and *electronic shot noise* (also called *quantum noise*) (Lange, 2000). *Electronic shot noise* is the most dominating noise source and cannot be suppressed. It describes the statistical Poisson-distributed nature of the arrival process of photons and the generation process of electron–hole pairs. It limits the theoretically reachable signal-to-noise ratio (SNR) and the accuracy involved. Because the measurement principle is based on integrating discharged electrons from incoming light, the optical power influences the reachable precision. These electrons are collected within a conversion capacity, which can result in oversaturation if the integration time is too high (Lange, 2000). Recent cameras use burst modes to increase the power output for short intervals at the same energy level over time, which yields a better SNR while complying with eye-safety regulations.

Interreflection (also called multiple-ways reflection) occurs due to occlusions and in concave objects, e.g., corners or hollows. The signal can take multiple ways through reflection before returning to the receiver. For that the remitted NIR signal is a superposition of NIR light that has traveled a different distance, called *multimodal reflection*. Hollows and corners appear rounded off and occluding shapes with a smooth transition (see Figure 2). This error arises as a consequence of a diverging measurement volume. For instance, the SwissRanger SR-3k device has a 0.27×0.27 deg field of view for each pixel. The width of a pixel's measurement volume therefore depends on the distance (e.g., 1 m: 4.7 mm/5 m: 23.5 mm).

Light scattering is a general physical process that forces light to deviate from a straight trajectory by one or more localized nonuniformities. This effect occurs in the lenses of an optical device and depends on the amount of incident light. For ToF cameras the secondary reflections of light reflected by near bright objects superpose the measurements from the background, which appear closer, in consequence (Mure-Dubois & Hügli, 2007).

3.2.2. Systematic Errors

In contrast to the aforementioned random errors, systematic errors do not tend to have a null arithmetic mean when a measurement is repeated several times. Usually, systematic errors either cause a constant bias or relate to the measured value or to the value of a different quantity.

Circular distance errors (also called distance-related or wiggling errors) stem from the emitted

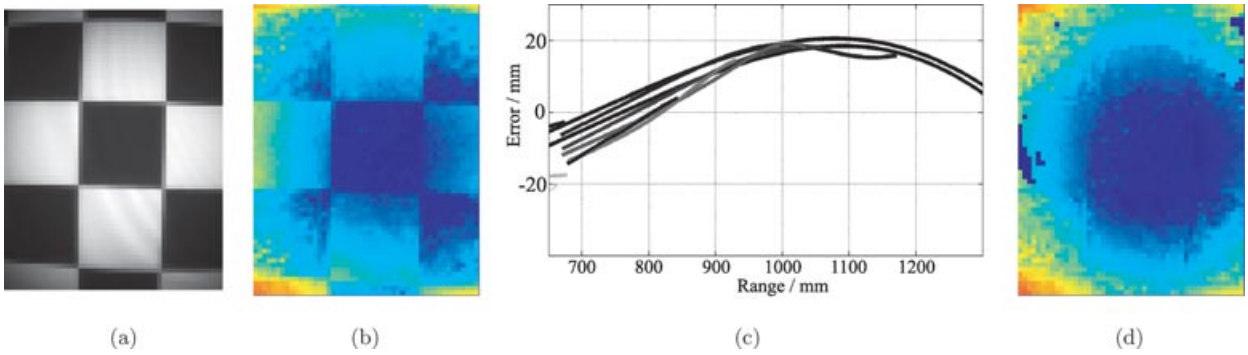


Figure 3. Amplitude-related error. (a) Amplitude image of checkerboard taken with an IFM O3D100 device. The inhomogeneous illumination can clearly be seen (decreasing illumination in the peripheral area). (b) The related miscolored depth image. Ideally, the checkerboard pattern should not be noticeable because it is a plane. But due to the amplitude-related error, dark regions are somewhat nearer to the observer. (c) The diagram plots the identified amplitude- and distance-related error by a calibration step. (d) The measurement is corrected, and the checkerboard pattern is no longer noticeable.

square-wave signal additionally distorted by the asymmetric response of a NIR LED signal. The result is a nonharmonic sinusoidal NIR signal. Because a harmonic sinusoidal illumination is the basic assumption in the principle of modulation interferometry, the computed phase delay and distance, respectively, are inaccurate. The amplitude of the circular distance error for the SwissRanger SR-3k was determined to be between −150 and 10 mm with a wavelength of 2,000 mm (see Section 3.3.2).

Amplitude-related errors are caused by nonlinearities of the pixel's electronic components. Incident photons induce electrons in the capacitances of the photosensitive layer. These voltages are read out and amplified before signal processing and digitalization. Both charging of capacitances and amplification are nonlinear. The arrival of different numbers of photons at a constant distance results in different distance measurements (see Figure 3).

Inhomogeneous image illumination depends on the configuration of the LEDs, the optics, and the camera's field of view. The illumination decreases in the peripheral area of the image as a result of two effects, the inhomogeneous scene illumination through the sensor's active light emitters and the vignetting effect induced by the sensor's optics [see Figure 3(a)]. This entails amplitude-related errors and different SNRs (contribution of systematic and nonsystematic errors in different proportions).

Fixed-pattern noise (FPN) has to be considered in two ways. First, each pixel has an individual characteristic due to an imperfect manufacturing process and different material properties in each CMOS gate, which yields a pixel-individual fixed measurement offset. Second, the triggering of each pixel depends on the position on the chip and its distance to the signal generator. Pixels aligned in rows or columns are connected in series and cause a gradual phase shift. As a consequence the measurement is distorted by an increasing offset. This offset is also called fixed-pattern phase noise. In the strict sense the term "noise" is not correct for this kind of signal propagation delay. But due to the "fixed" occurrence it is often ranked as such.

3.3. Error Handling

The concurrence of the above-mentioned error sources significantly distorts the measurements. Without any consideration of these errors, the raw depth measurements do not allow sophisticated applications such as mapping. The following sections present some methods that handle the random errors as well as the systematic errors by filtering and by calibration.

3.3.1. Filtering

The accuracy of measurements in unknown scenes varies considerably, due to the above-mentioned effects. It can be rated with respect to the amount of light returning to the sensor (amplitude data) and allows at least the applicability of filtering. But a pure amplitude-based filtering is disadvantageous in

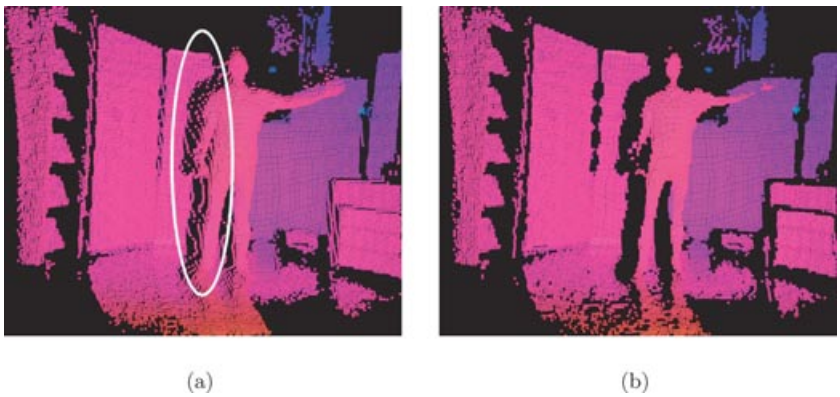


Figure 4. Jump edges occurring at the boundary of occluding shapes. (a) Unfiltered scene: A transition between the person in the foreground and the wall in the background can clearly be seen. (b) Filtered scene: Jump edges are reliably removed.

terms of narrowing the field of view. Owing to the inhomogeneous image illumination, measurements in the peripheral part of the field of view would be discarded primarily. However, such a filter has a practical use for the map generation process, in which only measurement points providing high amplitudes are added to a 3D map.

Among the nonsystematic errors, we further focus on the occurrence of so-called *jump edges*, i.e., when the transition from one to another shape appears to be 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 (see Interreflection effect in Section 3.2.1). This effect can be seen in Figure 4(a) between the person in the foreground and the wall in the background. The appearance depends on the perspective view and induces a large error when considered for the estimation of motion (see Figure 5).

Jump edge filtering. Several approaches have been proposed to overcome the identification and/or correction of these mismeasurements. Pathak, Birk, and Poppinga (2008) proposed a Gaussian analysis for correcting multimodal measurements. The drawback of their method is the integration over 100 images for each frame. This reduces significantly the frame rate. Additionally, the process of estimating the parameters for the Gaussian fitting takes 8.5 min

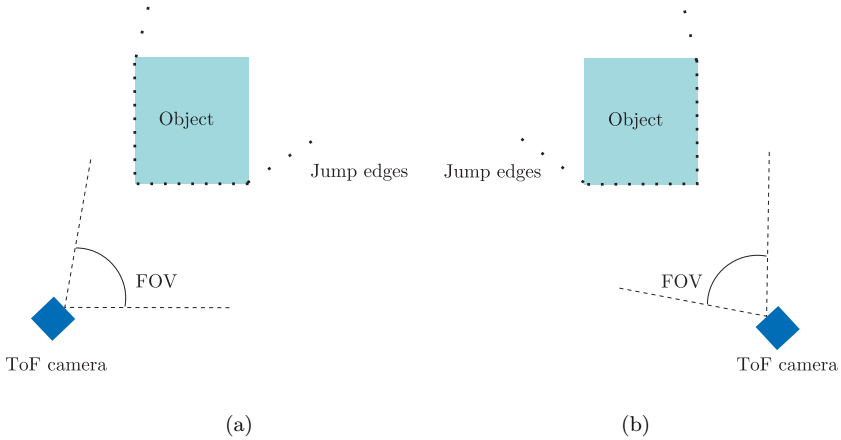


Figure 5. Jump edge appearance depends on the perspective view. The matching of both point clouds can easily result in a wrong registration.

per frame for a Matlab implementation. This is far from being applicable in real time.

Sappa, Restrepo-Specht, and Devy (2001) presented an approach that is aimed at identifying and classifying edges. It uses the fitting of polynomial terms to approximate scan lines. These scan lines are connected at edge points. The strength of this approach is that it also performs a classification of edges in jump edges and *crease edges*: "Crease edges are those points in which a discontinuity in the surface orientation appears," e.g., in corners or hollows.

Focusing on the identification of jump edges, sufficient results are achieved with local neighborhood relations. From a set of 3D points $P = \{\boldsymbol{p}_i \in \mathbb{R}^3 | i = 1, \ldots, N_p\}$, jump edges J can be selected by comparing the opposing angles $\theta_{i,n}$ of the triangle spanned by the focal point $\boldsymbol{f} = \boldsymbol{0}$, point $\boldsymbol{p}_i$ and its eight neighbors $P_n = \{\boldsymbol{p}_{i,n} | i = 1, \ldots, N_p : n = 1, \ldots, 8\}$ with a threshold θ_{th}:

$$\theta_i = \max \arcsin \left(\frac{\|\boldsymbol{p}_{i,n}\|}{\|\boldsymbol{p}_{i,n} - \boldsymbol{p}_i\|} \sin \varphi \right), \tag{5}$$

$$J = \{\boldsymbol{p}_i | \theta_i > \theta_{th}\}, \tag{6}$$

where φ is the apex angle between two neighboring pixels.

The application of this filter can be seen in Figure 4(b). 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 applied first. Figure 6(a) depicts a scene

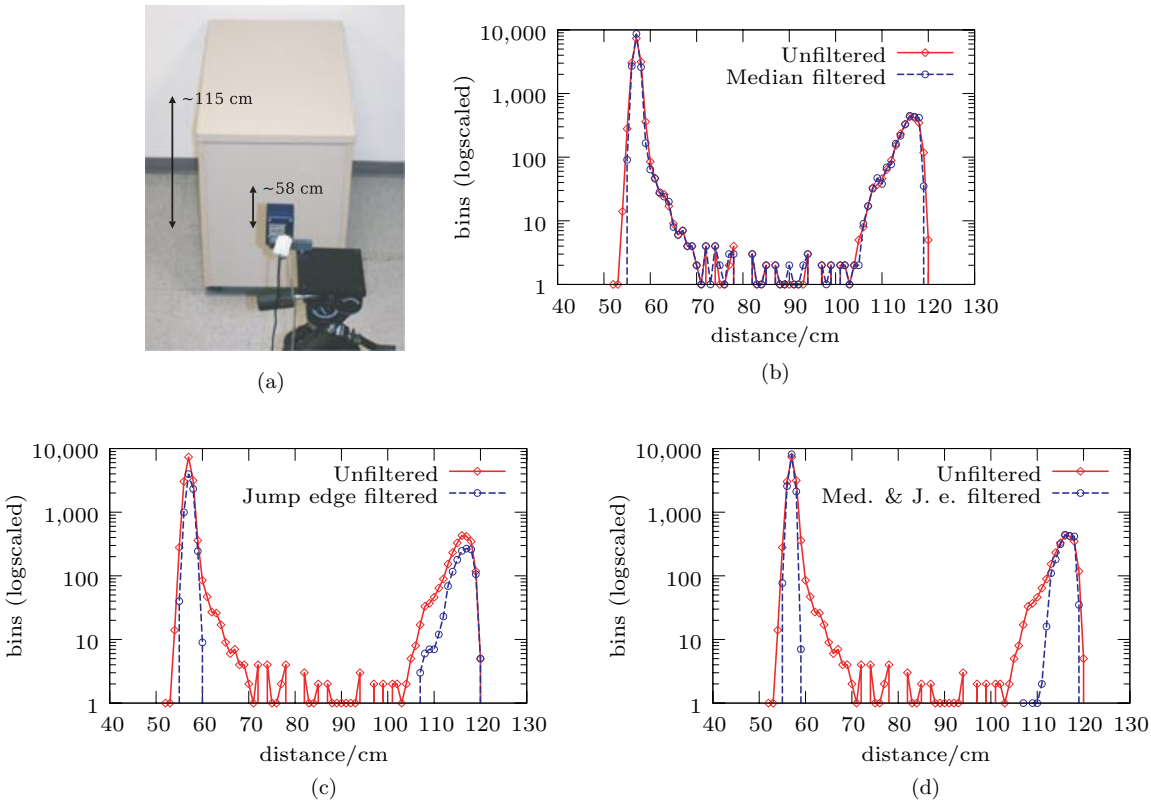
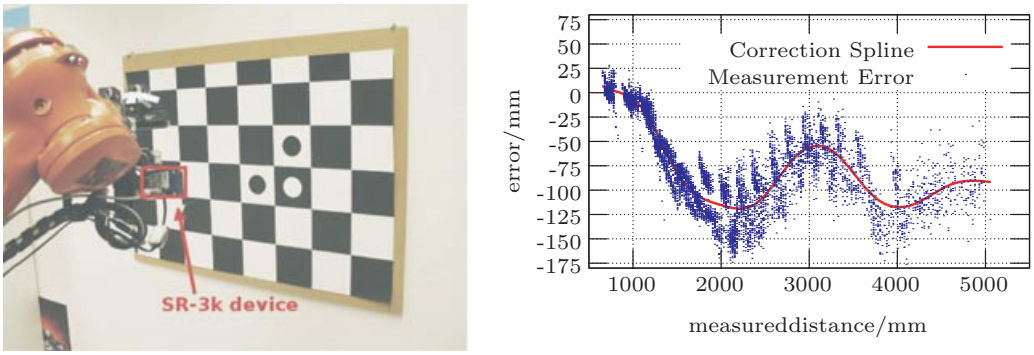


Figure 6. (a) Distance measurement of bimodal scene (two planar objects with different z distances, i.e., background and filing cabinet). (b)–(d) Histograms showing the influence of different filters. (b) The bimodal measurement is represented by two peaks at ≈58 and ≈118 cm. Measurement values deviating from these values indicate noisy or wrong measurements. The application of median filtering reduces noise on surfaces but does not have an impact on wrong measurement data (jump edges). (c) The jump edge filter is sensitive to noise and besides jump edges discards too many valid points located at surfaces (cf. the reduction of valid points at both peaks). (d) The subsequent application of both filters properly determines jump edge measurements without discarding surface points.



