

3D time-of-flight cameras for mobile robotics

Stefan May, Björn Werner, Hartmut Surmann and Kai Pervölz
Fraunhofer Institute for Autonomous Intelligent Systems (AIS)
Schloss Birlinghoven
D-53754 Sankt Augustin, Germany
{smay, surmann, pervoelz}@ais.fraunhofer.de

Abstract—Recently developed 3D time-of-flight cameras have an enormous potential for mobile robotic applications in particular for mapping and navigation tasks. This paper presents a new approach for online adaptation of different camera parameters to environment dynamics. These adaptations allow the usage of state-of-the-art 3D cameras reliably in real world environments and enables capturing of 3D scenes with up to 30 frames per seconds. The performance of the approach is evaluated with respect to different robotic specific tasks in changing environments.¹

I. INTRODUCTION

Automatic environment sensing and modeling is a fundamental scientific issue in mobile robotics regarding the capacity to interact with three-dimensional real world environments. The problem domain of autonomous robotics comprises two major tasks with different key aspects. The first one covers exploration issues while creating accurate three dimensional maps. Here, high resolutions with precise 3D data as well as fast and accurate matching algorithms are required to create consistent scenes. The second task covers the question of exploration and navigation in known and unknown terrains. Real-time 3D computation of the scene in the moving direction of a robot is required to ensure obstacle avoidance, whereas the precision is secondary. The real-time capability is also mandatory for mapping and surveying tasks if environment dynamics are considered. The enhancements of photonic mixer devices, that enable 3D image grabbing within a few milliseconds give an important impulse in visual 3D sensing. Today, 3D cameras with a resolution of round about 20 thousand pixels and a frame rate with up to 30 frames per second are available [1][2]. Beside the high data rate its low weight and small size makes it a very interesting sensor for the mobile robotics community. But similar to passive visual sensors, this type of camera is very fragile to changing lighting conditions which leads not only to imprecise data but also to completely wrong measurement data. These errors are influenced by the physical properties of the sensor as well as by the environment conditions. For robotic issues, primarily the environmental influences are important since the real world is unpredictable. Without algorithms for online environment-adaptation and data preprocessing this sensor is not suitable for autonomous mobile robotic tasks.

This paper is structured as follows. In the next sub-section a brief state-of-the-art of 3D sensors for mobile robotics is

given. Section II briefly describes the CSEM Swiss Ranger SR-2 camera [3] and its required pixel-wise calibration for the experimental setup. In section III the performed experiments and the resulting approach for online calibration is described. Section IV discusses the results and concludes the paper.



Fig. 1. The Swiss Ranger SR-2 3D camera mounted on a pan-tilt unit.

A. State of the art

Since the problem of sensing the spatial properties of the environment is fundamental for autonomous robotics, many different groups have been working in this field for a long time. Today the most common techniques for 3D sensing are CCD- or CMOS-camera, laser scanner or recently 3D time-of-flight camera based. Due to its physical properties, the resolution and precision of ultra sonic sensors have been shown to be not suitable for fast and save interaction of autonomous system with its environment.

As CCD- or CMOS-camera based approaches to 3D robot vision mainly stereo cameras or structure from motion techniques are known [4], [5]. Both have difficulties providing reliable navigation or mapping information for a mobile robot in real-time and like all passive visual sensors, they are difficult to handle in real world environments with changing lighting conditions. In addition to that, some groups try to solve 3D modeling by using a planar 2D laser scanner and cameras, e.g., in [6].

A few other groups use highly accurate, expensive 3D laser scanners [7], [8], [9]. The RESOLV project aimed at modeling interiors for virtual reality and tele-presence [9]. They used a RIEGL laser range finder on robots and the ICP algorithm for scan matching [10]. The AVENUE project develops a robot for modeling urban environments [7], using a CYRAX scanner and a feature-based scan matching approach for registering

¹A video, showing the performance of the approach, is available on: <http://www.ais.fraunhofer.de/~surmann/videos/iros2006.mpg>

the 3D scans. Nevertheless, in their recent work they do not use data of the laser scanner in the robot control architecture for localization [8]. The group of M. Hebert has reconstructed environments using the Zoller+Fröhlich laser scanner and aims to build 3D models without initial position estimates, i.e., without odometry information [11]. The usability of those systems on mobile robots is limited because of their size, weight and power consumption.

