Calibration and Registration for Precise Surface Reconstruction with TOF Cameras

Stefan Fuchs and Stefan May

DLR, Institute of Robotics and Mechatronics,
E-mail: stefan.fuchs@dlr.de

Fraunhofer Institut, Intelligente Analyse- und Informationssysteme,
E-mail: stefan.may@iais.fraunhofer.de

Abstract: This paper presents a method for precise surface reconstruction with time-of-flight (TOF) cameras. A novel calibration approach which simplifies the calibration task and doubles the camera's precision is developed and compared to current calibration methods. Remaining errors are tackled by applying filter and error distributing methods. Thus, a reference object is circumferentially reconstructed with an overall mean precision of approximately 3mm in translation and 3° in rotation. The resulting model serves as quantification of achievable reconstruction precision with TOF cameras. This is a major criteria for the potential analysis of this sensor technology, that is firstly demonstrated within this work.

Keywords: Time-of-flight cameras, calibration, surface reconstruction

1 Introduction

Since their invention nearly a decade ago, TOF-cameras have attracted attention in many fields, e.g. automotive engineering, industrial engineering, mobile robotics and surveillance. So far, 3D laser scanners and stereo camera systems are mostly used for these tasks due to their high measurement range and precision. In contrast, TOF cameras allow for higher frame rates and thus enable the consideration of motion. However, the high frame rate has to be balanced with large fluctuations in precision depending on external interfering factors (e.g. sunlight) and scene configurations, i.e. distances, orientations and reflectivities. These influences cause systematic errors besides noise and have to be handled by the application. Without controversy, TOF cameras enable the handling of environment dynamics. Accurate measurements and robustness to these influences ease the research of related more sophisticated tasks, such as tracking and object recognition. This paper presents a method for surface reconstruction using TOF cameras. Surface reconstruction is a basic task for appli-

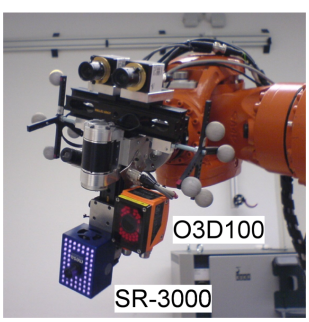

Figure 1 Experimental setup of DLR 3D-Modeller with TOF cameras

cations, i.e. object detection, manipulation, environment modelling or bin picking. The presented method includes a calibration approach, which addresses the error model, as well as necessary post processing methods tackling remaining inaccuracies. As a result, this work enhances the measurements by accurate pre-processing, showing the precision of an arbitrary object reconstruction. The outline of this paper is as follows: Section 2 presents related work with TOF cameras, i.e. calibration methods as well as implemented applications for these devices. Section 3 develops the error model, which is used in section 4 to derive an optimal calibration method. Section 5 and 6 comprise the reconstruction task while handling remaining imprecision. Experiments and results of the entire approach are presented in section 7 and finally concluded in section 8.

2 Related Work

This section presents related work on TOF cameras in the context of surface reconstruction. First an overview of calibration approaches will be given. The second part discusses applications of TOF cameras.

2.1 Camera Calibration

In relation to grey-scaled cameras that are defined by a pinhole camera model, calibration circumscribes the estimation of intrinsic parameters, i.e. distortion coefficients, focal length and shifting of the optical centre. Zhang presented a flexible technique to easily compute these projection parameters of a camera [14]. Furthermore, calibration also includes the pose estimation of the camera frame with respect to (w.r.t) the robot end-effector frame (also known as Tool Center Point (TCP) frame). This pose is important for dynamic applications, i.e. if several measurements have to be related to each other. Strobl et al. depicted a calibration method for eye-in-hand systems in order to estimate the hand-eye and the robot-world transformation [11].

Since TOF cameras provide grey-scaled images and are described by a pinhole camera model, these named methods in estimating the intrinsic and extrinsic parameters are feasible. In parallel to intensity images TOF cameras provide depth information, that is erroneous and has to be corrected. Currently, only few authors considered these both sides of the calibration task. Lindner et al. as well as Kahlmann et al. both estimated the intrinsic parameters of a TOF camera using grey-scaled shots of a chequerboard [7] and a planar testfield with Near-Infra-Red (NIR) LEDs respectively [5]. Both authors also investigated the faulty depth measurements by mounting the camera on a precise measurement rack and directed it towards a white smooth wall. With the known wall distance a systematic circular distance error was detected for the whole measurement range of 7500 mm. Lindner et al. approximated this error with cubic B-splines, whereas Kahlmann et al. compensated the error by a look-up-table. A per-pixel precision of at least 10 millimeters is achieved.