(a) (b)

Figure 7. (a) The improved calibration is performed against a checkerboard pattern (image taken at the DLR laboratory). Pattern sizes 1,190 × 840 mm. (b) The revealed depth measurement error varies from −150 to 10 mm. For the most part the camera provides too-large-distance data. The spline approximates this error for subsequent correction purposes.

in which the z coordinate of measurements should be bimodal, i.e., only two z distances are measured over the entire scene. The background has a z distance of ≈115 cm and the front surface of the filing cabinet of ≈58 cm, both determined with a measuring tape. In the z-value histograms depicted in Figure 6 two peaks can be observed. Different z distances result mainly from jump edges. Jump edge filtering without noise reduction discards too many valid points on surfaces. Both peaks are considerably reduced.

The noise that overlies range measurements follows a Gaussian distribution (Gabriel, 2006), whereby a median filter is preferable in terms of preserving edges. The subsequent application of median and jump edge filtering achieved reliable results. The ratio between jump edge points and surface points in a range image is small (mostly below 5%). With respect to the computational complexity, experiments in Section 5 are confined to discarding jump edge points instead of correcting them.

3.3.2. ToF Camera Calibration

The projection onto the image plane of ToF cameras is described by the pinhole camera model. Thus, a common photogrammetric calibration is performed in order to determine the intrinsic parameters. These parameters characterize aberrations caused by the optics in, for instance, lens distortion, skew, focal length, and optical center.

Furthermore, the above-mentioned systematic distance errors are identified in a depth calibration step according to Fuchs and May (2007) (extended model in Fuchs & Hirzinger, 2008), which is called *improved calibration* in the course of this article. The depth calibration is combined with the intrinsic calibration, and hence it is easy to apply and achieves considerable improvements in terms of accuracy.

Initially, a number of amplitude and depth images of a checkerboard are captured from different points of views. In the first step the amplitude images are used for an intrinsic camera calibration. In the second step the amplitude and depth images are used to identify the depth correction model, comprising circular distance errors, amplitude-related errors and fixed-pattern phase noise.

Figure 7(a) illustrates the ToF camera calibration using an industrial robot. The robot provides external positioning data, which sustain the estimation of the calibration pattern pose. This pose is needed as ground truth for depth calibration. If no position data are available, the pattern pose can also be estimated from every single capture but is more error-prone. Figure 7(b) plots the identified distance-related error. As a result the calibration allows for an accuracy of at least 3 mm (Fuchs & Hirzinger, 2008).

4. 3D MAPPING

Thrun et al. (2005) count SLAM among the "core competencies of truly autonomous robots." The goal in SLAM is to generate a map of previously unknown territory without external localization aids. SLAM is

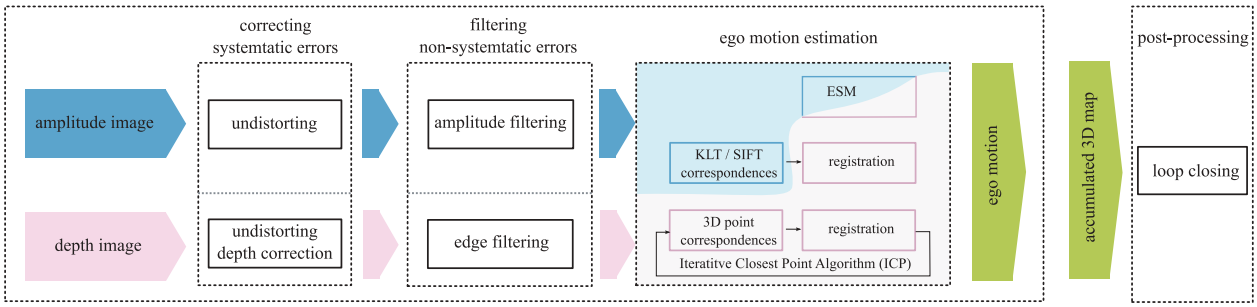


Figure 8. The diagram outlines the 3D mapping process. Systematic and nonsystematic errors are treated by undistortion, depth correction, and appropriate filtering. Several approaches are used to estimate the ToF camera's ego motion. The hybrid ESM algorithm applies both amplitude and depth data. KLT and SIFT are implemented in a two-stage method, in which first the amplitude data are used to track features and second, the associated depth values are registered. The ICP employs only depth images and iteratively registers consecutive point clouds until the algorithm converges. Finally, the estimated ego motions are accumulated for 3D mapping purposes.

challenging because it is a chicken-and-egg problem: To build a truthful and consistent map, one would have to know the accurate pose; however, to know the pose with no external localization aids, one would need to have an accurate map for localizing.

This section presents some approaches for building 3D maps out of ToF camera data. Obviously, if some ToF camera data sets are "somehow" assembled into a truthful and consistent 3D map, then each and every camera pose is also localized with respect to the map. However, most of the methods coming next do not handle this localization explicitly. Therefore, we are using the term "mapping" rather than "SLAM." The last part of the section is an exception to this principle: We will use past ToF camera poses explicitly here in order to reduce accumulated registration errors. Unsurprisingly, the method used there is borrowed from a SLAM loop-closing method.

Figure 8 outlines the complete mapping process. First, in a preprocessing step, the systematic and nonsystematic errors of the depth and amplitude images are considered. Second, the consecutive point clouds are registered to each other, whereas the displacement of these consecutive point clouds is synonymous with the ToF camera's ego motion. Four different methods have been implemented. A common and straightforward approach in the context of point cloud registration is the iterative closest point (ICP) algorithm. The ICP algorithm employed here is exclusively based on range data. A more sophisticated approach also comprises the amplitude data. Two of the most frequently used feature tracking techniques, namely the *scale invariant feature transform* (SIFT) matching and *Kanade–Lucas–Tomasi* (KLT) feature tracking, are applied for feature tracking in consecutive images. The 3D registration can be performed on the basis of these feature correspondences and their associated depth values. Efficient second-order approximation method (ESM) tracking is a hybrid approach that considers depth and intensity data jointly. Finally, the created map is refined by means of loop-closing techniques.

4.1. ICP-Based Ego Motion Estimation

The ICP algorithm, which was developed independently by Besl and McKay (1992), Chen and Medioni (1991), and Zhang (1992), is the most popular approach for range image registration. It aims at obtaining an accurate solution for the alignment of two point clouds by means of minimizing distances between point correspondences. Corresponding points are obtained by a nearest neighbor search in two point sets. Let $M = \{\boldsymbol{m}_j | j = 1 \dots N_m\}$ be a set of points from a previous data take, called *model point set*, and $G = \{\boldsymbol{g}_i | i = 1 \dots N_g\}$ a set of points from the most recent data take, called *scene point set*. Then, every point in the scene point set is taken into consideration to search for its closest point in the model point set:

$$\boldsymbol{m}_k = \underset{j=1\dots N_m}{\operatorname{argmin}} \|\boldsymbol{g}_k - \boldsymbol{m}_j\|. \tag{7}$$

The ICP algorithm aims at finding a rigid transformation, comprising a rotation $\mathbf{R}$ and a translation $\boldsymbol{t}$, by performing an iterative least-square minimization. In each iteration step the transformation minimizing the mean squared error function

$$E(\mathbf{R}, \boldsymbol{t}) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_m} \omega_{i,j} \|(\mathbf{R}\boldsymbol{g}_i + \boldsymbol{t}) - \boldsymbol{m}_j\|^2 \qquad (8)$$

is to be determined, where $\omega_{i,j}$ denotes weights that are of value 1 if (g_i, m_j) are found to be corresponding and 0 otherwise.

Expressing the corresponding point pairs as a set of N tuples $\{(\boldsymbol{g}_k, \boldsymbol{m}_k)|k = 1 \ldots N\}$, the error function can be reduced to

$$E(\mathbf{R}, \boldsymbol{t}) = \sum_{k=1}^{N} \|(\mathbf{R}\boldsymbol{g}_k + \boldsymbol{t}) - \boldsymbol{m}_k\|^2. \qquad (9)$$

There are four known algorithms calculating the closed-form solution for $\mathbf{R}$ and $\boldsymbol{t}$. An explanation of these algorithms can be found in Lorusso, Eggert, and Fisher (1995). Rusinkiewicz and Levoy (2001) give a detailed analysis of efficient variants of the ICP, discussing the closed-form solutions, point-to-point vs. point-to-plane metrics, and different point rejection rules.

Point correspondence rejection has a strong impact on the registration result. The original formulation of the ICP approach assumes that the scene point set is completely covered by the model point set (Besl & McKay, 1992). In the case that the scene point set includes points that are not part of the model point set (from a nonoverlapping area), wrong correspondences are assigned for these points that might distort the registration result (Fusiello, Castellani, Ronchetti, & Murino, 2002). The simplest solution is the employment of a distance threshold. Corresponding tuples are rejected if their Euclidean distance exceeds this value. Several strategies are possible to determine suitable thresholds, e.g., a gradually decreasing threshold with respect to the iteration step. In general, these thresholds increase the registration performance on partially overlapping point clouds significantly. For convenience, the original formulation of the ICP approach including a distance threshold is called the *vanilla ICP* approach in the following.

Many extensions to the vanilla ICP approach have been published addressing the determination of valid point correspondences from overlapping parts. Prusak et al. (2007) employed the *trimmed ICP (TrICP)* approach (Chetverikov, Svirko, Stepanov, & Krsek, 2002). It employs a parameter representing the degree of overlap, i.e., the number of corresponding points n. Only the first n correspondence pairs ordered ascending by point-to-point distance are considered for estimating the optimal transformation. Prusak et al. employed a coarse registration step performed on data of a second sensor, a spherical camera, in order to obtain an estimate for the degree of overlap.

Several approaches have been proposed to overcome the registration problem with unknown degree of overlap. Fusiello et al. (2002) employed the X84 rejection rule, which uses robust estimates for location and scale of a corrupted Gaussian distribution. It aims at estimating a suitable rejection threshold concerning the distance distribution between corresponding points. Niemann, Zinper, and Schmidt (2003) proposed a rejection rule that considers multiple point assignments *(picky ICP algorithm)*. If multiple points from the scene point set are assigned to the same corresponding model point, only the scene point with the nearest distance is accepted. The other pairs are rejected. Pajdla and Van Gool (1995) proposed the inclusion of a reciprocal rejection rule *(iterative closest reciprocal point algorithm; ICRP)*. For a corresponding point pair $(\boldsymbol{m}_k, \boldsymbol{g}_k)$, which has been determined by searching the nearest neighbor of $\boldsymbol{g}_k$ in M, the search is reversed subsequently, i.e., for $\boldsymbol{m}_k$ the nearest neighbor in G is determined. This need not be the same scene point and is denoted with $\boldsymbol{g}'_k$. The point correspondence is rejected if $\boldsymbol{g}_k$ and $\boldsymbol{g}'_k$ have a distance larger than a certain threshold. A disadvantage of the ICRP approach is the higher computational effort, because the nearest neighbor search, which is the most time-consuming task, is performed twice as much as for all other approaches.