Other 3D laser scanner approaches are based on 2D laser range finders, actually designed for industrial application, which are expanded by an additional rotation axis for acquiring 3D information. In [12] Wulf et al. analyzed scanning setups with a SICK 2D laser range finder and different rotation axis. Rotating the scanner around the horizontal axis is very common and used by, e.g., Surmann et al. [13]. To receive consistent 3D data with these methods, normally the data acquisition is performed in a stop-scan-go mode which means that the robots does not move during the 3D scanning. Recently, different groups employ continuous rotating laser range finders for acquiring 3D data [14], [15]. Wulf et al. let the scanner rotate around the vertical axis and acquire 3D data while moving, thus the quality of the resulting map crucially depends on the pose estimate that is given by inertial sensors, i.e., gyros. In combination with recently developed very small 2D laser scanners [16] these techniques allow to utilize 3D spatial information on small mobile robot systems.

Instead of using rotating 2D scanners, which yield consistent 3D scans in the first place, some groups have attempted to build 3D volumetric representations of environments with 2D laser range finders. Thrun et al. [17], Früh et al. [18] and Zhao et al. [19] use two 2D laser range finders for acquiring 3D data. One laser scanner is mounted horizontally, the other vertically. The latter one grabs a vertical scan line which is transformed into 3D points based on the current robot pose. Since the vertical scanner is not able to scan sides of objects, Zhao et al. use two additional, vertically mounted 2D scanners, shifted by 45° to reduce occlusions [19]. The horizontal scanner is used to compute the robot pose. The precision of 3D data points depends on that pose and on the precision of the scanner.

A very new and promising technique are the already mentioned 3D cameras, which are based on the photon mixer device technology [1], [2]. First approaches in robot navigation with an evaluation prototype of a Swiss Ranger SR-2 were presented by Weingarten et al. in [20]. They have shown, that obstacle avoidance with the Swiss Ranger has many advantages in contrast to a 2D laser scanner. In comparison to laser scanner based 3D sensors the SR-2 provide 3D information with a framerate of about 30 fps and thus is much faster. But they also pointed out, that camera calibration and data preprocessing were necessary to get stable sensor data. Outside of the robotics community Oliver Gut focused on surveying and mapping tasks with the SR-2, where high precision is required, but he has shown that it is not achieved with this current version of the 3D camera [21]. He uncovered several erroneous influences based on environmental or sensorical influences.

II. THE SWISS RANGER SR-2 CAMERA

Our experimental setup consisted of a Swiss Ranger SR-2 device connected via the USB 2.0 interface to a workstation running SuSE Linux 9.3. The swiss ranger device was mounted on a rotatable rack using a servo motor, which was used to adjust the device to several positions. Concatenating images while pivoting the device enabled a virtual 180 degree view.

A. Inside the Swiss Ranger SR-2

The camera belongs to the group of active sensors. It uses the phase-shift principle to determine distances. While the environment is being illuminated with infrared flashes, the reflected light is measured by a CCD/CMOS sensor. It provides amplitude data, intensity data and distance data, which are weakly addicted to each other. An image can formally be defined as: Dependencies between these data values will be investigated in detail in section III. Amplitude data represents the incoming waves amplitude, intensity its offset and distance its phase shift. For a detailed description of the measurement principle, please refer to [20]. The camera comes with a resolution of 124x160 pixel. All measurements are being organized by a FPGA, which provides an USB interface to access the data values. The FPGA can be configured setting one or more of its eleven registers. The most important register concerns the adjustment of integration time. It has a range from 1 to 255, which are multiple of $255 \mu\text{s}$. Finding the optimal value will be investigated in section III-A.