These investigations are very valuable for describing the error model of TOF cameras, but are laborious and not required for most applications which have a delimitable working range, e.g. grasping or bin packing. Furthermore, the TOF camera is calibrated to a known distance that is assumed to be the ground truth

but may be erroneous, because the camera is manually attached to the high precision rack. Since calibration is related to camera adjustments, especially to the adjustment of exposure time, this paper proposes a novel calibration method for a small working range (e.g. 500mm to 1500mm), while keeping these parameters constant. The calibration method concurrently estimates the transformation between camera and the TCP as well as the depth correction parameters. In contrast to the cited approaches no special pattern or ground truth is needed. The calibration is only based on depth images. Thus, manual errors are eliminated and the effort of a practical calibration is minimised.

2.2 Surface Reconstruction

Surface reconstruction is a basic task for object detection, manipulation and environment modelling. Generally, the object's surface is reconstructed by merging measurements from different views. For this approach the depth data and the pose of the sensor is needed. When both, the pose and the depth, are unknown, structure from motion is a solution. Corresponding features in consecutive images are used to estimate the ego-motion of the sensor. Based on this ego-motion information the depth is estimated. If only the depth information but no pose is given, i.e. by using a stereo camera or a laser scanner system without inertial sensors, an *Iterative Closest Point* (ICP) algorithm can be used to register the point clouds from different views [1]. Finally, if pose and depth are known, the registration procedure is dispensable and the data can simply be merged. In any case: The quality of surface reconstruction depends on the precision of sensor pose estimation and depth measurement.

There are already precise surface reconstruction applications available using laser scanners, e.g. Bodenmueller et al. presented an online surface reconstruction approach with a hand-guided scanner system [2]. The precision of registering one scan in the global coordinate system is in the order of some millimeters in translation and below 1° in rotation. A 3D mapping approach tackling large environments was presented by Nuechter et al. [8] using a 3D laser scanner mounted on a mobile robot. Imprecision of inertial sensors is handled with an ICP approach, both for registering consecutive scans and for *closing the loop*. The pose displacement between the first and the last scan of a loop quantifies the error summed up scan by scan and can be used to be distributed among all scans. The performance is one of the main issues, since millions of points are measured for a map with a trajectory length of 250 m.

In 2006 Ohno et al. used a TOF camera for estimating a robot's trajectory and reconstructing the surface of the environment [9]. An ICP algorithm registered the distance measurements to each other and provided the robot's trajectory. The calculated trajectory was compared to precise reference data in order to demonstrate the algorithm's precision. The estimation error for the robot pose was up to 15 percent in translation and up to 17 percent in rotation respectively. These results are far beyond the precision achieved with the scanner system referenced in [2], but sufficient for building maps of large environments, which is quite different from object reconstruction especially for small objects. In this paper range data registration is used to give a quantity of achievable precision for surface reconstruction tasks with TOF cameras, especially of small objects (100mm to 500mm).

3 Description of TOF Camera Errors

Lange detailed the different error sources, that falsify the TOF camera depth measurements [6]. In this context only those errors are mentioned which are considered by the presented surface reconstruction technique. First of all, there is an error in distance measurement depending on the NIR illumination of the scene. Low illumination results in a bad signal-to-noise ratio and distorts the measurement. This problem can be actively solved by carefully increasing the exposure time or the illumination power. A passive solution is filtering out the measurands with low amplitudes, which indicate a faulty depth. Furthermore, imprecision arises also for data acquired on jump edges. The measured distance is determined by a mixture of reflected light from fore- and background objects. It is aimed to discard this type of data by applying special edge filters. Finally, the electronic layout of the sensor and the special measurement principle of time-of-flight cameras create several systematic errors. Because the modulation of the transmitted NIR light is not perfectly sinusoidal, a *Circular Error* arises in the distance measurement with an amplitude between 60mm and 120mm and a wave length of 2000mm (see [7] [5]). Within the CMOS-gates of the sensor signal propagation delays are generated, which cause an offset or pixel dependent distance error. While the first mentioned errors are handled by filters, that are presented in section 5, the latter systematic errors are considered by a calibration step, which is explained in the following section.