The method proposed here to overcome an unknown degree of overlap stems from 3D computer graphics and is called *frustum culling* (Lengyel, 2000). A frustum defines the volume that has been in the range of vision while acquiring the model point set [see Figure 9(a)]. Luck, Little, and Hoff (2000) used frustum culling for prefiltering based on an initial pose estimate. The iterative registration process was then performed on the reduced data set. On the contrary, we employ no initial pose estimate. Therefore, the frustum culling is embedded in the iteration process by employing the pose estimate of the previous iteration step. Scene points outside of the model

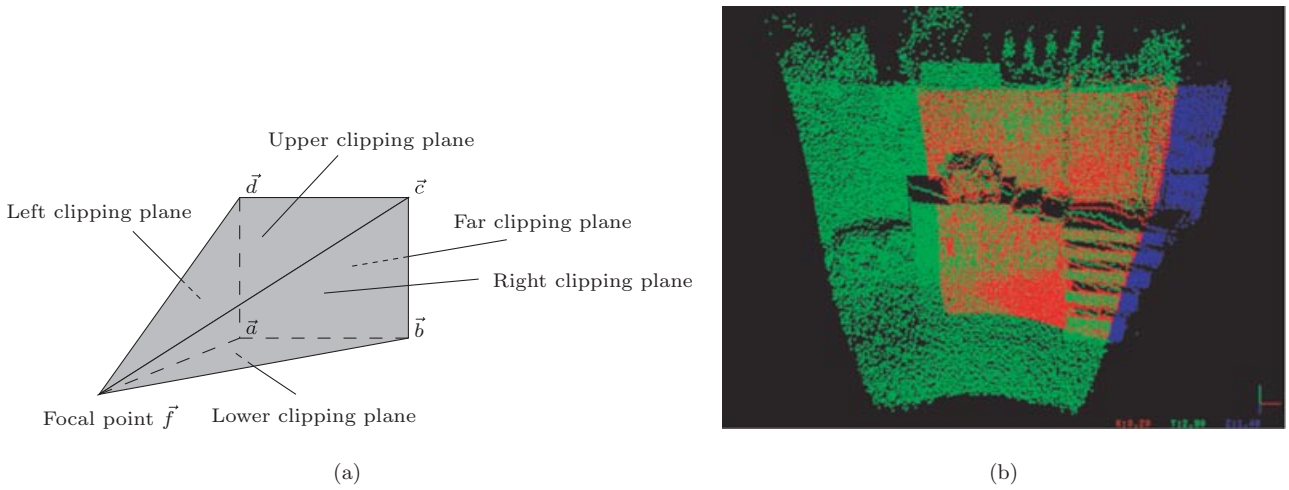


Figure 9. Frustum culling technique. (a) Definition of clipping planes. The clipping planes are defined by the model data set. (b) Clipping of scene points (red/blue) outside of a model (green) frustum. Blue points are not considered for the nearest neighbor search.

frustum are filtered by testing against clipping planes before performing nearest neighbor searching. This extension is called the *frustum ICP* approach and is described in the following.

Let $\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}\}$ be the vectors from the origin to the four edge points of the far clipping plane and $\boldsymbol{f} = \boldsymbol{0}$ be the focal point. The lateral clipping planes are then spanned by the focal point and two edge points.

The normal vectors $\boldsymbol{n} = (n_x, n_y, n_z)^T$ of each lateral clipping plane

$$\begin{aligned} \boldsymbol{n}_b &= \boldsymbol{b} \times \boldsymbol{a}, \\ \boldsymbol{n}_u &= \boldsymbol{c} \times \boldsymbol{d}, \\ \boldsymbol{n}_r &= \boldsymbol{b} \times \boldsymbol{c}, \\ \boldsymbol{n}_l &= \boldsymbol{d} \times \boldsymbol{a} \end{aligned} \tag{10}$$

can be used to check whether a point $\boldsymbol{x} = (x, y, z)^T$ is inside the frustum by

$$x n_x + z n_z < 0 \tag{11}$$

for the left and right clipping plane (with $\boldsymbol{n}_l$ and $\boldsymbol{n}_r$) and

$$y n_y + z n_z < 0 \tag{12}$$

for the upper and lower clipping plane (with $\boldsymbol{n}_u$ and $\boldsymbol{n}_b$).

This test is nonparametric and iteratively removes scene points from nonoverlapping 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 especially the problem of a restricted field of view [see Figure 9(b)]. This extension is called the *frustum ICP* approach in the following. It is not in contradiction to other rejection rules and can be jointly employed with them, e.g., when the rejection of point correspondences is needed through independent motion in the scene.

4.2. Feature-Based Ego Motion Estimation

Range image registration based on the ICP approach is computationally complex. The most time-consuming task is the search for nearest neighbors, which has a naïve complexity of $\mathcal{O}(n^2)$. Several measures have been proposed to reduce the computational effort, e.g., the use of search trees or the subsampling of points (see Nüchter et al., 2007, for an overview).

Subsampling and assignment of point correspondences can also be achieved without the iteration scheme while identifying and tracking discriminative features in the reflectance image. Two fairly recent approaches are used in terms of achieving real-time applicability, KLT feature tracking (Tomasi & Kanade,

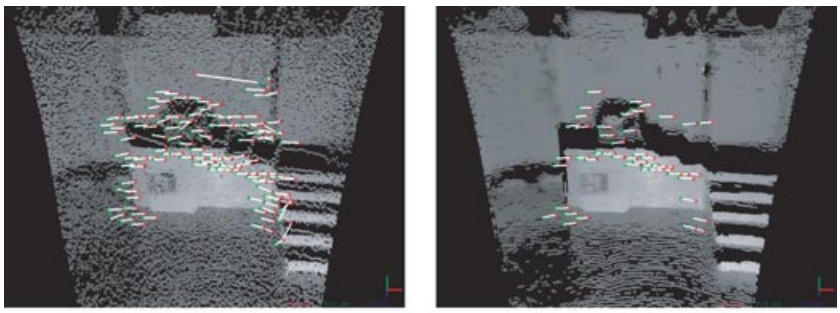
1991) and SIFT-based feature tracking (Lowe, 2004).

KLT feature tracking models image motion by constraining the brightness. Unfortunately, this assumption is violated relying on ToF camera reflectance images. The brightness depends on the squared distance (Lange, 2000), the integration time, and the location in the image (see inhomogeneous image illumination in Section 3.2.2). The first issue is of minor influence because image motion is assumed to be small. A change in integration time is clearly noticeable in the brightness from one frame to the next. This influence is also known from video cameras employing autoexposure in order to handle a high dynamic range in illumination, especially in outdoor scenes where sunlight has a strong impact on the brightness. An approach addressing this problem with an extension for KLT feature tracking can be found in Kim, Frahm, and Pollefeys (2007). The third problem is caused by inhomogeneously illuminating the scene by the active light emitters of ToF devices and the vignetting effect. A gradient in illumination is jointly moved with the camera under ego motion and influences the matching quality. Figure 10(a) depicts the feature tracking based on KLT features in reflectance images of a SwissRanger device.

SIFT feature tracking employs an image pyramid convolved with a difference-of-gaussian function. A search for possible feature locations is performed over all scales in the pyramid image stack. Those feature locations are robust with respect to changes in scale and orientation. At the candidate locations, a detailed model is fitted in order to determine precisely the location and scale. Only key points are selected that provide a certain measure of stability. An orientation and scale assignment is performed to each key point location with respect to local image gradient magnitudes and directions. This step achieves rotation and scaling invariance. Finally, a key point descriptor is created that allows the identification even with significant levels of local shape distortion and change in illumination. The latter aspect is important concerning the modalities of a ToF camera, for which the change in illumination is a matter of principle (see explanations for KLT feature tracking). Figure 11(a) depicts the feature tracking based on SIFT features in reflectance images of a SwissRanger device.

Ego motion estimation based on stereo vision needs the matching between image pairs in order to obtain 3D coordinates. The triangulation of feature pairs is error prone. Some approaches propose the use of only 2D image projections for the minimization process, e.g., Nister, Naroditsky, and Bergen (2004) and Sünderhauf and Protzel (2006). Contrary to stereo vision, we can make direct use of 3D information provided by the depth images. The ego motion estimation, i.e., estimation of $(\mathbf{R}, \boldsymbol{t})$, is performed by a least-squares method as for the ICP-based methods.

To achieve a robust feature tracking, two measures have to be applied. First, features related to jump edges are discarded. Second, outliers are identified and rejected. The most common approach to outlier detection is RANSAC (RANdom SAmple Consensus) (Fischler & Bolles, 1981). A random



(a) (b)

Figure 10. KLT feature tracking in ToF reflectance data. Feature points from model and scene are connected with a white line. (a) Initial set of tracked KLT features. (b) Remaining point correspondences after jump edge removal and RANSAC outlier detection.

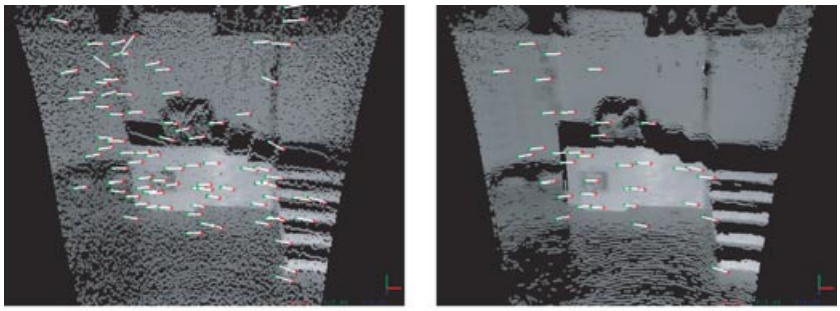


Figure 11. SIFT feature tracking in ToF reflectance data. Feature points from model and scene are connected with a white line. (a) Initial set of tracked SIFT features. (b) Remaining point correspondences after jump edge removal and RANSAC outliers detection.

subsampling of tracked features is used to estimate $(\mathbf{R}, \boldsymbol{t})$. Only those features providing the estimated model (consensus set), i.e., by having a small Euclidean distance between corresponding points after applying the transformation, are chosen for the next iteration. From the remaining feature set a random subsampling is used to restock rejected features. The application of RANSAC outlier detection on KLT and SIFT features is shown in Figures 10(b) and 11(b).

4.3. ESM-Based Ego Motion Estimation

Unlike feature-based methods, the ESM for visual tracking does not rely on extracting features and finding correspondences based on certain feature matching criteria. It aims at finding an optimal transformation between two subsequent data sets or between parts of it, i.e., the visual tracking of rigid and deformable surfaces (Malis, 2007). Here we show how the ESM can be applied to data acquired by ToF cameras.

The basic model assumes a pinhole camera model. Two pixels in the current and reference image are related by a warping function w:

$$\boldsymbol{u}' = w(\boldsymbol{u}, \boldsymbol{\eta}) \tag{13}$$

that allows us to obtain the current coordinates $\boldsymbol{u}'$ as a function of the reference coordinates $\boldsymbol{u}$ and the parameters $\boldsymbol{\eta}$. The vector $\boldsymbol{\eta}$ contains the intrinsic and extrinsic camera parameters $\mathbf{T_i}$ and $\mathbf{T_e}$ as well as surface parameters $\boldsymbol{s}$. If surface parameters are known and intrinsic camera parameters are kept fixed, only the transformation matrix $\mathbf{T_e}$ is considered in the estimation process. Indeed, for ToF cameras the surface parameters can be measured directly from depth images $\boldsymbol{s} = f(\boldsymbol{d})$.

The basic version of the ESM approach assumes constancy in brightness by setting

$$I[w(\boldsymbol{u}, \boldsymbol{\eta})] = I'(\boldsymbol{u}). \tag{14}$$

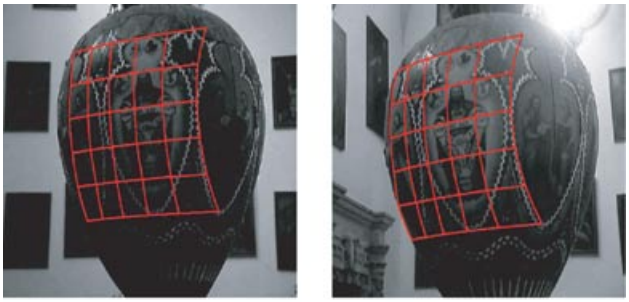
This modeling provides one equation per pixel and is iteratively solved by an efficient second-order approximation method (see Malis, 2007, for details). Real-time applicability is expected for the low-resolution images of ToF cameras (here 176×144 pixels).

As stated in Section 4.2., the brightness constancy assumption is violated for reflectance images due to inhomogeneous image illumination and changes in exposure time. Because the ESM approach determines a solution for a larger template (in contrast to KLT feature tracking), it is not expected that moderate changes in illumination influence the robustness significantly (see Figure 12). If needed, arbitrary illumination changes (even specular reflections) can be handled by an extended ESM version (Silveira & Malis, 2007).

The ESM approach can be specialized to consider all ToF camera data assuming constancy in depth (i.e., we observe a rigid object):

$$d[w(\boldsymbol{u}, \boldsymbol{\eta})] = d'(\boldsymbol{u}). \tag{15}$$

In this case, we have to solve a multiobjective optimization problem considering both Eqs. (14) and (15).



(a) (b)

Figure 12. ESM visual tracking. A template is reliably tracked even under moderate illumination changes (see details in Malis, 2007).

In this paper, we focus on finding the solution of an optimization problem based on Eq. (14) only. However, we use depth measurements to compute the warping function in Eq. (13). To our knowledge, this is the first attempt to adapt the ESM to ToF cameras.

4.4. Loop Closing for Consistent 3D Mapping

To register two overlapping ToF camera data sets, Eq. (9) is minimized. However, while successively aligning data with their local predecessors in a global coordinate system, small registration errors sum up. Loop closing is a remedy for this problem known from SLAM work. The idea is this: Once the sensor comes back again into a region of which data have been recorded earlier (the sensor has finished a loop), a new data set is registered not only with the ToF images taken just before, but also with the map data resulting from ToF images previously taken. Realizing this idea of course requires that the ToF camera pose estimation is explicitly considered (leading from basic mapping to SLAM) and that the pose estimation is at least sufficiently good to allow loop-closing events to be estimated—albeit imprecisely.

We use a probabilistic SLAM approach, i.e., a network-based global relaxation method for 3D point cloud data (Borrmann, Elseberg, Lingemann, Nüchter, and Hertzberg, 2008). It extends the GraphSLAM approach (Thrun & Montemerlo, 2006) to six degrees of freedom and restricts it to estimating poses instead of poses and features. The nodes of the SLAM graph are the ToF camera poses; the arcs connect those nodes that are known or estimated to cover spatial regions of the environment with sufficiently high overlap. For example, subsequent frames are connected until the pose change by ego motion is such that a new frame and a predecessor frame cover no or only a few points in 3D space. Closing a loop means establishing (or hypothesizing) an arc between nodes corresponding to ToF camera poses quite distant in the time they were taken but presumably close in the spatial region that they cover.

