A Comprehensive System for 3D Modeling from Range Images Acquired from a 3D ToF Sensor

Agnes Swadzba[1], Bing Liu[2], Jochen Penne[3], Oliver Jesorsky[2], and Ralf Kompe[2]

[1] Faculty of Technology, Bielefeld University, 33594 Bielefeld, Germany
[2] Elektrobit Automotive Software, 91058 Erlangen, Germany
[3] Chair for Pattern Recognition, University Erlangen-Nuremberg, 91058 Erlangen, Germany

Abstract. Developing a system which generates a 3D representation of a whole scene is a difficult task. Several new technologies of 3D time-of-flight (ToF) imaging, which overcome various limitations of other 3D imaging systems, such as laser/radar/sonar scanners, structured light and stereo rigs have been developed in recent years. However, only limited work got published upon computer vision applications based on such ToF sensors. We present in this paper a new complete system for 3D modeling from a sequence of range images acquired during an arbitrary flight of a 3D ToF sensor. Comprehensive preprocessing steps are performed to improve the quality of range images. An initial estimate of the transformation between two 3D point clouds – computed from two consecutive range images – is achieved through feature extraction and tracking based on three kinds of images delivered by the 3D sensor. During the initial estimation, a RANSAC sampling algorithm is implemented to filter out outlier correspondences. At last the transformation is further optimized through registering the two 3D point clouds using a robust variation of the Iterative Closest Point (ICP) algorithm, the so-called Picky ICP. Extensive experimental results are provided in the paper and show the efficiency and robustness of the proposed system.

1 Introduction

Range imaging provides a direct way for acquiring 3D information from objects and scenes. For building a complete model of a 3D scene, one needs to acquire images from different points of view, register them and integrate the sequence of range images to obtain a single digital 3D surface representation of the scene.

A great deal of work has been published on 3D modeling from range images. A best-known algorithm for registering two 3D point clouds – computed from two consecutive range images – is the Iterative Closest Point (ICP) method [1]. A number of variations of the ICP has been proposed to improve its convergency criterion, to deal with partially overlapping cases, to remove outliers, and to reduce computational costs [2] [3] [4] [5]. A good initial relative pose estimate is required for ICP and its variations. This problem is usually solved by detecting and tracking distinctive features between consecutive range images, such as edges, lines, curves, or surface curvatures [6] [7] [8] [9]. Dorai gives an extensive overview of already existing reconstruction systems [10]. However, most of the

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

methods above deal with range images captured by specialized sensor hardware, e.g. laser/radar/sonar range scanners, structured light or stereo rigs. Such systems have various restrictions on the environment and require high development and computational costs. E.g. a laser scanner is not able to provide a matrix of 3D points in real time or stereo rigs need textures for computing 3D points.

The Photonic Mixer Device (PMD) technology enabling the real-time acquisition of a 2D matrix of depth values has been developed in recent years [11]. The PMD sensor measures the time-of-flight (ToF) of an infrared signal that is sent from the sensor to an object and reflected back to the sensor. The ToF is determined by the distance between an object and the optical center of the sensor. Another 3D sensor called Swiss Ranger, developed by the Swiss Center for Electronics and Microtechnology (CSEM) [12], is based on a similar technology as PMD. Nowadays, the frame rate of such 3D sensors is up to 30 frames per second.

Our system is able to generate a complete 3D surface representation of a scene using a sequence of frames acquired during an arbitrary camera flight of such a 3D ToF sensor. Due to the low resolution and low depth accuracy, existing 3D modeling techniques cannot work directly upon such 3D ToF sensor systems. The PMD sensor yields for each frame an intensity, an amplitude and a distance image. The amplitude image reflects the relative accuracy of the measurement of each pixel, and the 3D structure of the scene can be computed directly from the distances. A minimum amplitude filter is implemented using a threshold which is computed statistically from the image to remove noisy pixels, and a distance-adaptive algorithm is proposed to smooth the images (Section 2). Afterwards an initial estimate of the transformation between two 3D point clouds, computed from two consecutive frames, is achieved by feature extraction and tracking based on the three types of images. During the process, a RANSAC sampling algorithm is used to filter out outlier correspondences, to get a robust initial estimate (Section 3 – Coarse Registration). Then the transformation is further optimized by registering the two 3D point clouds using the Picky ICP algorithm [13](Section 3 – Fine Registration). The proposed system is tested with two 3D sensors: PMD®19k with a resolution of 120×160 pixels and PMD®1k-S with a resolution of 64×16 pixels. The efficiency and the robustness of the system are evaluated with extensive experimental results (Section 4).