4 Calibration Step

The intrinsic and extrinsic parameters as well as the systematic errors of the camera have to be precisely estimated for achieving a high level of precision. Initially, the intrinsic parameters of the camera are computed via a chequerboard pattern [3]. The extrinsic parameters could also be calculated via this pattern. However, the limited precision of corner localisation due to the low resolution of TOF cameras only allows for inaccurate pose estimation. Therefore, a new procedure for computing the extrinsic parameters is proposed, that simply needs depth images of a planar surface and supplementary parametrises the two discussed additional systematic errors: The circular error and the signal propagation delays. Let $\mathbf{P} = \{\mathbf{v}_j\}$ represent the depth image coordinates where $\mathbf{v} = (row, col)$ and j denotes the pixel index. The camera measurements of distorted distances are given by $(D^i(\mathbf{v})|i = 1, ..., n)$ where i denotes the image number. The circular error is modelled by a polynomial term $E_c(D^i(\mathbf{v})) = \sum_{k=0}^{m} p_k[D^i(\mathbf{v})]^k$. The error caused by signal propagation delays is linear dependent on the pixel location in the CMOS-array on the chip and geometrically described by $E_d(\mathbf{v}) = b_0 + b_1 r + b_2 c$. Summed up, the depth error is $C(D^i(\mathbf{v}), \mathbf{v}) = E_c(.) + E_d(.)$. Based on this assumptions, the desired sensor-to-tcp transformation, ${}^t\mathbf{T}_s$, is estimated. Thereto, the camera, which is mounted on a robot, is moved to different poses ${}^w\mathbf{T}_t^i$. These poses are given by the robot control. The calibration plane is defined by its normal $\mathbf{n}_c$ and its distance d_c to the origin of world coordinate system. Now, the unknown error-function parameters $(p_0, ..., p_m, b_0, b_1, b_2)$, the calibration plane pose $(\mathbf{n}_c, d_c)$ and the sensor-to-tcp transformation $({}^t\mathbf{T}_s)$ are estimated through a nonlinear least-squares fitting.

Thereto, an error function

$$F(p_0, ..., p_m, b_0, b_1, b_2, \mathbf{n}_c, d_c, {}^t\mathbf{T}_s) = \frac{1}{2}\sum_{i=1}^{n}\sum_{\mathbf{v}\in P} f_{\mathbf{v}}^i(.)^2 \tag{1}$$

is minimised. $A(D^i(\mathbf{v}), \mathbf{v}) \in \mathbb{R}^3$ forms the projection of a given distance and its image coordinates into the three-dimensional coordinate system.

$$f_{\mathbf{v}}^i(.) = \mathbf{n}_c^T[\,{}^w\mathbf{T}_t^i\,{}^t\mathbf{T}_s\,A(D^i(\mathbf{v}) - C(D^i(\mathbf{v}), \mathbf{v}), \mathbf{v})] - d_c \tag{2}$$

denotes the individual error per measurement. The error function $F(.)$ is minimised by trust-region methods for nonlinear minimisation in order to estimate the desired parameters.

5 Filtering the Depth Images

Calibrating the TOF camera increases the measurement precision for defined range intervals and camera configurations by adjusting systematic errors. The remaining errors caused by low illumination or jump edges, are treated by filtering. A high confidence is related to a high amplitude (to be precise: this statement is only a compromise to what the camera provides; see [6] for a description of error influences). Thresholding the amplitude discards primarily data resulting from objects with lower infrared reflectivity, higher distance or from objects which are located at the border area of the measurement volume. Edge detection in range images can be done by a variety of methods (see [4] for an overview). It is important to use an approach that can distinguish between different types of edges, i.e. jump and crease edges. Crease edges can be further classified into concave or convex roof and concave or convex nonroof edges. Measurement errors resulting from crease edges are difficult to handle in a post processing stage and should have been tackled sufficiently with a proper calibration. Therefore, it is sufficient to detect jump edges for filtering. The scan line approximation technique referenced in [4] represents a good approach in detecting distinguishable edges. From a range image $z(x, y)$, scan lines of different orientations are taken and split into quadratic curve segments. The computational effort depends on how many edges are being found. Especially noisy areas will be decomposed into many segments, e.g. background areas while focusing near bright objects. Confining on jump edges, a similar result is achieved with a more lightweighted approach. Data points are tagged as jump edges, if one of the opposing angles of the triangle spanned by the focal point and two neighboring data points exceeds a threshold, see figure 5.