Given this network with $n+1$ nodes $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_n$ representing the poses $\boldsymbol{v}_0, \ldots, \boldsymbol{v}_n$ and the directed edges $\boldsymbol{d}_{i,j}$, the relaxation algorithm aims at estimating all poses optimally. The directed edge $\boldsymbol{d}_{i,j}$ represents the *change* of the pose $(x, y, z, \theta_x, \theta_y, \theta_z)$ that is necessary to transform one pose $\boldsymbol{v}_i$ into $\boldsymbol{v}_j$; i.e., $\boldsymbol{v}_i = \boldsymbol{v}_j \oplus \boldsymbol{d}_{i,j}$, thus transforming two nodes of the graph. For simplicity, the approximation is made that the measurement equation is linear, i.e.,

$$\boldsymbol{d}_{i,j} = \boldsymbol{x}_i \oplus \boldsymbol{x}_j.$$

A detailed derivation of the linearization is given in Nüchter (2009). An error function is formed such that minimization results in improved pose estimations:

$$W = \sum_{i \to j} (\boldsymbol{d}_{i,j} - \boldsymbol{d}_{i,j})^T C_{i,j}^{-1} (\boldsymbol{d}_{i,j} - \boldsymbol{d}_{i,j}), \qquad (16)$$

where $\boldsymbol{d}_{i,j} = \boldsymbol{d}_{i,j} + \Delta\boldsymbol{d}_{i,j}$ models random Gaussian noise added to the unknown exact pose $\boldsymbol{d}_{i,j}$. This representation involves resolving the nonlinearities resulting from the additional roll and pitch angles by Taylor expansion. The covariance matrices $\mathbf{C}_{i,j}$ describing the pose relations in the network are computed based on the paired closest points (Nüchter, 2009). The error function equation (16) has a quadratic form and is therefore solved in closed form by sparse Cholesky decomposition.

5. EXPERIMENTS AND RESULTS

This section describes experiments addressing the impact of calibration, filtering, and ego estimation approaches for improving accuracy, precision, and robustness. Influences of the light-scattering effect are rated in a separate experiment. Then, several ego motion estimation approaches are applied or adapted, respectively, in order to provide a performance benchmark for registering ToF camera data.

All experiments were carried out with a SwissRanger SR-3k device. This model features a

resolution of 176×144 pixels and provides depth and intensity data with a frame rate of up to 30 Hz. The unambiguity range is 7,500 mm.

5.1. Evaluation Measures

The quality of a 3D mapping approach can be rated by comparing either the created 3D map or the estimated path (ego motion) of the sensor with a ground truth measure.

5.1.1. Evaluation of Localization

Primarily, the presented 3D mapping approaches are benchmarked by analyzing the accuracy of the estimated ego motion absolutely and incrementally. For this purpose, an industrial robot arm with an accurate positioning system, i.e., a repeatable accuracy of 1 mm and 0.1 deg, is used for moving the SR-3k through the scene.

Let $\{\mathbf{T}_{r,i} | i = 0 \dots N\}$ be a set of poses provided by the robot control. The pose change between two consecutive poses can be determined by

$$\Delta\mathbf{T}_{r,i} = \begin{pmatrix} \Delta\mathbf{R}_{r,i} & \Delta\boldsymbol{t}_{r,i} \\ \mathbf{0} & 1 \end{pmatrix} = \mathbf{T}_{r,i-1}^{-1}\mathbf{T}_{r,i}, \tag{17}$$

where $\mathbf{T}_{r,0}$ is initialized with $\mathbf{T}_{r,1}$ (in order to provide $\Delta\mathbf{T}_{r,1} = \mathbf{1}$).

The registration of two consecutive range images results in an estimation matrix $\Delta\mathbf{T}_e$. Let $\{\Delta\mathbf{T}_{e,i} | i = 1 \dots N\}$ be the set of estimated pose changes for the whole trajectory. The estimated pose is then calculated recursively with

$$\mathbf{T}_{e,i} = \mathbf{T}_{e,i-1}\Delta\mathbf{T}_{e,i}. \tag{18}$$

The axis angle representation of a rotation matrix is used to evolve evaluation measures for the benchmark of different calibration and ego motion estimation approaches. The notations θ and $\boldsymbol{a} = (a_x, a_y, a_z)^T$ are used for the angle and the rotation axis of an arbitrary rotation matrix $\mathbf{R}$. Absolute and incremental pose changes are derived from this definition.

Absolute measures consider the pose change with respect to an absolute reference system (here the robot base):

$$\Delta\mathbf{T}_{\text{abs},i} = \begin{pmatrix} \Delta\mathbf{R}_{\text{abs},i} & \Delta\boldsymbol{t}_{\text{abs},i} \\ \mathbf{0} & 1 \end{pmatrix} = \mathbf{T}_{r,i}\mathbf{T}_{e,i}^{-1}. \tag{19}$$

The axis angle $\Delta\theta_{\text{abs},i}$ of $\Delta\mathbf{R}_{\text{abs},i}$ is used to define the *absolute angular error measure*

$$e_{\text{abs},\Delta\theta} = \Delta\theta_{\text{abs},i} = f_\theta(\Delta\mathbf{R}_{\text{abs},i}). \tag{20}$$

The *absolute translational error measure* is assigned to the magnitude of translation:

$$e_{\text{abs},\Delta\boldsymbol{t}} = \|\Delta\boldsymbol{t}_{\text{abs},i}\| : \tag{21}$$

and constitutes the second measure that can be consulted. However, both values provide only weak objectivity because registration errors can be compensated by other registration errors. Therefore, incremental measures are more suitable in terms of validity.

Incremental measures sum up the magnitude of registration errors in a stepwise way. To that end, the incremental error registration matrix is computed by

$$\Delta\mathbf{T}_{\text{inc},i} = \begin{pmatrix} \Delta\mathbf{R}_{\text{inc},i} & \Delta\boldsymbol{t}_{\text{inc},i} \\ \mathbf{0} & 1 \end{pmatrix} = \Delta\mathbf{T}_{r,i}\Delta\mathbf{T}_{e,i}^{-1}. \tag{22}$$

Summation over the magnitudes of incremental translation errors defines the *incremental distance error measure*:

$$e_{\text{inc},\Delta\boldsymbol{t}} = \sum \|\Delta\boldsymbol{t}_{\text{inc},i}\|. \tag{23}$$

Unfortunately, the rotation axis of $\Delta\mathbf{T}_{\text{inc},i}$ is not the same for each increment; thus, angles cannot be simply added up. A different rotation axis defines a different unit. For quantifying rotation angles around different axes, a transformation in vector representation is needed.

Using the axis angle representation, rotation angles around axes $\boldsymbol{a}_{e,i}$ and $\boldsymbol{a}_{r,i}$ are determined by

$$\Delta\theta_{e,i} = f_\theta(\Delta\mathbf{R}_{e,i}), \tag{24}$$

$$\Delta\theta_{r,i} = f_\theta(\Delta\mathbf{R}_{r,i}). \tag{25}$$

$\Delta\theta_{e,i}$ and $\Delta\theta_{r,i}$ are then used to find the normalized axes:

$$\boldsymbol{a}_{e,i} = f_{\boldsymbol{a}}(\Delta\mathbf{R}_{e,i}), \tag{26}$$

$$\boldsymbol{a}_{r,i} = f_{\boldsymbol{a}}(\Delta\mathbf{R}_{r,i}). \tag{27}$$

The *incremental angular error measure* is then determined as the Euclidean distance of axis vectors

multiplied with the magnitude of rotation angles:

$$e_{\mathrm{inc},\Delta\theta} = \sum \| \, |\Delta\theta_{r,i}|\boldsymbol{a}_{r,i} - |\Delta\theta_{e,i}|\boldsymbol{a}_{e,i} \, \|. \qquad (28)$$

[Note the similarity to Eq. (23).] These definitions provide a uniform measure for accumulating translational and rotational errors.

5.1.2. Evaluation of Mapping

Absolute accurate ego motion parameters do not inevitably yield a perfect 3D map. Even if the sensor was localized exactly, it could provide corrupted depth data, resulting in an inaccurate map. On this account, two additional measures are implemented in order to evaluate the quality of the generated 3D map.

Root mean square (RMS) errors express a measure for the fitting of data sets to a certain model. The quadratic distances of the point-to-point correspondences between a data set and a certain model are summed up, and the root mean error expresses the consistency. The RMS error is computed by Eq. (9) while applying the frustum culling technique in order to exclude wrong point correspondences resulting from nonoverlapping areas. Here, this measure is used mainly to evaluate the consistency of consecutive depth measurements. This consistency can be distorted by multiple ways reflections, light scattering, or overexposure and would yield a higher RMS error.

Isometry is the only measure that is used for rating the created map as a whole in this context. For this purpose, some characteristic distances in the scene were manually measured and compared to the corresponding ones in the 3D map [see Figure 17(a) later in the paper].

5.2. 3D Mapping of a Laboratory Environment

The following experiments aim at rating the impact of calibration, light scattering, exposure time, and registration approaches separately by performing different trajectories in modified scene configurations. For the purpose of ground truth evaluation the ToF camera was attached to the tool center point (TCP) of an industrial robot [see Figure 14(a)], which provided the sensor pose with an accuracy of 1 mm in translation and 0.1 deg in rotation.

Three setups were designed in order to evaluate the accuracy and their dependencies. In the first setup [*SI*; see Figure 13(a)] the camera was rotated by 90 deg around a basic geometric Styrofoam object located in a distance of 600 mm, such that it was kept centered in the field of view. Thus, the length of the performed trajectory covered a distance of 950 mm.

In the second setup [*SII*; see Figure 13(b)], the camera was translated by 50 mm and rotated by 22 deg. Contrary to *SI*, the observed Styrofoam object did not appear to be centered in the field of view all the time but moved from the right to the left side.

The third setup (*SIII*; see Figure 14) was enlarged and more complex. Several Styrofoam objects

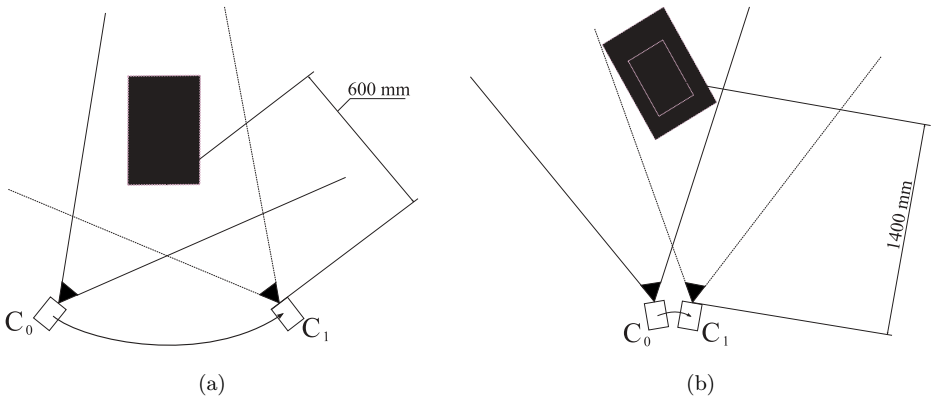


Figure 13. (a) Experimental setup *SI*: The camera was rotated by 90 deg around an object in a distance of 600 mm while it was kept centered in the field of view. The covered distance between C_0 and C_1 measured 950 mm. (b) Experimental setup *SII*: The camera was translated and rotated from C_0 to C_1 by 50 mm and 22 deg, respectively.

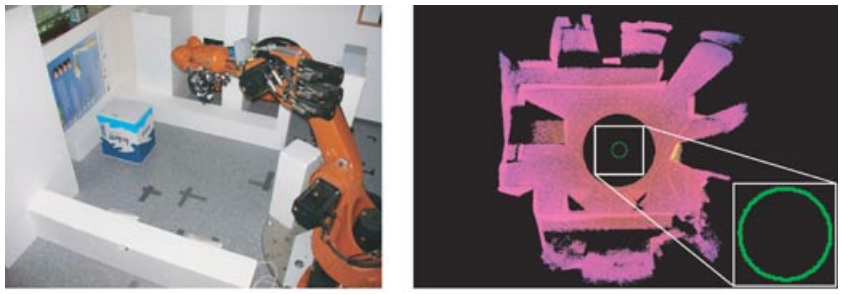


Figure 14. (a) Laboratory scene used for the ground truth evaluation. The ToF camera was mounted on an industrial robot arm (KUKA KR 16). (b) Bird's-eye view of the 3D map created by the use of provided robot poses (false color code relates distance to origin of coordinate system). The performed trajectory of circular shape is drawn at the lower right.

were assembled in a square, which measured approximately 1,800 mm. The performed trajectory in this experiment describes a circular path (diameter ≈180 mm) with a radially outward-looking camera mounting. A total of 180 range images have been taken equally distributed along this trajectory, i.e., in 2-deg steps.

5.2.1. Impact of Calibration

First, the impact of calibration was investigated. The calibration procedure outlined in Section 3.3.2. was applied. This so-called *improved calibration* considers circular errors and signal propagation delays. The effect of amplitude-related errors decreases with larger distances as a matter of decreasing bandwidths for amplitude values. Thus, the amplitude-related error was neglected. Figure 7(b) shows the identified circular error. While running the ToF camera at a working range of 500–1,800 mm in the laboratory setups, the default (manufacturer's) calibration provides depth data drifting between −150 and 10 mm.

The trajectories in *SI*, *SII*, and *SIII* were performed twice: once with default calibration and once with improved calibration. In *SI* and *SII* the calibration reduced errors in ego motion estimation by ≈25%. The translational errors decreased from 39.7 to 28.2 mm and from 12.8 to 9.9 mm (see Table I). But compared to the length of both trajectories, the translational error in *SII* was significantly larger (20%) than in *SI* (3%). The rotational error decreased only in *SI*, from 4.8 to 2.4 deg, whereas in *SII* it stayed nearly constant. We conclude that the type of movement was crucial for the results. Especially in *SII*, the observed object was located at the margins of the field of view. The low resolution and small apex angle handicapped ego motion estimation.

In *SIII* the results were less clear. Figure 15 shows that the incremental error accumulated to the same values (≈35 deg in rotation and ≈800 mm in translation). The localization results were only marginally affected by the improved calibration. Furthermore, the consistency was evaluated with respect to the RMS error, i.e., the indication of geometrical fit of subsequent range images. Again, improvements over the default calibration were only small (see Figure 16). The RMS error decreased by 0.26 mm on average.

The major impact of the improved calibration concerns the isometry of the resulting 3D map. Figure 17 depicts a 3D map of the laboratory scene

Table I. For both setups *SI* and *SII* the improved calibration minimizes the translational error by ≈25%. The rotational error decreases only in *SI*.