2 Preprocessing

This section describes in detail each preprocessing step as shown in Figure 1. The PMD cameras used provide three images per frame containing distance, intensity, and amplitude values. The distance value describes the distance between the optical center of the camera and a certain world point. The corresponding intensity value is the grayscale value of this world point and the amplitude value accords to the amplitude of the signal measured by the PMD sensor. In the following the smoothing of the three output images, the rejection of pixels with bad distance measurements and the computing of 3D coordinates will be described.

Smoothing of the Intensity, Amplitude and Distance Images The principal sources of noise in digital images arise during image acquisition influenced by

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

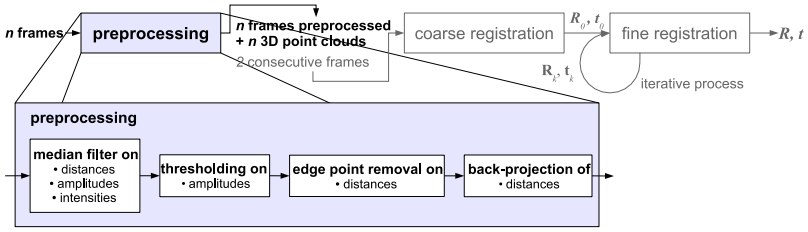


Fig. 1. Overview of the preprocessing steps.

environmental conditions [14]. Especially the PMD sensor, which works with active illumination, has to deal with scattering effects overlaying the distance, intensity, and amplitude values with noise. Gabriel showed that this noise overlaying the distance values follows a Gaussian distribution [15]. A commonly-used approach to suppress such distributed image noise is to apply a median filter to the image. Smoothing of images with preserving of edges and corners for further image processing, e.g. feature tracking, is an advantage of the median filter.

However, a 3D sensor provides also a distance value for each image pixel, which refers to an amount of 3D area a pixel corresponds to. Obviously, the larger the distance value, the larger is the area represented by the pixel. The physical size of the sensor is significant especially for the PMD sensor: 0.003mm^2 for PMD®19k and 0.0016mm^2 for PMD®1k-S. Some numbers for the areas covered at different distances are given in Table 1 (left).

Therefore, a *distance-adaptive median filter* is proposed in this paper to smooth the distance, amplitude, and intensity images, respectively. The median filter operates on each pixel with different sizes of filter mask according to its distance value. Generally, pixels with larger distance value are filtered with smaller filter masks, and vice versa, so that significant structures at large distances are not blurred, and at the same time, noise at small distances can be removed. Table 1 (right) lists three distance ranges and the corresponding filter mask sizes – 3×3, 5×5, and 7×7 – that are used in the system proposed.

dis (mm)	area (mm^2)	
	1k-S	19k
0	0.033	0.0016
1000	127.88	11.111
2000	511.52	44.444
3000	1150.9	100

area (mm^2)	dis (mm)		filter size
	1k-S	19k	
up to 100	884.30	3000	7×7
up to 400	1768.6	6000	5×5
from 400	> 1768.6	> 6000	3×3

Table 1. (left) Size of the area projected on one PMD pixel of PMD®1k-S (1k-S) and of PMD®19k (19k) depending on the distance (dis) between the image plane and the area in 3D. (right) Determination of the distance threshold for the mask sizes of the distance-adaptive median filter.