6 Range Data Registration

In the context of range images, registration means to merge different point sets into a common coordinate system. With n datasets $(D^i|i = 1, ..., n)$, it is aimed to find the transformation matrices of those datasets $\mathbf{T}^i = [\mathbf{R}_i|t_i]$, which are composed of a rotation matrix $\mathbf{R}_i \in \mathbb{R}^3$ and a translation vector $t_i = (x, y, z)$. With the start transformation ${}^w\mathbf{T}_t^i\,{}^t\mathbf{T}_s^i$, given by the calibration (sensor to TCP) and the robot encoders, it is aimed to find the remaining displacement ${}^\theta\mathbf{T}_w^i$ completing in $\mathbf{T}^i = {}^\theta\mathbf{T}_w^i\,{}^w\mathbf{T}_t^i\,{}^t\mathbf{T}_s^i$. A common solution to the registration problem is the use/application of the ICP algorithm or one of its variants [1]. Since the pose of

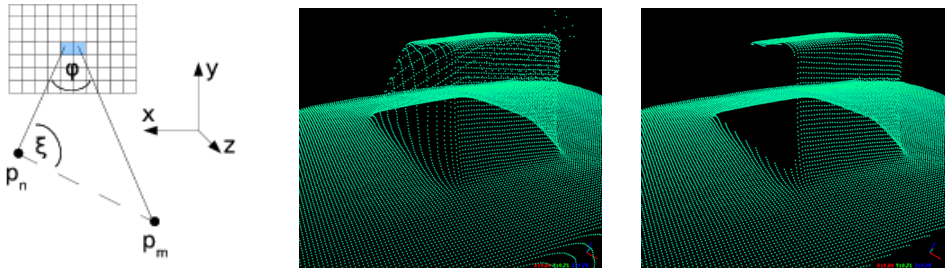

Figure 2 Left: Simplified filter for jump edge removement. p_n and p_m are neighboring data points and φ the angle between them and the focal point. A triangle is spanned between these points. The angle ξ is used for thresholding jump edges. Mid: Unfiltered image. Right: Appliance of gaussian blurring and jump edge filtering.

all datasets is determined with the same precision it is aimed to distribute the displacement of all surfaces with an approach called *simultaneous matching* [13]. The scene S is sequentially assigned to one of the datasets D_i. Therefore, the model M is composed of all datasets except the one that is taken as scene.

7 Experiments and Results

Two cameras were used for the experiments: the IFM O3D100 (www.ifm.de) and the SR-3000 (www.swissranger.ch). The O3D100 camera provides 50x64 pixels, whereas the SR-3000 has 176x144 pixels. Both cameras have been added to our experimental setup of the DLR 3D-Modeller [12], which was attached to the TCP of a standard industrial robot, type KUKA KR 16 (see figure 1 for an image of the sensor configuration on the rails). The robot has moved to different positions to take depth images for calibrating the cameras as well as for reconstructing the objects. The TCP pose for every depth image was given by the robot control with a precision of 1mm in translation and 0.1° in rotation.

7.1 Calibration Results

About 10 depth images with significant angle and different distances to the plane were used for the estimation of the desired sensor-to-tcp frame and the coefficients of the error model. Additional shots were taken to verify the calibration results. Every shot in this calibration step was taken 20 times to eliminate the noise by averaging the depth. Thus, the standard deviation for one pixel at an integration time of 8000μs was decreased from 4.1mm to 3.4mm (SR-3000) and from 2.2mm to 0.9mm (O3D100). In a verification step the corrected measurements from the different shots (respectively planes) were projected into the world coordinate system. If these measurements were error free and the sensor-to-tcp transformation correct, the measured planes would fit perfectly. The distances Δ and angles α between the measurements of a plane from different perspectives would then be zero. Table 1 shows the photogrammetric calibration[11] (PP) and the depth calibration (PD) results for the two TOF cameras. Obviously, the single photogrammetric calibration causes distance errors of 18mm to 35mm. The proposed error model proves to be valuable and the applied calibration nearly halves the errors. For both cameras a circular error and a signal propagation delay were detected. Within a limited working range the circular error is approximated by a linear function. As seen in

	SR-3000		**IFM O3D100**	
Type	Δ/mm	$\alpha/^\circ$	Δ/mm	$\alpha/^\circ$
PP[1]	17.76	3.48	27.71	3.14
PP[2]	18.71	3.94	35.90	3.36
PD[1]	10.32	1.58	11.90	1.48
PD[2]	3.86	1.58	9.05	1.37

Table 1 Verification on all taken shots of one smooth white wall from different perspectives signed with ([1]) and on the calibration shots signed with ([2]) (parameters taken from: PP - photogrammetric calibration, PD = depth calibration)

table 1, the results obtained from the O3D100 camera are nearly as good as those of the SR-3000. Due to the low resolution and the small amount of data points, which are used for the calibration of the O3D100 camera, there is a difference in translational error of about 6mm. In conclusion, the measurements show that a calibration step is necessary and helpful, but does not totally compensate all error sources.