Calibration	*SI* rot. err. (mm)	*SI* trans. error (deg)	*SII* rot. err. (mm)	*SII* trans. err. (deg)
Default	39.7	4.8	12.8	1.2
Improved	28.2	2.4	9.9	1.3

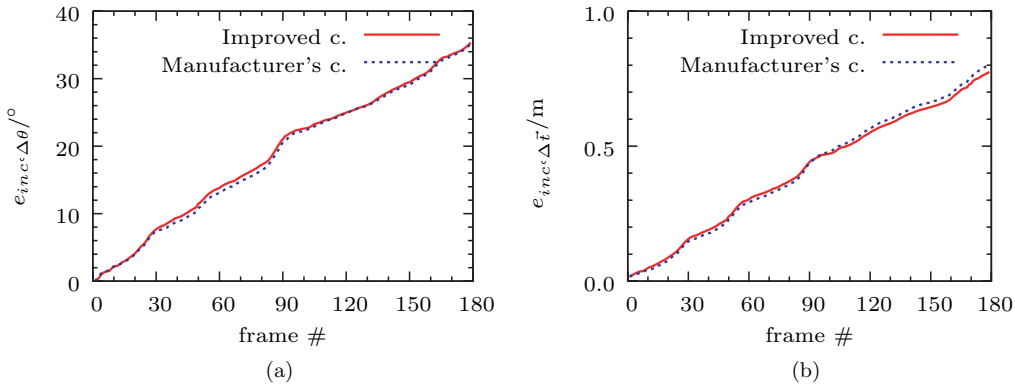


Figure 15. Incremental error measures for different calibrations in the laboratory experiment (frustum ICP registration). (a) Incremental angular error measure. (b) Incremental distance error measure.

from a bird's-eye view and compares the isometry between registration results obtained with default and improved calibration. Two distances between opposing surfaces in the environment were measured in order to relate them to ground truth. The isometry error decreased from 85 mm in the horizontal measure and 165 mm in the vertical measure to 35 and 20 mm, respectively.

Though the improved calibration significantly reduced ego motion estimation errors in *SI* and *SII*, an impact is hardly recognizable in *SIII*. This contradiction can be explained through an additional (unmodeled) error source, which was dominant for the performed trajectory and scene configuration in this experiment. It is assumed that the light-scattering effect completely adumbrated related results. A separate investigation is provided in the next section.

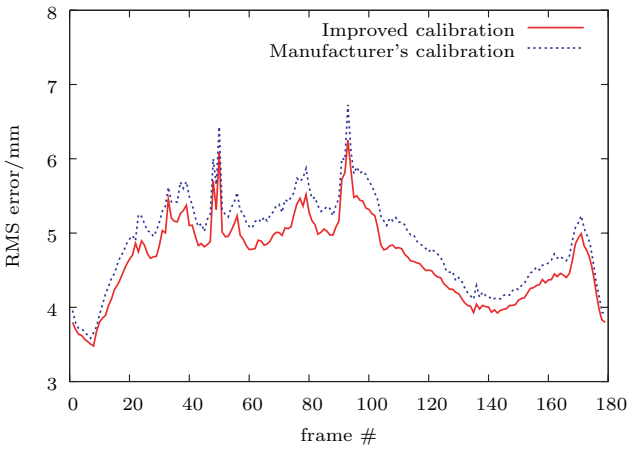


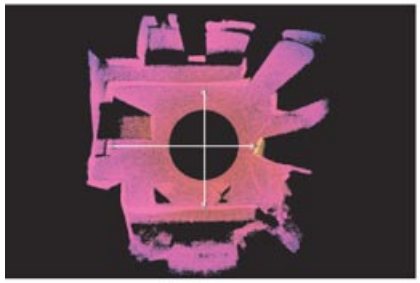
Figure 16. RMS error comparison of scene-to-model fitting employing default calibration ($\overline{\mathrm{RMS}_m} = 4.92$ mm) and improved calibration ($\overline{\mathrm{RMS}_i} = 4.66$ mm).

5.2.2. Impact of Light Scattering

The second investigation concentrated on the light-scattering effect. For this purpose, scattering effects were induced by adding an object to *SII* at several distances. Figure 18 shows the experimental setup and registration results. The closer the disturbing object was placed to the camera (from 1,250 mm down to 650 mm), the more the ego motion estimation results degraded in translation (from 9.9 to 42.1 mm). In particular, strong influences were noticeable when high reflective objects came into the field of view or when distances to objects in the scene were heterogeneous.

5.2.3. Impact of Integration Time

Third, the impact of the integration time was examined. For this purpose, the scene was captured twice, once with a fixed integration time and once having the exposure time controller activated. This controller adjusts the integration time so as to prevent overexposure and to maintain a good SNR. If it is possible to constrain dynamics in the field of view, a fixed integration time is preferable because depth measurements are affected in case of a change. On the other hand, improper adjustments of integration time may result in serious measurement errors (see Section 3.2.1.). Figure 19(a) contrasts the



(a)

	Hor. dist. (mm)	Vert. dist. (mm)
Ground truth (meas. tape)	1,715	1,405
Default calibration	1,800	1,570
Improved calibration	1,750	1,425

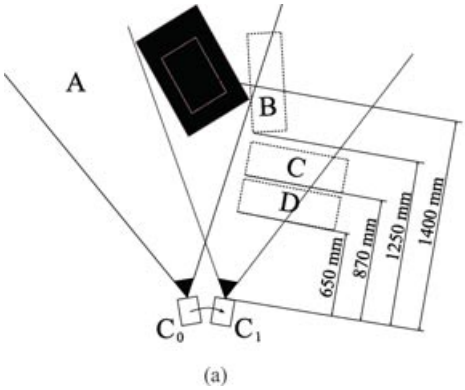
(b)

Figure 17. (a) Isometry of resulting 3D map of *SIII*. The distances of opposing walls were manually measured and assumed to be the ground truth. (b) The improved calibration reduces deviations from the ground truth and provides a more accurate mapping result.

RMS error for both cases. Even in this well-defined laboratory experiment it was noticeable that the integration time is an important parameter in terms of measurement accuracy. Unfortunately, the change of integration time influences the distance measurements. Figure 19(b) shows results of an experiment in which the integration time was periodically adjusted from 5,000 to 15,000 μs against a static target. Distance measurements altered by up to 20 mm. In conclusion, exposure time control is necessary in dynamic environments, but it is also a matter of calibration due to their influences on the accuracy. Nonetheless, exposure time control is given top priority. Thus, the automatic integration time controller was activated for the experiments in the following sections.

5.2.4. Evaluation of Registration Approaches

The evaluation of registration approaches was performed on the laboratory data set *SIII*. We concentrated on the most popular registration techniques. On the one hand, the ICP algorithm was applied to the distance data. On the other hand, two feature-based tracking methods, SIFT and KLT, were examined for their suitability for working on the reflectance images. Features in the reflectance domain



(a)

	Trans. error (mm)	Rot. err. (deg)
A	9.9	1.3
B	35.4	2.1
C	36.0	2.1
D	42.1	1.8

(b)

Figure 18. (a) Experimental setup for light-scattering investigations. The camera is moving from C_0 to C_1. Initially, the scene consists of two Styrofoam cuboids standing on top of each other (case A). Then (cases B, C, and D), an additional Styrofoam cuboid was put into the scene. (b) Ego motion estimation results degrade from 9.9 to 42.1 mm due to light-scattering effects.

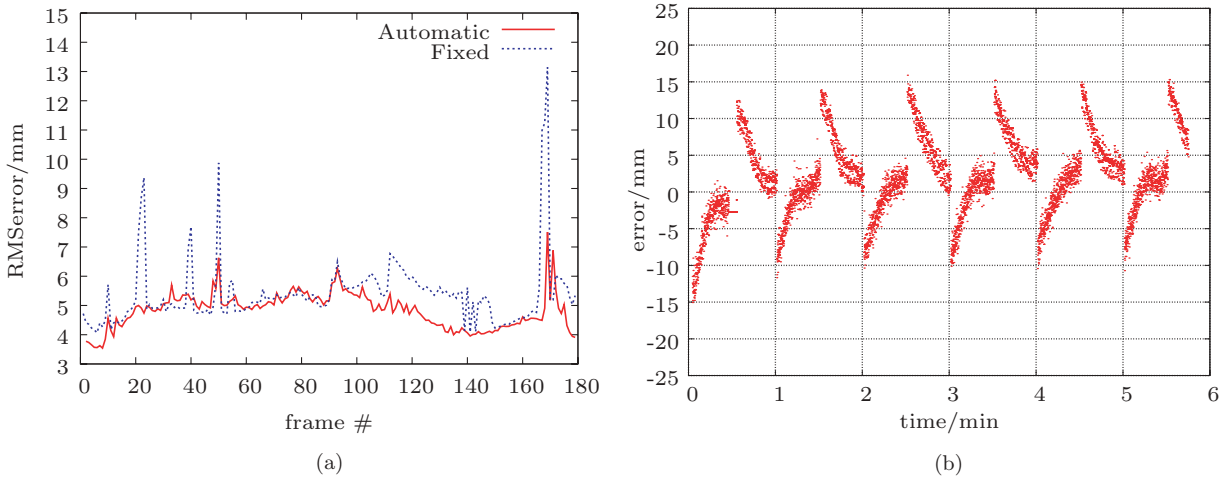


Figure 19. Impact of exposure time. (a) RMS error comparison of scene-to-model fitting between fixed integration time and automatic integration time exposure. Robot poses were used as ground truth for calculating the RMS error. Several peaks indicate a bad fit between model and scene points resulting from a nonoptimal integration time adjustment. (b) The automatic exposure time controller may cause steps of several thousand microseconds, especially if near objects come into the field of view. The right-hand diagram illustrates periodical steps of 10,000 μs. The error alters between −10 and 10 mm. After a short setting time, the error converges but a little drift remains.

were tracked in order to identify corresponding 3D coordinates computed from related depth measurements. Feature tracking approaches are computationally cheaper than ICP matching, which currently does not meet the timing requirements for real-time mapping applications based on ToF camera data (defining real time as to work on the full frame rate).

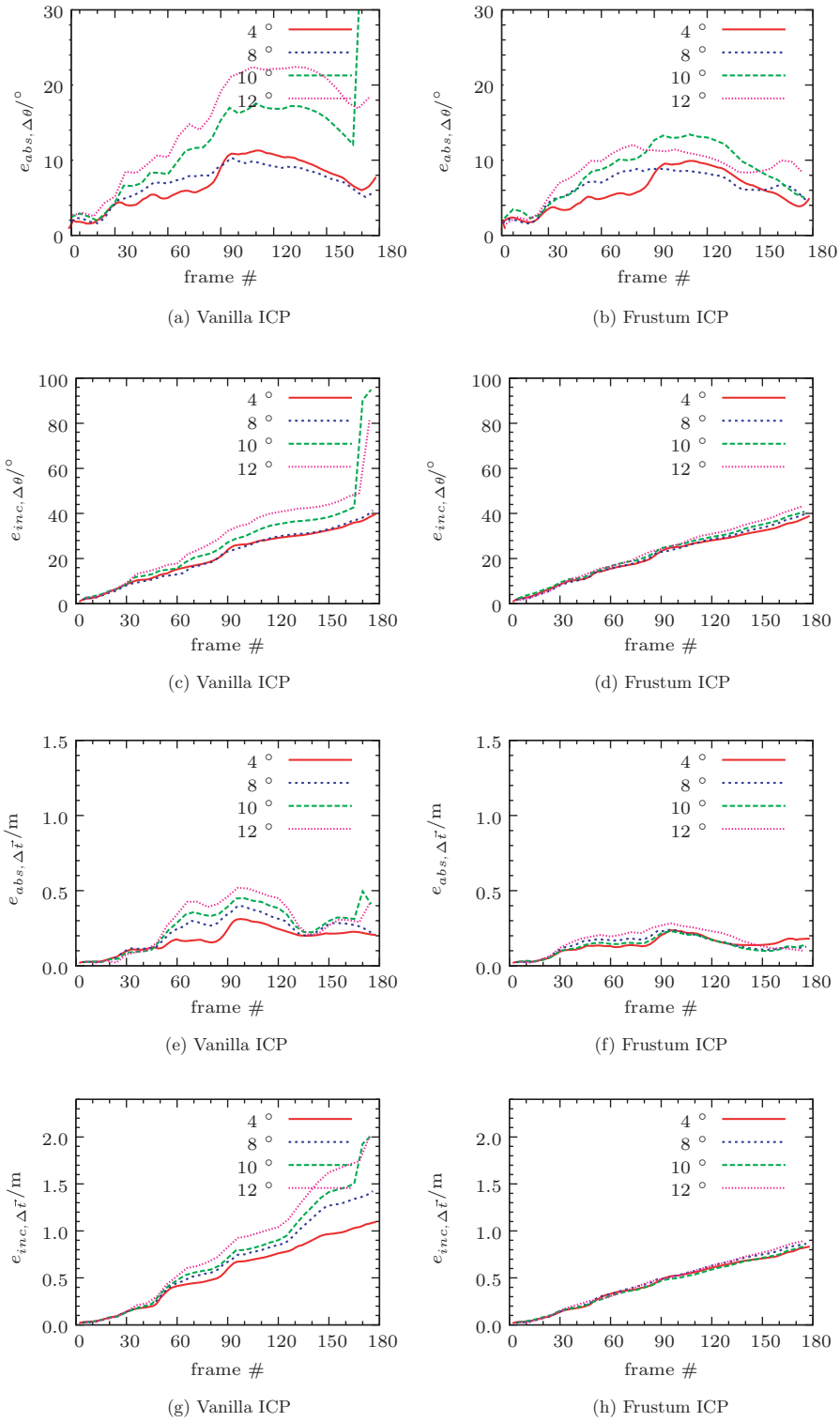
Beyond these "traditional" registration techniques, ESM tracking has been evaluated. This tracking method employs both dimensions—intensity and depth. Notably, ESM has been designed for real-time applicability in visual tracking tasks. This section benchmarks the performance, robustness and quality of ToF image registration for the above-mentioned approaches.