Thresholding on the Amplitude Values The distance measurement of a PMD sensor is based on the measurement of the phase delay between the signal sent and received. The signal detected is characterized by a phase angle equal to the phase delay measured and by a certain amplitude value. This amplitude characterizes the reliability of the measurement of the pixel. It gives information

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

about the amount of light reflected by an object. The less light is reflected, the less amount of light is detected and the smaller is the amplitude of the signal.

Signals from badly reflecting surfaces will be more influenced by noise, so that the distances measured will probably be wrong. Our approach introduces a threshold θ_{amp}. All distance measurements with amplitudes below this threshold are removed. The threshold varies for each frame. It is a fraction of the mean value of all amplitudes of one frame.

Edge Point Removal This preprocessing step deals with edge points, which arise in the case when rays from the foreground and the background hit the same pixel simultaneously. The measurement returned by the pixel will be a distance somewhere between the foreground and the background.

The idea for filtering such points, is to consider the pixels in the 8-neighborhood $\mathcal{N}_c$ of the current pixel $\boldsymbol{\chi}_c$. If there are at least two pixels $\boldsymbol{\chi}_i \in \mathcal{N}_c$ with distances to $\boldsymbol{\chi}_c$ smaller than a certain threshold θ_{edge} then the current pixel $\boldsymbol{\chi}_c$ has enough *near* neighbors. Otherwise, this pixel $\boldsymbol{\chi}_c$ is a so-called edge point as described above. Such pixels are removed from the point cloud.

3D Coordinates by Back-projection The given depth values are distances between the optical center of the camera and a world point in 3D. The 3D coordinates are generated out of these distances with regard to a 3D camera coordinate system. With the assumption of ideal perspective projection, the PMD camera works as a pinhole camera. Hence, the 3D coordinates can be computed from the distances measured using the mathematical idea of ray proportions in triangles.

3 Registration of Consecutive Frames

The main task of the framework presented here is to estimate the motion of the camera made during the data acquisition only based on the distance, amplitude, and intensity values of the PMD camera. As a set of frames $\boldsymbol{F}_0$, $\boldsymbol{F}_1$, ... is processed sequentially, it is sufficient to estimate the rigid transformations $(\boldsymbol{R}_i, \boldsymbol{t}_i)$ between all pairs of consecutive frames $(\boldsymbol{F}_{i-1},\ \boldsymbol{F}_i)_{i=1,\ \ldots}$ with $\boldsymbol{R}_i \in \mathbb{R}^{3\times 3}$ and $\boldsymbol{t}_i \in \mathbb{R}^3$. Reconstruction of a whole sequence is reduced to the problem of estimating the optimal motion between two frames $\mathcal{A}$ $(= \boldsymbol{F}_i)$ and $\mathcal{B}$ $(= \boldsymbol{F}_{i-1})$.

A typical sequence of PMD frames representing a scene recorded, consists of several point clouds with overlapping regions. The basic idea of registering such frames is to find pairs of corresponding points. The most popular algorithm to realize the concept is the *Iterative Closest Point* (ICP) algorithm [1]. The algorithm determines the point correspondences $\mathcal{C} = \{(i,j) \mid \boldsymbol{a}_i \in \mathcal{A} \text{ and } \boldsymbol{b}_j \in \mathcal{B}\}$ by searching for each point $\boldsymbol{a}_i$ for the nearest neighbor $\boldsymbol{b}_j$ concerning the Euclidean distance. Therefore, the ICP can only be used to estimate small motions. It is just applicable for *fine registration*. The ICP algorithm needs an initializing transformation $(\boldsymbol{R}_0, \boldsymbol{t}_0)$ to deal with bigger motion between two frames. This transformation results from the so-called *coarse registration* of two frames.