7.2 Surface Reconstruction Results

The visual impression of the merged range images give a first qualitative statement on the outcomes. Further, a cube of known size is used, which allows to quantify the remaining measurement error. For the five cube faces and the platform the individual plane parameters (normal of plane and distance to origin of world coordinate system) are computed for each different view. The deviation of the distances and angles between those planes from the ground truth distance of 140mm is used as an indicator of the calibration and registration performance. The performance is evaluated in three categories: The distance and angle between opposite cube faces (opp.), the angle and distance of interleaved cube faces (int.) and the angle between adjacent cube faces (adj.). Firstly, depth images were merged by the

Cal.	Cube	**SR-3000**				**IFM O3D100**			
Type	Faces	Δ/mm	σ/mm	$\Delta/^\circ$	$\sigma/^\circ$	Δ/mm	σ/mm	$\Delta/^\circ$	$\sigma/^\circ$
PP	opp.	94.99	32.54	7.33	2.94	137.94	6.88	6.33	5.48
PP	int.	7.66	3.89	7.73	2.54	5.87	5.00	4.69	2.50
PP	adj.	-	-	86.48	2.28	-	-	82.36	3.83
PD	opp.	134.27	4.19	1.69	1.29	139.28	5.34	4.89	5.08
PD	int.	5.18	1.57	3.94	1.77	4.04	2.61	3.31	2.12
PD	adj.	-	-	87.86	1.73	-	-	84.11	3.44
PD (ICP)	opp.	133.15	3.87	2.50	1.71	138.59	3.81	4.82	5.17
PD (ICP)	int.	0.76	0.58	4.69	2.01	2.49	1.71	3.10	2.09
PD (ICP)	adj.	-	-	87.56	1.83	-	-	83.89	3.27

Table 2 Merged results of opposed (opp.) interleaved (int.) and adjacent (adj.) cube faces.

means of undistorting the depth images, correcting the depth values and projecting 3D points into the world coordinate system based on the calibration alone. For the SR-3000 camera the cube reconstruction with extrinsic depth calibration (PD) is highly enhanced with respect to photogrammetric calibration (PP) (see table 2).

However, the distance between opposite faces diverges from the ground truth by 6mm (45mm before). This shrinking of the object results from the mean calibration error of 10mm (see Table 1). The remaining distance error causes a lateral error and mainly affects the size of the merged object but not the orientation. In contrast to photogrammetric calibration (PP) the faces are well orientated, so that the angle decreased from 7° to 2°. For the O3D100 camera, the distance between opposite faces almost agrees with the ground truth. Despite an average calibration error of 12mm, a shrinking of the object is not observed at the same order of magnitude. This effect scales with the aperture angle, which is 34° in contrast to the 47.5° of the SR-3000.

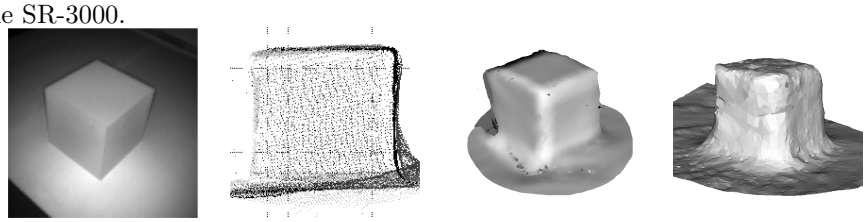

Figure 3 Cube surface reconstruction with both TOF cameras: SR-3000 (single intensity shot [left], merging point cloud and shaded object [2 × centre]), O3D100 (shaded object [right])

The ICP approach was used to relax errors among all measurements. Table 2 reports the performance indicators for the ICP application and shows a marginal improvement for all indicators. The highest gain is observed for the interleaved faces whose mean distances to each other decrease from 6.8mm to 2.2mm for the SR-3000 and from 4mm to 2.5m for the O3D100. The distance between the the opposite faces decreases, too. In case of the ICP algorithm the only constraint for the cube size is the width of a cube face in a single shot. Since the edge filtering truncates the widths of the cube faces by removing edges, the ICP algorithm can push the point clouds closer together than required. The reconstructed model shows a precision of 6mm up to 1mm in translation and 1.7° to 4.9° in rotation. The ICP further increases the consistency of the object with an accuracy of the interleaved cube faces up to 2.5mm in translation and 3.1° in rotation. Figure 4 presents the reconstruction results for an arbitrary model of a camel.