ICP-based range image registration: The application of ICP was evaluated in terms of robustness and accuracy. A basic assumption of the original formulation of the ICP algorithm is that the scene point set is a subset of the model point set. In 3D mapping applications, in which the basic task is to align 3D point sets with only partial overlap, this assumption is usually violated. Owing to the small angular steps of 2 deg in the laboratory experiment, the degree of overlap is high. Therefore, the vanilla ICP approach (employing a distance threshold) performed well for the basic setup. With increasing angular step widths, i.e., with faster angular motion, the registration got to be more and more of a problem, because the shrinking area of overlap induced a large portion of wrong point correspondences, which could not be determined completely by means of a distance threshold.

Owing to the small apex angle of ToF cameras, those "misassignments" can influence the registration result significantly. This depends on the geometric distinctiveness of measurements in the field of view, e.g., facing a large planar wall vs. a curved staircase. [See Figures 33(a) and 33(b) later in the paper.]

Clipping points from nonoverlapping areas increased the robustness, especially in scenes of poor geometric structure. Figure 20 contrasts the vanilla ICP approach with frustum ICP. It can clearly be seen that vanilla ICP achieves good results for this scene up to angular steps of 8 deg. For larger angular steps the approach converged more often to wrong local minima. The frustum culling variant still performed well on configurations with angular steps up to 12 deg.

An evaluation of minima to which ICP-based methods converged reveals that depth measurement errors have a deep impact on the registration results. Figure 21 shows the RMS error of fitting consecutive



(a) Vanilla ICP (b) Frustum ICP

(c) Vanilla ICP (d) Frustum ICP

(e) Vanilla ICP (f) Frustum ICP

(g) Vanilla ICP (h) Frustum ICP

Figure 20. Incremental error measures comparing vanilla ICP with frustum ICP. Frustum culling increases the robustness with respect to larger angular steps. (a)–(d) Incremental angular error measure. (e)–(h) Incremental distance error measure.

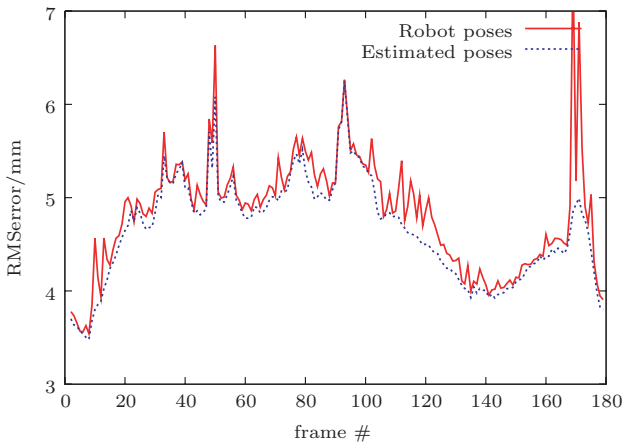


Figure 21. RMS error comparison of scene-to-model fitting employing robot poses ($\overline{\text{RMS}_r} = 4.84$ mm) and estimated poses from the ICP algorithm ($\overline{\text{RMS}_e} = 4.66$ mm).

point sets in dependency of their registration. The RMS errors were computed from robot poses (which are highly accurate) and poses estimated by frustum ICP. Though the estimated ICP poses were erroneous, the RMS error related to these ICP poses was always lower than or equal to the RMS error related to robot poses. This means that the error criterion of the ICP approach is not able to detect errors induced by the related measurement principle of ToF cameras and thus converges to a wrong result.

Figure 22 confirms this. Even using the ground truth robot poses as initial estimations does not help. Though a few scene point sets were "trapped" in local minima without the correct initial guess, namely, some frames around numbers 15 and 50, all remaining estimations were nearly identical, as indicated by the fact that all plots run mostly in parallel, except for the few mentioned frames. Knowing the true pose, therefore, does not guarantee ending up in an optimal registration by ICP.

Figure 23 shows the maps reconstructed using frustum ICP. The trajectory is distorted especially in the first part, where near bright objects (Styrofoam cubes) partially occluded the background. The trajectory runs from the lower right-hand corner in counterclockwise rotation. The second half of the trajectory reveals a semicircle.[1] Even employing highly accurate poses provided by the robot control as initial guesses for the ICP approach did not improve the pose estimation process in this experiment. The reconstructed trajectories have nearly the same shape and indicate that remaining deviations are due to inherent measurement errors, i.e., nonsystematic errors that are difficult to model (see Section 3.2.1).

The translation estimation is problematic in this experiment, as the single steps that were performed are tiny: The movement on the defined circular path with a diameter of 180 mm results in translational steps of 3–19 mm (for angular steps of 2–12 deg). Considering the incremental translational error, one can see that the measurement error is significantly larger on average than the performed translation. That means that a translational comparison of algorithms is not meaningful in this experiment and is consequently skipped.

Feature-based SLAM: Both feature tracking techniques, KLT and SIFT, can be applied during a mission with a reduced frame rate. The mean time elapsed for KLT feature tracking was $\bar{t}_{\text{KLT}} = 356$ ms; SIFT performed with a mean time of $\bar{t}_{\text{SIFT}} = 293$ ms.[2]

Applying feature tracking techniques has the advantage of a lower computational effort compared with range image registration based on ICP. The performance of both feature tracking approaches can be seen in Figure 24. The quality of matching was better for the SIFT approach. The reason is the robustness of SIFT against changes in illumination. A change in brightness is caused primarily by the automatic integration time exposure and the inhomogeneous image illumination. The latter effect was especially noticeable in KLT feature tracking. The inhomogeneous illumination gradient, which is "moved" with the sensor, showed a high responsiveness for the KLT feature detector, especially in scenes where near bright objects were in the field of view. Figure 25 depicts the reconstructed maps based on both feature trackers. The map obtained with the KLT approach

[1]Even this is not necessarily a meaningful indicator for a good reconstruction, because the diameter can be different.

[2]Note that the KLT approach is normally reported to be faster. The run time and quality of feature matching of both approaches depend on their parameterization, which, for instance, influences the number of tracked features. The parameters used here were optimized in terms of higher accuracy of registration results and were empirically found because the influence of measurement errors and change in brightness had to be considered.

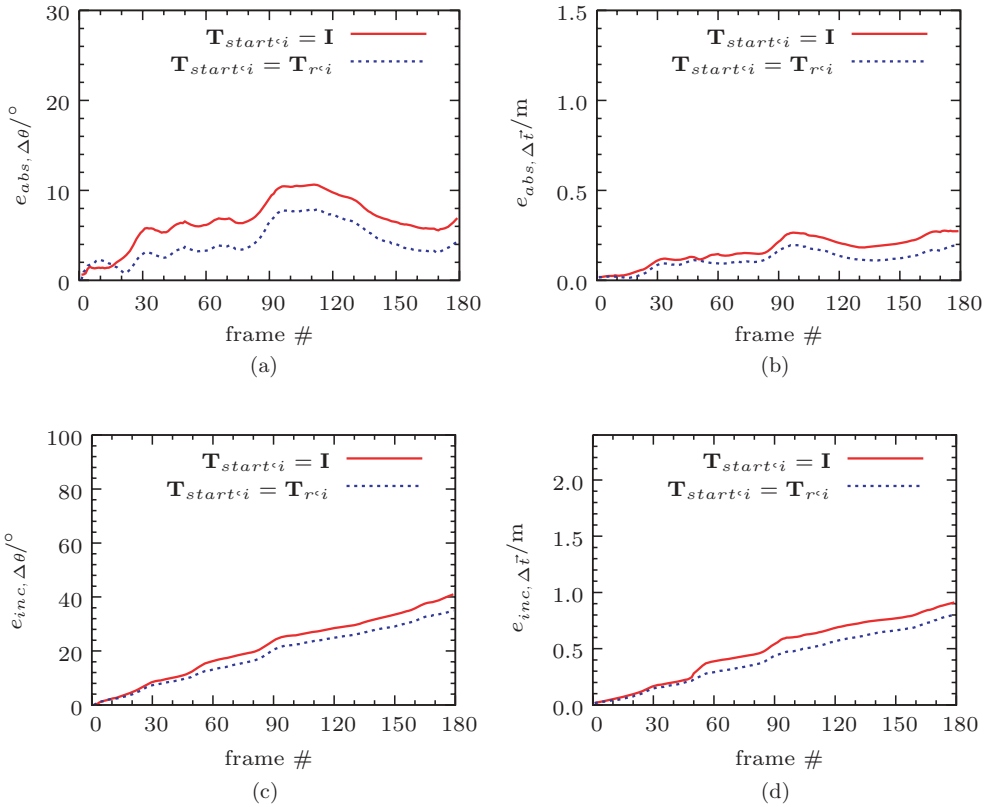


Figure 22. Absolute and incremental error measures comparing ICP convergence with or without employing ground truth robot poses as initial guesses. The profiles of the curves are nearly equal (except pose change estimations for frames around frames 15 and 50). Only a few range images were "trapped" in different minima.

shows considerably more distortions than the reconstruction based on SIFT.

The reconstruction quality for both feature trackers is lower compared to ICP registration. There are mainly two reasons for this fact. First, focusing on a small subset of the image, information is more likely to be distorted by the notorious depth measurement errors. Second, results depend on the amount of

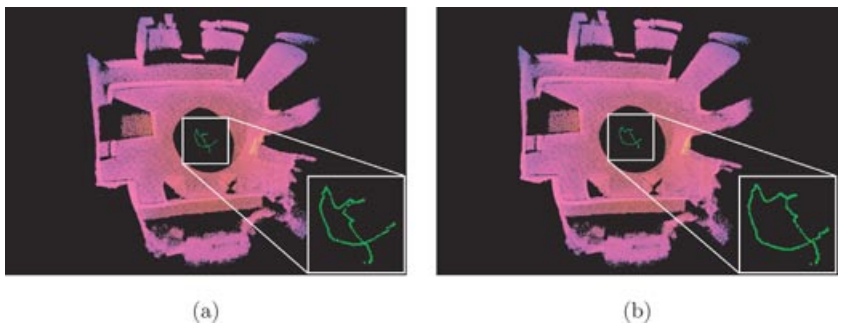


Figure 23. Reconstructed 3D maps and trajectory plots computed with the frustum ICP approach. (a) Two-degree angular displacement using no initial pose estimate. (b) Two-degree angular displacement using ground truth robot poses as initial guesses. The trajectory has nearly the same distortions as without providing accurate robot poses as initial guesses.

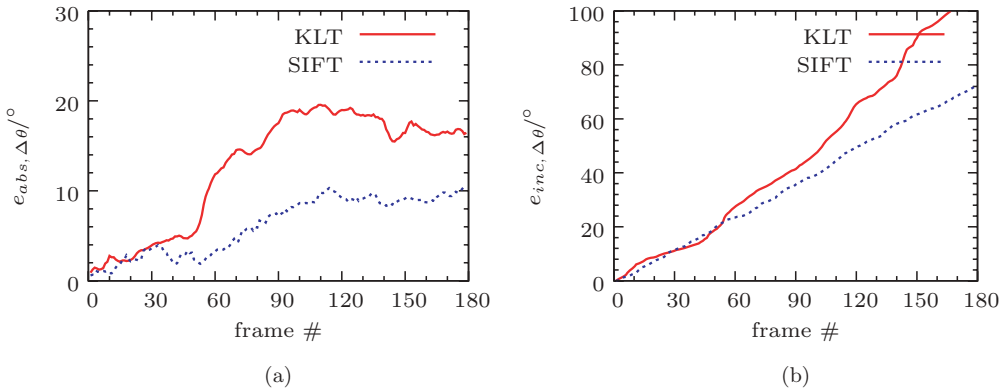


Figure 24. Error measures showing the performance of pose estimation based on KLT feature and SIFT feature tracking. (a) Absolute angular error measure. (b) Incremental angular error measure.

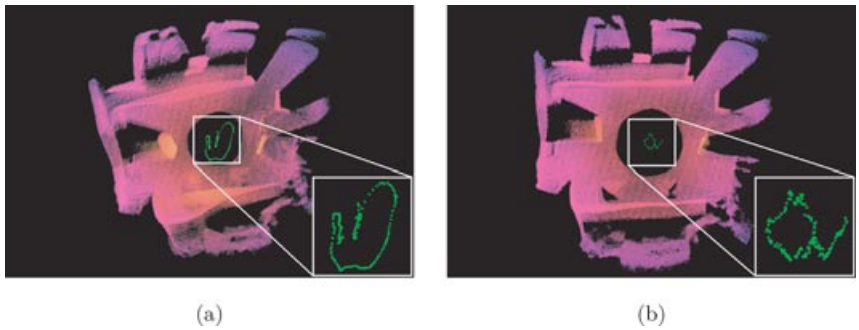


Figure 25. Reconstructed 3D maps with feature-based ego motion estimation. (a) 3D map and trajectory obtained with KLT feature tracking. (b) 3D map and trajectory with SIFT feature tracking.

texture in amplitude images. On the contrary, an ICP-based approach is more sensitive to structure in-depth images.

Direct registration with the ESM: The ESM visual tracking approach was implemented in Matlab and applied using 10 iterations per image. Considering that the number of unknown parameters is six, an optimized implementation would be able to run in real time at 30 ms/image. (This is the time for estimating eight unknown parameters of a homography matrix using the ESM visual tracking implemented in C.[3]) The direct registration outperformed the above-mentioned feature tracking techniques in the incremental angular error measure (see Figure 26) but had difficulties in the translational dimension, due to the remaining depth measurement errors (e.g., the light-scattering effect).

These errors have no influence on image intensities. Thus, we obtained an accurate amplitude image registration even with wrong depth measurements. Figure 27(a) shows the zero-mean normal cross correlation (ZNCC) score of the registration between subsequent images. The ZNCC was almost 1 for the whole sequence, thus indicating good image registrations. Figure 27(b) depicts the 3D map and the trajectory obtained with the ESM approach. The translational error can be seen on the left-hand side of the map. Here a large displacement between measurements from the same cuboid, which was in the range of sight in the first and last frames, is noticeable.

[3]The ESM visual tracking for planar objects is available at http://esm.gforge.inria.fr/ESMdownloads.html.