For a given set of point correspondences $\mathcal{C}$ the optimal transformation $(\boldsymbol{R}, \boldsymbol{t})$ is computed such that the mean square objective function

$$f_{\min}(\boldsymbol{R}, \boldsymbol{t}) = \frac{1}{|\mathcal{C}|} \sum_{(i,j)\in\mathcal{C}} \| \boldsymbol{b}_j - \boldsymbol{R}\boldsymbol{a}_i - \boldsymbol{t} \|^2 \tag{1}$$

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

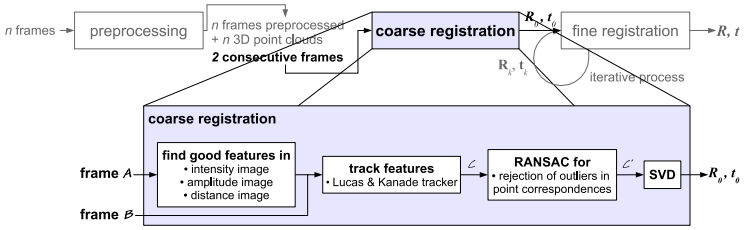


Fig. 2. Details of the coarse registration in the 3D reconstruction chain.

is minimized [16]. A *Singular Value Decomposition* (SVD) [17] using the special case treating by Umeyama [18] solves this optimization problem in our system.

Coarse Registration The coarse registration provides a good initial transformation $(\boldsymbol{R}_0, \boldsymbol{t}_0)$ which will support the convergence of the ICP algorithm to a global minimum. It is essential to find valid point correspondences, which means that a pair of points $(\boldsymbol{a}_i, \boldsymbol{b}_j) \in \mathcal{C}$ corresponds to the same point in the real 3D world. Figure 2 points out the four main steps of the coarse registration between two frames $\mathcal{A}$ and $\mathcal{B}$. The three output matrices containing intensity, distance, and amplitude values can be interpreted as normal grayscale images. Consequently, the coarse registration here is using methods dealing with 2D data.

The most important step is the extraction of robust and distinctive feature points. Significant structures, e.g. edges and corners, are suitable for features and are determined using the *structure tensor* operator [19]. In this way a set of distinctive features is provided in frame $\mathcal{A}$.

In the next step the new positions of these features in frame $\mathcal{B}$ are detected. Movement of feature positions results from the camera motion during the data acquisition. *Optical flow* algorithms are able to cope with this motion [20]. The estimation of optical flow in image sequences is mostly based on differential methods, which can be classified into global [21] and local methods. The local method of Lucas and Kanade [22] is used here, which assumes that the optical flow is constant within a certain neighborhood $\mathcal{N}$.

The result of the feature extraction and tracking is a set of point correspondences $\mathcal{C}$. As feature tracking via optical flow is based on solving an optimization problem that can result in wrong point correspondences, such outlier pairs have a disturbing influence on computing the initial transformation $(\boldsymbol{R}_0, \boldsymbol{t}_0)$ via SVD. Therefore, a method based on the *RANdom SAmple Consensus* (RANSAC) algorithm of Fischler and Bolles [23] is applied to reject outliers. As three non-linear point correspondences are sufficient for computing a transformation in 3D, the RANSAC algorithm chooses as an initial data set three point pairs randomly from the set $\mathcal{C}$ and enlarges this set with consistent point pairs if possible.

Fine Registration The task of the fine registration is to optimize the initial transformation $(\boldsymbol{R}_0, \boldsymbol{t}_0)$ computed during the coarse registration. Besl and McKay [1] developed the so-called *Iterative Closest Point* (ICP) algorithm to deal with this problem. For each point $\boldsymbol{a}_i$ of frame $\mathcal{A}$, the corresponding point $\boldsymbol{b}_j$ of frame $\mathcal{B}$ is the nearest neighbor in the Euclidean sense.