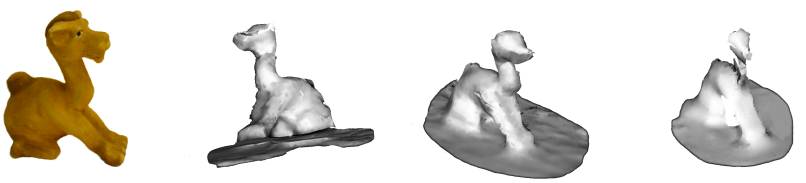

Figure 4 Surface reconstruction of a camel (digital image [left]) with both TOF cameras: SR-3000 (shaded object [2 × centre]) and O3D100 (shaded object [right]);

8 Conclusion

In this paper a surface reconstruction application using TOF cameras was presented. Two measures are important in order to reach high precision. First, a novel advantageous calibration method, which estimates the depth correction parameters as well as the eye-to-hand transformation simultaneously, was developed. This estimation simplifies the calibration requiring only distance measurements. The

second measure comprises necessary data processing, i.e. smoothing, edge filtering and pose relaxation. These measures have been illustrated by the reconstruction of a concrete example, a cube model. An overall mean precision and accuracy of approximately 3mm in translation and 3° in rotation has been achieved, that enables a simple merging of shots for surface reconstruction with high precision, especially by the means of magnitude and pose of the object, which is demanded for robotic interaction, i.e. visual servoing or grasping. A further ICP processing increases the consistency and accuracy respectively required by tracking or modelling applications.

References

[1] Besl, P. and McKay, N. (1992) 'A Method for Registration of 3-D Shapes', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, (February 1992) pp. 239-256.

[2] Bodenmueller, T. and Hirzinger, G. (2004) 'Online Surface Reconstruction From Unorganized 3D-Points For the DLR Hand-guided Scanner System', *2nd Symposium on 3D Data Processing, Visualization, Transmission*, Thessaloniki, Greece, 2004.

[3] CalLab 2005 and CalDe. Inst. of Robotics and Mechatronics, German Aerospace Center. [Online]. Available: http://www.robotic.dlr.de/callab/

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

[5] Kahlmann, T., Remondino, F., Ingensand, H. (2006) 'Calibration for increased accuracy of the range imaging camera SwissRanger'. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI, part 5, pp. 136-141, Dresden, Germany, 2006.

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

[7] Lindner, M. and Kolb, A. (2006) 'Lateral and Depth Calibration of PMD-Distance Sensors'. *In Proceedings of the ISCV, Second International Symposium on Advances in Visual Computing*, Lake Tahoe, USA, pp. 524-533, November 2006.

[8] Nüchter, A., Lingemann, K., Hertzberg, J. and Surmann H. (2006) '6D SLAM – Mapping Outdoor Environments', *International Workshop on Safty, Security and Rescue Robotics (SSRR '06)*, (CDROM Proceedings), Gaithersburg, Maryland, USA, August 2006.

[9] Ohno, K., Nomura, T. and Tadokoro, S. (2006) 'Real-Time Robot Trajectory Estimation and 3D Map Construction using 3D Camera', *In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Beijing, China.

[10] Sheh, R., Kadous, M. W. and Sammut, C. (2006) 'On building 3D maps using a Range camera: Applications to Rescue Robotics', *Technical Report UNSW-CSE-TR-0609*, UNSW, Sydney, Australia.

[11] Strobl, K. H. and Hirzinger, G. (2006) 'Optimal Hand-Eye Calibration', *In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Beijing, China, pp. 4647-4653, October 2006.

[12] Suppa M., Kielhfer S., Langwald J., Hacker F., Strobl K., and Hirzinger, G. (2007) 'The 3D-Modeller: A Multi-Purpose Vision Platform', . *International Conference on Robotics and Automation*, ICRA, Rome, Italy, April, 2007

[13] Surmann, H., Nüchter, A., and Hertzberg, J. (2003) 'An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments', *Journal Robotics and Autonomous Systems*, 45(3 - 4), pp. 181 - 198.

[14] Zhang, Z. (2000) 'A Flexible new Technique for Camera Calibration'. *IEEE Trans. on Pattern Analysis and Machine Intelligence* , vol. 22, no. 11, pp. 1330 - 1334, November 2000.