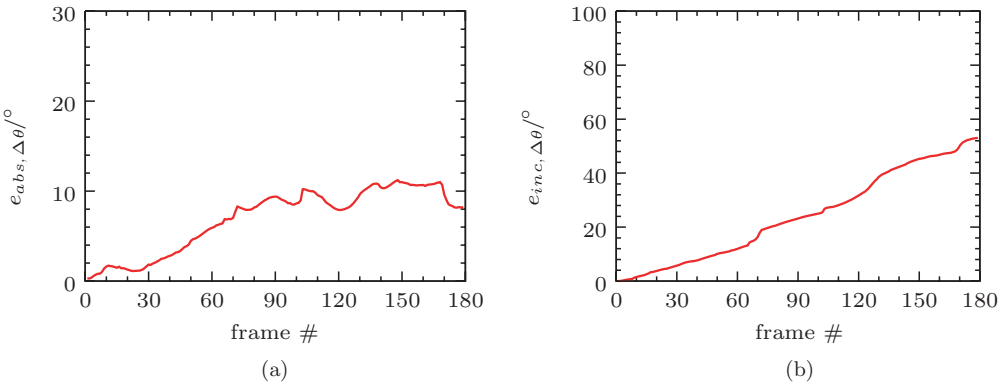


Figure 26. Error measures showing the performance of pose estimation based on ESM visual tracking. (a) Absolute angular error measure. (b) Incremental angular error measure.

5.2.5. Loop Closing

Once a loop closing is detected, or rather hypothesized, the accumulated error can be distributed uniformly in order to increase the consistency of a resulting 3D map. GraphSLAM was applied on poses calculated by the frustum ICP approach. The absolute and incremental error measures are depicted in Figure 28. Although the trajectory is still not of circular shape (see Figure 29), the accumulated error could be reduced significantly.

5.2.6. Concluding Remarks

Figure 30 contrasts the absolute and incremental error measures for all employed ego motion estimation approaches. ICP achieved the best results, but its run time does not allow it to be used at frame rate.

In comparison to the ICP approach, three real-time-applicable approaches were employed. The feature tracking techniques achieved good results. KLT feature tracking needs an extended modeling coping

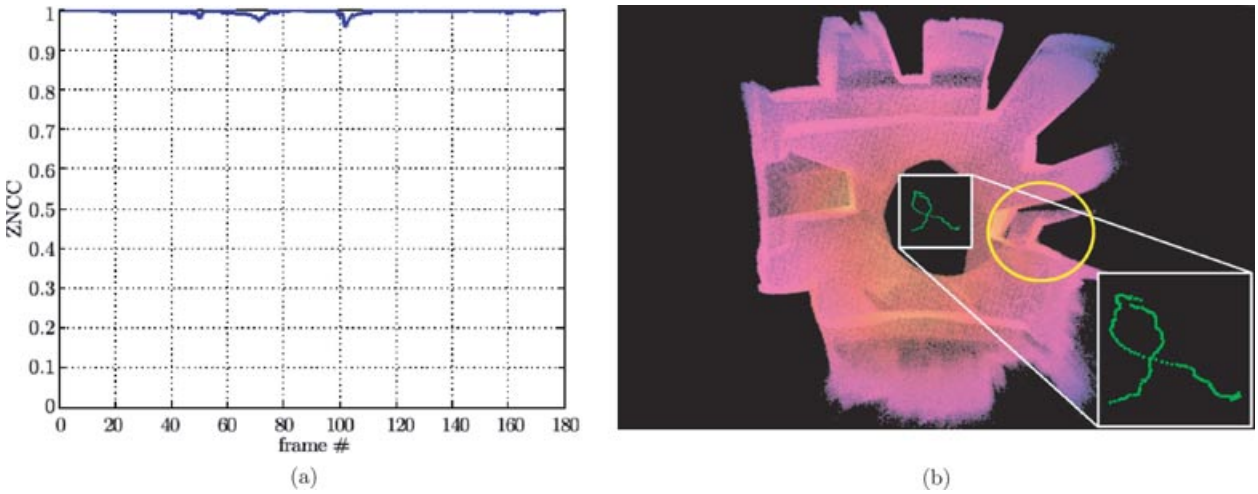


Figure 27. (a) The ZNCC between any two subsequent images registered with ESM. The ZNCC is almost 1 for each image pair, thus indicating a good image registration. (b) Reconstructed 3D map and trajectory with the ESM approach. The trajectory is most similar to the ICP reconstruction. The accumulated translational error can be seen by the displacement between measurements from the same cuboid, which was in the range of sight in the first and last frames (circular spot).

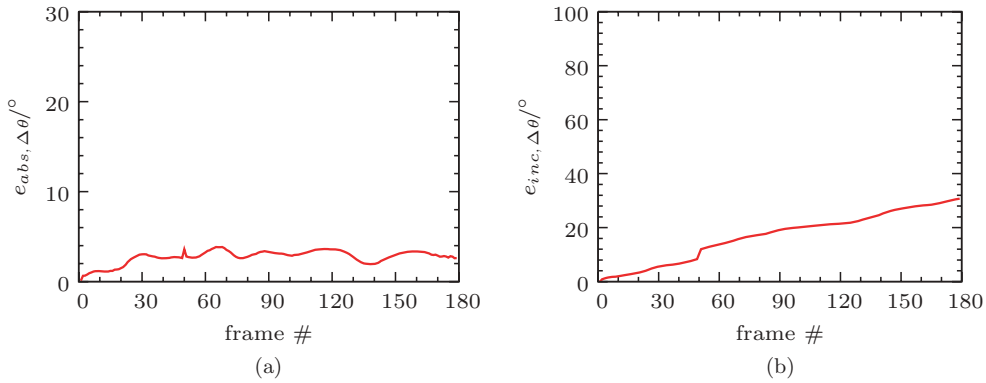


Figure 28. Error measures showing the performance of pose estimation based on frustum ICP (4-deg step width) after loop closing. (a) Absolute angular error measure. (b) Incremental angular error measure.

with changes in illumination (e.g., Kim et al., 2007). The SIFT feature tracking approach was robust against these issues because the SIFT descriptor is designed for that purpose. The ESM approach outperformed feature tracking techniques in the incremental angular error measure. It turned out to be valuable, employing both reflectance and distance data jointly for the estimation process. However, in order to avoid errors in estimating translational steps, the ESM method should be used with a robust method for discarding outliers in depth measurements. Robust estimation techniques have already shown to be effective for handling errors in the image intensities (Malis & Marchand, 2006). Finding robust extensions for error handling in the depth measurements is not straightforward and is a topic for future work.

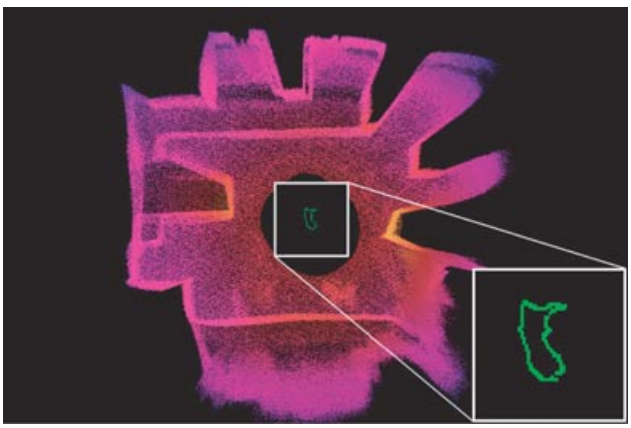


Figure 29. Reconstructed 3D map and trajectory with ICP approach after loop closing.

The experiments in this section showed that handling nonsystematic errors in ToF ranging is highly recommended for the implementation of accurate mapping applications. Trajectories are significantly distorted, although the RMS error measure is low. This means that subsequent frames fit well and the registration error is a matter of depth measurement distortions. These issues can be addressed either on the sensory level, e.g., by measures reducing the light-scattering effect, or on the methodical level, e.g., by employing robust statistics to estimate nonsystematic influences. Additionally, the relaxation of errors showed significant improvements. The inclusion of a frustum check in global scan relaxation techniques is a topic for future work.

5.3. 3D Mapping of Larger Environments

A further experiment focused on the mapping of a larger indoor environment. For this purpose, a looping trajectory in the robotics hall at the Fraunhofer Institute IAIS [see Figure 31(a)] was performed. The ToF camera was carried along manually, so no odometry was available for prior pose estimation. The hall size is 19.4 m in the longest distance (diagonal from corner to corner). Because the unambiguity interval is limited to 7.5 m, measurements appear closer (modulo 7.5 m) than they are, if the distance from camera to object exceeds this value. Mostly, the related amplitude can be used to discard those measurements, but

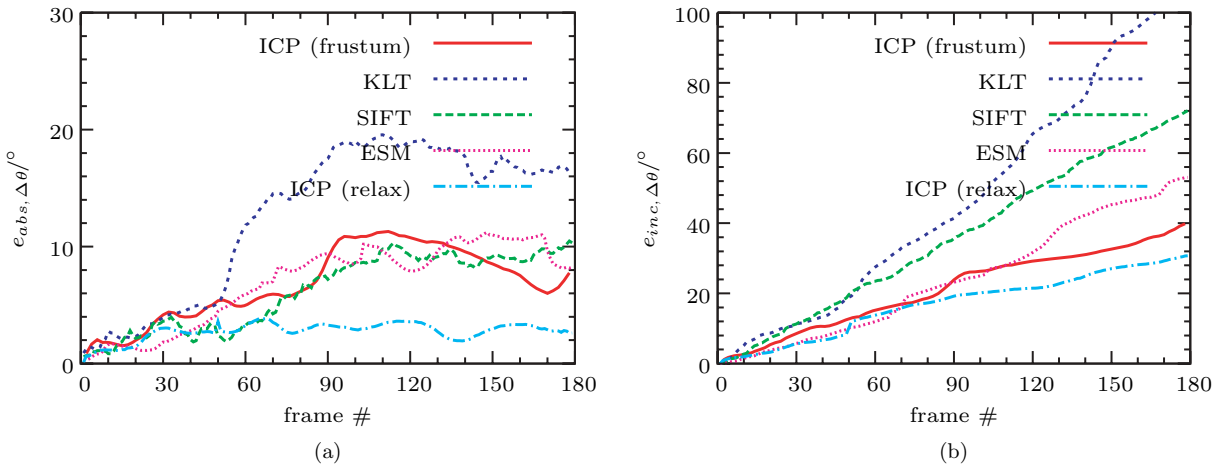


Figure 30. Error measures comparing the performance of all ego motion estimation approaches. (a) Absolute angular error measure. (b) Incremental angular error measure.

not in every case. Mismeasurements occur especially when specular surfaces are present in the field of view, e.g., mirrors, window panes, and metallic surfaces. The reflected signal then still has a high amplitude even at larger distances. Thus, amplitude-based filtering would not remove related measurements without discarding too many measurements in the close-up range.

This problem can be tackled reliably only on the sensory level, e.g., by employing coded binary sequences or multiple modulation frequencies. For the latter case, depth measurements are out of the unambiguity range, if they differ with several modulation frequencies. The follow-up model of the employed ToF camera, the SwissRanger SR-4k, allows for switching the modulation frequency in a large range with the consequence that one frame is lost in every frequency switching. If this reduction in frame rate is not acceptable, the camera can operate with a modulation frequency of 10 MHz, which enlarges the unambiguity range to 15 m. For the mapping experiment using the SR-3k, the application of amplitude filtering was sufficient after removing a few objects with specular reflectivity.

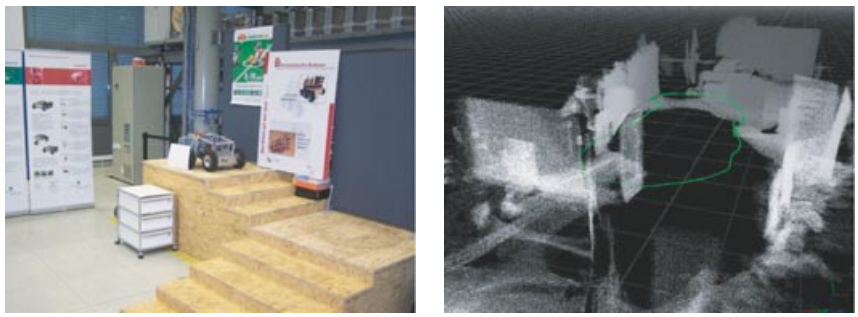


Figure 31. 3D mapping of a larger indoor environment—the robotics hall at Fraunhofer IAIS. (a) View of part of the hall. The environment contains different objects such as mobile filing cabinets, staircases, removable walls, and posters. (b) Perspective view of the registered 3D map of the hall. The performed trajectory (estimated poses of each frame) is drawn in green.

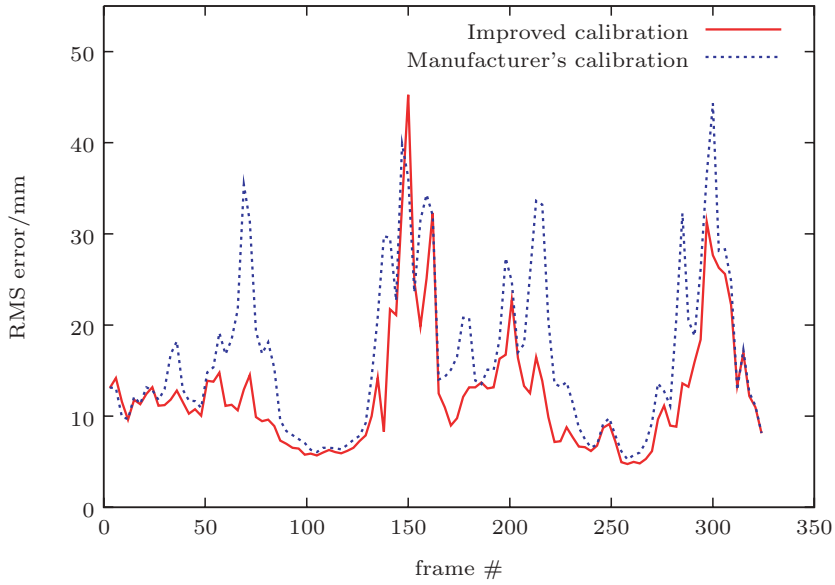


Figure 32. RMS error comparison (ICP registration) of model/scene fitting employing default calibration ($\overline{\text{RMS}_m} = 16.44$ mm) and improved calibration ($\overline{\text{RMS}_i} = 12.47$ mm).