As applying the nearest neighbor operation for each point of frame $\mathcal{A}$ in each iteration step k of the ICP algorithm is one of the most time consuming

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

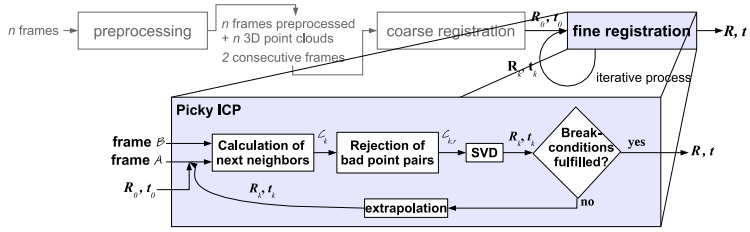


Fig. 3. General overview of the processing steps during the fine registration.

operations in the system, it is necessary to accelerate this search. The aim is to find in a fast way a small region of interest which contains a small number of potential nearest neighbor candidates. We use here the commonly accepted assumptions of localizing the principal point in the middle of the image plane and of using the details on focal length and pixel dimension given by PMDTec [24]. Using these intrinsic parameters of the camera a search method based on *reverse calibration* [25] is developed. There, a given point is projected on the image plane of the camera. The resulting pixel forms the center of a region of interest (ROI) in the image plane. The 3D points corresponding to the pixels within the ROI are potential nearest neighbor candidates. At last, the 3D point which is nearest to the given point is selected as the nearest neighbor.

It is obvious that a set of point correspondences generated via a nearest neighbor search contains a lot of invalid point pairs. As the standard ICP algorithm is sensitive to noise, a robust version of the ICP, the *Picky* ICP algorithm, is used here [13]. It introduces methods for rejecting bad point correspondences. First, point pairs with a distance between them greater than a certain threshold θ_{dis} $(= 2.5 \cdot 1.4826 \cdot \text{median}_{(i,j)\in\mathcal{C}}\{\| \boldsymbol{b}_j - \boldsymbol{a}_i \|\})$ [26] are removed from the set. Second, in the case of partially overlapping point clouds several points can have the same nearest neighbor point. All these point pairs are rejected apart from the pair with the smallest Euclidean distance.

Leaving out pairs causes that the convergence proof of the ICP by Besl and McKay is no more valid. Consequently, additional break conditions are necessary to ensure the termination of the Picky ICP algorithm. The algorithm terminates, if there are no big changes in the adjustment error $e = \frac{1}{|\mathcal{C}|}\sum_{(i,j)\in\mathcal{C}} \| \boldsymbol{b}_j - \boldsymbol{a}_i \|^2$ or in the computed transformation between the previous and the current iteration step or if the number of maximal iteration steps is exceeded.

Figure 3 summarizes the registration steps described. In the case when no break condition is fulfilled, an extrapolation of rotation $\boldsymbol{R}_k$ and translation $\boldsymbol{t}_k$ as already proposed by Besl and McKay with the extensions of Rusinkiewicz [2] and Simon [27] is applied to accelerate the convergence of the algorithm.

4 Experiments and Discussion

The experiments with the system are performed using the cameras PMD®19k (120×160 pixels) and PMD®1k-S (64×16 pixels). The system runs on a Pentium 4 with CPU 3.00GHz and 1.00GB RAM. A ground truth – the exact position of the PMD camera – for comparison with the motion estimated during the registration is determined by the tracking system as shown in Figure 4. The

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

camera is labelled with the target displayed in the figured. Figures 5(a) and 5(c) show two test scenes for reconstruction.

The accuracy of the estimated motion is determined by comparing the real motion $(\boldsymbol{R}_d, \boldsymbol{t}_d)$ to the estimated $(\boldsymbol{R}_e, \boldsymbol{t}_e)$ by computing a mean error ε. A given point set $\mathcal{P} = \{\boldsymbol{p}_i\}$ is transformed to the sets $\mathcal{P}^d = \{\boldsymbol{p}_i^d = \boldsymbol{R}_d\boldsymbol{p}_i + \boldsymbol{t}_d\}$ and $\mathcal{P}^e = \{\boldsymbol{p}_i^e = \boldsymbol{R}_e\boldsymbol{p}_i + \boldsymbol{t}_e\}$ with

$$\varepsilon = \frac{1}{|\mathcal{P}|}\sum_i \|\boldsymbol{p}_i^e - \boldsymbol{p}_i^d\| \tag{2}$$

defining the error between the resulting point sets. The smaller ε, the closer is the estimated motion to the real one.

Fig. 4. Tracking system which determines the position of the camera within a global world coordinate system. The camera is labelled with a target.

In this work several modules of the system are tested for their performance and accuracy. First, the module for searching the nearest neighbor is examined. The reconstruction results using the search method which determines a region of interest (f_{ROI}) are compared to results of the reconstruction considering all points of a point cloud when determining the nearest neighbor (f_{all}) for every 3D point. As can be seen in Table 2 (left) only in few cases the error increases when using a subset of points for determining the nearest neighbor. Mostly, the error stays nearly unchanged. Consequently, the method proposed for the nearest neighbor search is suitable, since this method accelerates the search by a factor of round 6 as shown in Table 2 (right).

	1k-S				19k			
	f_{all}		f_{ROI}		f_{all}		f_{ROI}	
	$\bar{e}$	σ	$\bar{e}$	σ	$\bar{e}$	σ	$\bar{e}$	σ
scene 1	10	5	15	22	59	44	29	22
scene 2	25	25	26	26	39	76	85	107

	CPU time (ms)	
	1k-S	19k
f_{all}	15	3041
f_{ROI}	2	242

Table 2. (left) Evaluation of both methods for the nearest neighbor search – considering all points of a frame (f_{all}) or only few points of a certain region of interest (f_{ROI}) – on sequences of PMD®1k-S (1k-S) and PMD®19k (19k). The mean error ($\bar{e}$ in mm) and its standard deviation (σ in mm) are displayed. (right) Computation times of the different methods for the nearest neighbor search.

Next, the influence of the preprocessing steps on the 3D reconstruction is examined. The results are shown in Table 3 (left). The distance-adaptive median filter leads to small improvements (at least 4mm) in the reconstruction results of frames from PMD®19k compared to a normal 3×3 median filter. For PMD®1k-S no difference between both filters can be observed as in both scenes the adaptive filter returns the 3×3 median filter for each pixel. The edge point removal has the biggest influence on the results of the small camera, since the point set is very small and can be strongly influenced by a few bad points. For the amplitude filtering two threshold values are tested: $\theta_{\text{amp}} = \frac{1}{3}\bar{\delta}$ and $\theta_{\text{amp}} = \frac{2}{3}\bar{\delta}$, where $\bar{\delta}$ is the mean value of all amplitudes of the current frame. Reconstruction

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

f_{med}	f_{rm}	f_{amp}	$\bar{e}$ (mm) 1k-S sc. 1	1k-S sc. 2	19k sc. 1	19k sc. 2
no	no	no	9	24	100	100
adapt	yes	$\frac{1}{3}\bar{\delta}$	11	19	29	86
3×3	yes	$\frac{1}{3}\bar{\delta}$	10	20	90	90
adapt	yes	$\frac{2}{3}\bar{\delta}$	18	25	29	86
adapt	no	$\frac{1}{3}\bar{\delta}$	19	21	30	87

modalities for f_{crs}	f_{RANSAC}	$\bar{e}$ (mm) 1k-S sc. 1	1k-S sc. 2	19k sc. 1	19k sc. 2
Int, Amp, Dis	yes	9	19	29	86
Int, Amp, Dis	no	11	22	27	85
Int	yes	12	19	45	86
Int	no	13	22	46	85
no	no	19	22	82	82

Table 3. (left) Influence of the different preprocessing steps – smoothing with median filters (f_{med}), rejecting of bad point pairs via edge point removal (f_{rm}), and amplitude filtering (f_{amp}) – on the reconstruction results of sequences of scene (sc.) 1 and 2 recorded with PMD®19k (19k) and PMD®1k-S (1k-S). (right) Influence on the reconstruction results of the coarse registration (f_{crs}) used with different image modalities and of the rejection of invalid point pairs via RANSAC (f_{RANSAC}).

results of PMD®19k do not show any difference between both values. But for PMD®1k-S $\frac{2}{3}\bar{\delta}$ produces worse results, because too many points are rejected which are necessary for a camera with low resolution.