5.3.1. Evaluation

Ground truth data given by an external positioning system were not available. For demonstrating the robustness of the entire approach, two measures were evaluated. First, the pose deviation (obtained by first-to-last range data registration) is provided to demonstrate that the error accumulation is small enough to perform loop closure in a larger scaled scenario. The second measure concerns the RMS error indicating consistency of subsequent registration results.

Figure 32 contrasts the scene-to-model fitting employing the default and the improved calibration. The impact of the improved calibration is more noticeable in this experiment than for the mapping of the lab scene. The circular distance error correction has a greater influence due to the coverage of a larger range interval and larger translational steps.

A second aspect that can be observed in this experiment concerns the increased robustness through frustum culling. Incorrect range image registration is often attributed to a narrow field of view. Figure 33 depicts two measurements made in this experiment. Point clouds shown in the left-hand image were taken subsequently against a staircase during ego motion [see Figure 31(a)]. Both data takes (model and scene) provide a high degree of geometric structure, so ICP-based approaches can be expected to perform well. Point clouds depicted in the right-hand image were taken a few steps later against the movable walls visible at the left-hand border of Figure 31(a). Geometric structure is poor, so vanilla ICP achieves poor results. In these scenes especially the performance gain of frustum culling was noticeable. Frustum ICP performed well for the whole data take. Figure 31(b) shows the reconstructed 3D map and the estimated trajectory. Errors accumulated after 325 range image registrations to $\|\Delta \boldsymbol{t}\| \approx 0.21$ cm and $\Delta\theta \approx 5$ deg.

5.3.2. Concluding Remarks

This experiment showed that 3D mapping of indoor environments is feasible by employing a ToF camera. The frustum ICP approach increased the robustness especially in scenes of poor geometric structure, while decreasing the computational effort.

A larger dynamic range made the impact of an improved calibration more noticeable. Also, the activation of the automatic integration time exposure showed a positive influence. Several preceding experiments with a fixed integration time provided worse results. The range of values adjusted by the integration time controller was between 8.9 and 65.5 ms, which indicates a large dynamic range for the experiment, i.e., the incidence of close-up

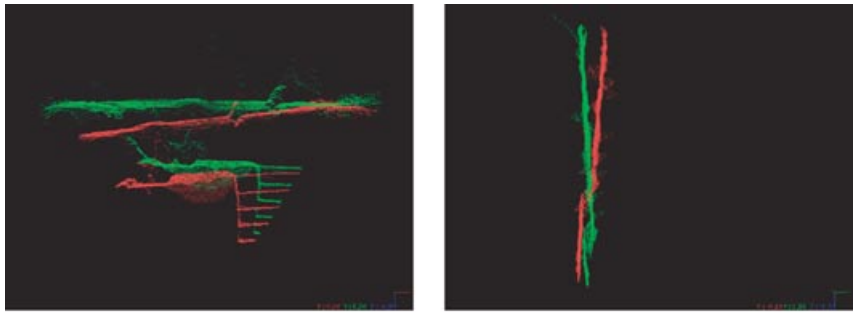


Figure 33. Matching model and scene point sets with ICP is difficult if measurements are of poor geometric structure. (a) Ego motion estimation can be performed robustly based on range image data while facing the staircase. (b) Reflectance-based methods perform better in scenes providing low geometric structure but texturedness, e.g., facing the movable poster walls visible at the left-hand border in Figure 31(a).

scenes comprising high-reflective surfaces as well as wide-range scenes. Preceding tests with fixed integration times often resulted in either low-illuminated or oversaturated measurements. Both situations have to be avoided.

6. CONCLUSION AND FUTURE WORK

This article investigated the applicability of ToF cameras in the context of autonomous mobile robotics, especially SLAM. Environment dynamics, as considered here, comprise a series of aspects, i.e., the sensor motion during data acquisition, changes in illumination as a matter of exposure time control, a high dynamic range in object reflectivity, and an unconstrained working range. In ToF camera-based applications, dealing with these environment conditions is often avoided or restrained, e.g., by defining a limited working range in manipulation tasks.

Performing robust 3D mapping with ToF cameras succeeds only if the special sensor characteristics are considered, for which this work provided three contributions. First, the impact of calibration has been shown with respect to the resulting 3D map. A higher accuracy in isometry could be achieved. Three different scenarios have been provided in order to rate improvements in accuracy for ego motion estimation and to demonstrate the influences of remaining error sources, i.e., mainly the light-scattering effect. It became evident that random errors affect results, depending on the performed motion.

Second, a robust nonparametric extension to the ICP approach has been presented. It showed robustness even when dealing with larger displacements. Third, a benchmark of ego motion estimation approaches has been provided. Two of the most common feature tracking algorithms based on reflectance data—KLT and SIFT—were compared to a purely depth image–based ICP approach and a hybrid technique, the ESM. Experiments focused on the applicability of those approaches with respect to the special characteristics of ToF camera data. The ESM has been modified in order to incorporate reflectance and distance data. The coupling of both dimensions in this approach achieved accurate results.

On the whole, the experiments revealed three issues for future work. To start, the modeling of nonsystematic errors such as the compensation of the light-scattering effect or the treatment of multiple-way reflections by means of geometric measures could enhance 3D mapping accuracy significantly. Then, a joint minimization criterion in the reflectance and depth domain will improve the robustness with respect to unambiguity in structure or texture of the measurement volume. Finally, the sensor fusion of ToF camera and an inertial measurement unit gives rise to improved mapping results.

The data sets and ground truth poses taken in the laboratory scene are available at http://www.robotic.dlr.de/242/.

REFERENCES

Besl, P., & McKay, N. (1992). A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2), 239–256.

Borrmann, D., Elseberg, J., Lingemann, K., Nüchter, A., & Hertzberg, J. (2008). Globally consistent 3D mapping

with scan matching. Journal of Robotics and Autonomous Sytems, 65(2), 130–142.

Chen, Y., & Medioni, G. (1991, April). Object modelling by registration of multiple range images. In Proceedings of the IEEE Conference on Robotics and Automation (ICRA '91), Sacramento, CA (pp. 2724–2729).

Chetverikov, D., Svirko, D., Stepanov, D., & Krsek, P. (2002, August). The trimmed iterative closest point algorithm. In Proceedings of the 16th International Conference on Pattern Recognition (ICPR), Quebec, Canada (vol. 3, pp. 545–548).

Cole, D., & Newman, P. (2006, May). Using laser range data for 3D SLAM in outdoor environments. In Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006, Orlando, FL (pp. 1556–1563).

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), 381–395.

Fuchs, S., & Hirzinger, G. (2008, June). Extrinsic and depth calibration of ToF-camera. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Anchorage, AK.

Fuchs, S., & May, S. (2007, September). Calibration and registration for precise surface reconstruction with TOF cameras. In Proceedings of the Dynamic 3D Imaging Workshop in Conjunction with DAGM (Dyn3D), Heidelberg, Germany (vol. I).

Fusiello, A., Castellani, U., Ronchetti, L., & Murino, V. (2002). Model acquisition by registration of multiple acoustic range views. In A. Heyden, G. Sparr, M. Nielsen, & P. Johansen (Eds.), Proceedings of the 7th European Conference on Computer Vision (ECCV). Number 2351, Lecture Notes in Computer Science (pp. 805–819), Copenhagen, Denmark: Springer.

Gabriel, P. (2006). Passenger classification for airbag control with 3D camera technology. Bachelor thesis, University of Erlangen–Nürnberg.

Holz, D., Lörken, C., & Surmann, H. (2008, June). Continuous 3D sensing for navigation and SLAM in cluttered and dynamic environments. In Proceedings of the International Conference on Information Fusion, Cologne, Germany.

Huhle, B., Jenke, P., & Straßer, W. (2007, September). On-the-fly scene acquisition with a handy multisensor-system. In Proceedings of the Dynamic 3D Imaging Workshop in Conjunction with DAGM (Dyn3D), Heidelberg, Germany.

Kahlmann, T., Remondino, F., & Ingensand, H. (2006, September). Calibration for increased accuracy of the range imaging camera SwissRanger. In International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Dresden, Germany (vol. 5, pp. 136–141).

Kim, S. J., Frahm, J.-M., & Pollefeys, M. (2007, October). Joint feature tracking and radiometric calibration from auto-exposure video. In Proceedings of the Eleventh IEEE International Conference on Computer Vision (ICCV), Rio de Janeiro, Brazil.

Lange, R. (2000). 3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology. Ph.D. thesis, University Siegen.

Lengyel, E. (2000). A fast cylinder-frustrum intersection test. In Game programming gems. Boston, MA: Charles River Media.

Lindner, M., & Kolb, A. (2006, November). Lateral and depth calibration of PMD-distance sensors. In Proceedings of the Second International Symposium on Advances in Visual Computing (ISCV), Lake Tahoe, NV (pp. 524–533).

Lorusso, A., Eggert, D., & Fisher, R. (1995, September). A comparison of four algorithms for estimating 3-D rigid transformations. In Proceedings of the 4th British Machine Vision Conference (BMVC), Birmingham, England (pp. 237–246).

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), 91–110.

Luck, J., Little, C., & Hoff, W. (2000, April). Registration of range data using a hybrid simulated annealing and iterative closest point algorithm. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), San Francisco, CA.

Malis, E. (2007, October). An efficient unified approach to direct visual tracking of rigid and deformable surfaces. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), San Diego, CA.

Malis, E., & Marchand, E. (2006, October). Experiments with robust techniques in real-time robotic vision. In IEEE/RSJ International Conference on Intelligent Robots Systems, Beijing, China (pp. 223–228).

May, S., Pervölz, K., & Surmann, H. (2007). 3D cameras: 3D computer vision of wide scope. In Advanced robotic systems, applications (vol. 4, pp. 181–202). Vienna: Advanced Robotic Systems (ARS).

Mure-Dubois, J., & Hügli, H. (2007, March). Real-time scattering compensation for time-of-flight camera. In Proceedings of the ICVS Workshop on Camera Calibration Methods for Computer Vision Systems, Bielefeld, Germany.

Niemann, H., Zinßer, T., & Schmidt, J. (2003, September). A refined ICP algorithm for robust 3D correspondence estimation. In Proceedings of the IEEE International Conference on Image Processing, Barcelona, Spain.

Nister, D., Naroditsky, O., & Bergen, J. (2004, June). Visual odometry. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC.

Nüchter, A. (2009). 3D robotic mapping–The simultaneous localization and mapping problem with six degrees of freedom. In Springer Tracts in Advanced Robotics (STAR), vol. 52. Berlin: Springer.

Nüchter, A., Lingemann, K., Hertzberg, J., & Surmann, H. (2007). 6D SLAM—3D mapping outdoor environments. Journal of Field Robotics, 24, 699–722.

Ohno, K., Nomura, T., & Tadokoro, S. (2006, October). 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.

Pajdla, T., & Van Gool, L. (1995). Matching of 3-D curves using semi-differential invariants. In Proceedings of the Fifth International Conference on Computer Vision (ICCV), Boston, MA (pp. 390–395).

Pathak, K., Birk, A., & Poppinga, J. (2008). Sub-pixel depth accuracy with a time of flight sensor using multimodal gaussian analysis. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Nice, France (pp. 3519–3524).

Prusak, A., Melnychuk, O., Roth, H., Schiller, I., & Koch, R. (2007, September). Pose estimation and map building with a PMD-camera for robot navigation. In Proceedings of the Dynamic 3D Imaging Workshop in Conjunction with DAGM (Dyn3D), Heidelberg, Germany.

Rusinkiewicz, S., & Levoy, M. (2001). Efficient variants of the ICP algorithm. In Proceedings of the Third International Conference on 3D Digital Imaging and Modellling (3DIM '01), Quebec City, Canada (pp. 145–152).

Sabeti, L., Parvizi, E., & Wu, Q. J. (2008). Visual tracking using color cameras and time-of-flight range imaging sensors. Journal of Multimedia, 3, 28–36.

Sappa, A. D., Restrepo-Specht, A., & Devy, M. (2001, July). Range image registration by using an edge-based representation. In Proceedings of the International Symposium of Intelligent Robotic Systems (SIRS), Toulouse, France (pp. 167–176).

Sheh, R., Kadous, M. W., & Sammut, C. (2006). On building 3D maps using a range camera: Applications to rescue robotics (Tech. Rep. UNSW-CSE-TR0609). Sydney, Australia: University of New South Wales.

Silveira, G., & Malis, E. (2007, June). Real-time visual tracking under arbitrary illumination changes. In Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), Minneapolis, MN.

Sünderhauf, N., & Protzel, P. (2006, September). Towards using bundle adjustment for robust stereo odometry in outdoor terrain. In Proceedings of Towards Autonomous Robotic Systems (TAROS), Guildford, UK (pp. 206–213).

Swadzba, A., Liu, B., Penne, J., Jesorsky, O., & Kompe, R. (2007, March). A comprehensive system for 3D modeling from range images acquired from a 3D ToF sensor. In The 5th International Conference on Computer Vision Systems, Bielefeld, Germany.

Thrun, S., Burgard, W., & Fox, D. (2005). Probabilistic robotics (intelligent robotics and autonomous agents). Cambridge, MA: MIT Press.

Thrun, S., & Montemerlo, M. (2006). The graph SLAM algorithm with applications to large-scale mapping of urban structures. International Journal of Robotics Research, 25(5–6), 403–429.

Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., Fong, P., Gale, J., Halpenny, M., Hoffmann, G., Lau, K., Oakley, C., Palatucci, M., Pratt, V., Stang, P., Strohband, S., Dupont, C., Jendrossek, L.-E., Koelen, C., Markey, C., Rummel, C., van Niekerk, J., Jensen, E., Alessandrini, P., Bradski, G., Davies, B., Ettinger, S., Kaehler, A., Nefian, A., & Mahoney, P. (2006). Stanley: The robot that won the DARPA Grand Challenge. Journal of Field Robotics, 23(9), 661–692.

Tomasi, C., & Kanade, T. (1991). Detection and tracking of point features (Tech. Rep. CMUCS-91-132). Carnegie Mellon University.

Weingarten, J. W., Grüner, G., & Siegwart, R. (2004, September). A state-of-the-art 3D sensor for robot navigation. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '04), Sendai, Japan.

Zhang, Z. (1992). Iterative point matching for registration of free–form curves (Tech. Rep. RR-1658). INRIA–Sophia Antipolis, Valbonne Cedex, France.