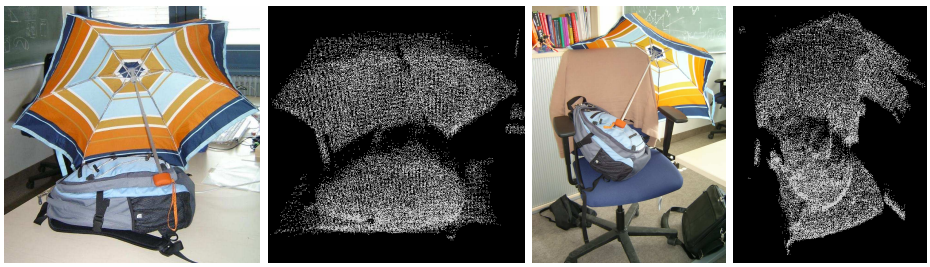
The last test, as shown in Table 3 (right), examine the influence of the coarse registration on the reconstruction results using different image modalities and point rejection via RANSAC for the registration. In most cases coarse registration performed on the three (intensity, distance, and amplitude) images together with bad point pair rejection via RANSAC provide the best reconstruction results. Even when the coarse registration does not provide an improvement of the results it accelerates the whole reconstruction in any case. For example, registering two frames of scene 1 using the PMD®19k takes about 39s without coarse registration and 11s using the computed initial transformation. In Figures 5(b) and 5(d) reconstruction results of sequences of scene 1 and 2 containing 20 frames acquired with PMD®19k are presented.

To summarize the results it can be stated that the distance-adaptive median filter is most suitable for smoothing during preprocessing. The coarse registration using all modalities and bad point pair rejection via RANSAC improves and accelerates the whole reconstruction process. Unfortunately, the range of scenes this reconstruction framework can deal with is limited. Scenes with huge homogeneous areas in distance and intensity values are difficult to register as there are no or not enough distinctive features that can be used for coarse and fine registration.

5 Conclusion and Outlook

In this paper a fully automatic system for 3D reconstruction of data from a 3D ToF sensor is proposed. Since all frames of the input sequences are processed sequentially, the entire problem of 3D reconstruction is reduced to a problem of registering two consecutive frames. The computed optimal transformation estimates the motion of the camera between the corresponding camera positions in 3D space. The system consists of preprocessing steps for the three (intensity, distance, and amplitude) images per frame, where these images are smoothed with

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de



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

Fig. 5. Digital photos of scene 1 (a) and scene 2 (c) are shown here. Additionally, reconstruction results of 20 frames recorded by PMD®19k of scene 1 (b) and scene 2 (d) are presented.

a distance-adaptive median filter and invalid 3D points are rejected. A coarse registration computes an initial motion estimate via detecting certain features and tracking them using optical flow. As some of the point correspondences may be illposed, a RANSAC based method is applied to reject bad point pairs. Finally, a fine registration optimizes the estimate using the Picky ICP. In future work it will be interesting to extend the coarse registration from 2D data to features in 3D and to test the registration on more different scenes. Additionally, the main aim will be to realize 3D reconstruction in real-time.

Acknowledgement.

I want to thank Elektrobit Automotive Software (former 3Soft GmbH) for giving me the chance to write my Diploma thesis, which forms the background for this paper, in cooperation with them. Additionally, I want to thank Prof. Hornegger from LME of the University Erlangen-Nuremberg for enabling this cooperation.

References

1. Besl, P., McKay, N.: A method for registration of 3D shapes. In: IEEE Trans. PAMI. Volume 14. (1992) 239–256
2. Rusinkiewicz, S., Levoy, M.: Efficient variants of the ICP algorithm. In: Proc. of the 3rd Intl. Conf. on 3D Digital Imaging and Modeling. Volume 00. (2001) 145–152
3. Pajdla, T., Gool, L.V.: Matching of 3D curves using semi-differential invariants. In: ICCV '95: Proc. of the 5th Int. Conf. on Computer Vision. (1995) 390
4. Sharp, G.C., Lee, S.W., Wehe, D.K.: Invariant features and the registration of rigid bodies. In: IEEE Proc. ICRA. Volume 2. (1999) 932–937
5. Dalley, G., Flynn, P.: Range image registration: A software platform and empirical evaluation. In: Proc. of the 3rd Intl. Conf. on 3D Digital Imaging and Modeling. Volume 00. (2001) 246–253
6. Chua, C.S., Jarvis, R.A.: 3D free-form surface registration and object recognition. Int. Journal of Computer Vision **17**(1) (1996) 77–99
7. Sappa, A., Restrepo-Specht, A., Devy, M.: Range image registration by using an edge-based representation. In: Proc. of the 9th Int'l Symp. Intelligent Robotic Systems. (2001) 167–176

8. Wyngaerd, J.V., Gool, L.V.: Automatic crude patch registration: Toward automatic 3D model building. In: Comput. Vis. Image Underst. Volume 87. (2002) 8–26
9. Yamany, S.M., Farag, A.A.: Free-form surface registration using surface signatures. In: ICCV '99: Proc. of the Int. Conf. on Computer Vision. Volume 2. (1999) 1098–1104
10. Dorai, C., Wang, G., Jain, A.K., Mercer, C.: Registration and integration of multiple object views for 3D model construction. IEEE Trans. PAMI **20**(1) (1998) 83–89
11. Xu, Z., Schwarte, R., Heinol, H., Buxbaum, B., Ringbeck, T.: Smart pixel – Photometric Mixer Device (PMD) / new system concept of a 3D-imaging-on-a-chip. In: 5th Int. Conf. on Mechatronics and Machine Vision in Practice. (1998) 259–264
12. Weingarten, J., Gruener, G., Siegwart, R.: A state-of-the-art 3D sensor for robot navigation. In: Proc. of IROS 2004. (2004)
13. Zinßer, T., Schmidt, J., Niemann, H.: A refined ICP algorithm for robust 3D correspondence estimation. In: Proc. of the IEEE Int. Conf. on Image Processing. Volume 2. (2003) 695–698
14. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. (2001)
15. Gabriel, P.: Passenger classification for airbag control with 3D camera technology. Master's thesis, Universität Erlangen-Nürnberg (2006)
16. Schönemann, P.: A generalized solution of the orthogonal procrustes problem. Psychometrika **31**(1) (1966) University of North Carolina.
17. Lorusso, A., Eggert, D.W., Fisher, R.B.: A comparison of four algorithms for estimating 3D rigid transformations. Machine Vision and Applications **9**(5) (1997) 272–290
18. Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns. In: IEEE Trans. PAMI. Volume 13. (1991) 376–380
19. Förstner, W., Gülch, E.: A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: ISPRS Intercommission Workshop. (1987) 281–305
20. Gibson, J.: The perception of the visual world. Riverside Press (1950)
21. Horn, B.K., Schunck, B.G.: Determining optical flow. In: Artificial Intelligence. Volume 17. (1981) 185–204
22. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Proc. of the 7th Int. Joint Conf. on Artificial Intelligence (IJCAI '81). (1981) 674–679
23. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In: Commun. ACM. Volume 24. (1981) 381–395
24. Ringbeck, T., Lange, R., Hagebeuker, B.: PMD[vision]19k 3D video range camera preliminary (2005) http://www.pmdtec.de/; 12.06.06.
25. Blais, G., Levine, M.D.: Registering multiview range data to create 3D computer objects. In: IEEE Trans. PAMI. Volume 17. (1995) 820–824
26. Masuda, T., Yokoya, N.: A robust method for registration and segmentation of multiple range images. In: Computer Vision and Image Understanding. Volume 10. (1995) 295–307
27. Simon, D.: Fast and accurate shape-based registration. PhD thesis, Carnegie Mellon University Pittsburgh, Robotics Institute (1996)

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de