B. Camera calibration

A per-pixel calibration is best to enlarge accuracy. The camera must be mounted in a defined distance towards a white smooth wall. First, some captures has to be taken to ensure the camera to be at the right running temperature. The calibration run needs an optimal adjusted integration time. Oversaturation will falsify measurements. The distance offset register is set to 0 during calibration. We propose to determine a calibration matrix in polar coordinates to simplify measurement corrections. This matrix is used later for distance corrections by subtracting related measurement values. Because of the cameras phase-shift principle, you run into a "modulo 2π problem". To avoid negative values, that can appear after calibration matrix subtraction, we propose the following correction method.

$$l_{i,j} = (m_{i,j} - o_{i,j} + r_e) \bmod r_e, \quad (1)$$

where i and j are row and column indices, $m_{i,j}$ the measured and $l_{i,j}$ the corrected polar distance value of its related pixel. $o_{i,j}$ describes the pixel specific offset value and r_e the effective range of sensor (7.5 m).

III. EXPERIMENTS AND RESULTS

Weingarten et al. used an empirically found solution to suppress noise by decomposing the space into cells [20]. A cell with a minimum number of data points is to be considered occupied. This simplification fits well for navigation tasks but will lack for mapping issues. First, we analyzed inaccurate

data points to verify the conditions under which they appear. The examinations of Gut [21] provides an informative basis. It focuses surveying and mapping tasks, where high precision are required. The most interesting irritating effect is that of scattering light on near objects. As already mentioned above, the provided data values are weakly addicted to each other. Intensity information of an object, for example, depends on its distance, its alignment in relation to the sensor and its surface properties, like color and texture. Amplitude and intensity values allow to predict the accuracy of distance values. Some test scenarios should represent common environments a robot will typically act in and therefore include different compositions of objects to demonstrate their influences.

A. Setting up the integration time

First of all, the integration time has to be set up. It is one of the most important parameters to get stable data. It has to be adjusted in relation to each scene, otherwise too high saturation could cause erroneous effects. Figure 2(a) and 2(b) show the same scene taken with the Swiss Ranger device running at two different integration times. To illustrate 3D data in this paper, we chose a false color representation using the HSV model starting at 60 degrees in reverse order to 240 degrees. Near objects will appear yellow, far objects blue. Objects with medium distances appear red or magenta.

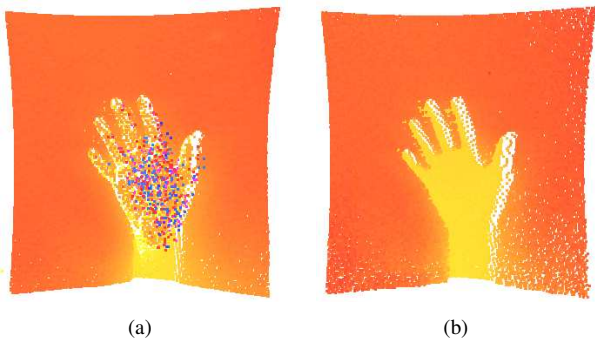


Fig. 2. Two sample measurements using different integration times. 2(a) Measurement of a close hand with an integration time of 15 ms, what was definitely too high for that scene. 2(b) Measurement of a close hand with an integration time of 4 ms, that fits better for that scene.

The noisy area is caused by a near bright object, i.e. a close hand in this case. The second measurement (with smaller integration time) provides better results, which means, that less noise superposes the image. A series of measurements to determine the dependency between integration time and data values of the Swiss Ranger device can be seen in figure 3. Each curve represents the determination of data values from the same scene varying the integration time. The significant range with constant distance data is also represented by high mean amplitude values. The gap between the intensity graph and its linearisation as well between the amplitude graph and its linearisation determines the degree of oversaturation. Only values in the linearly parts features accurate distances.

Oggier et al. already reported the development of an algorithm to automatically select the best integration time for

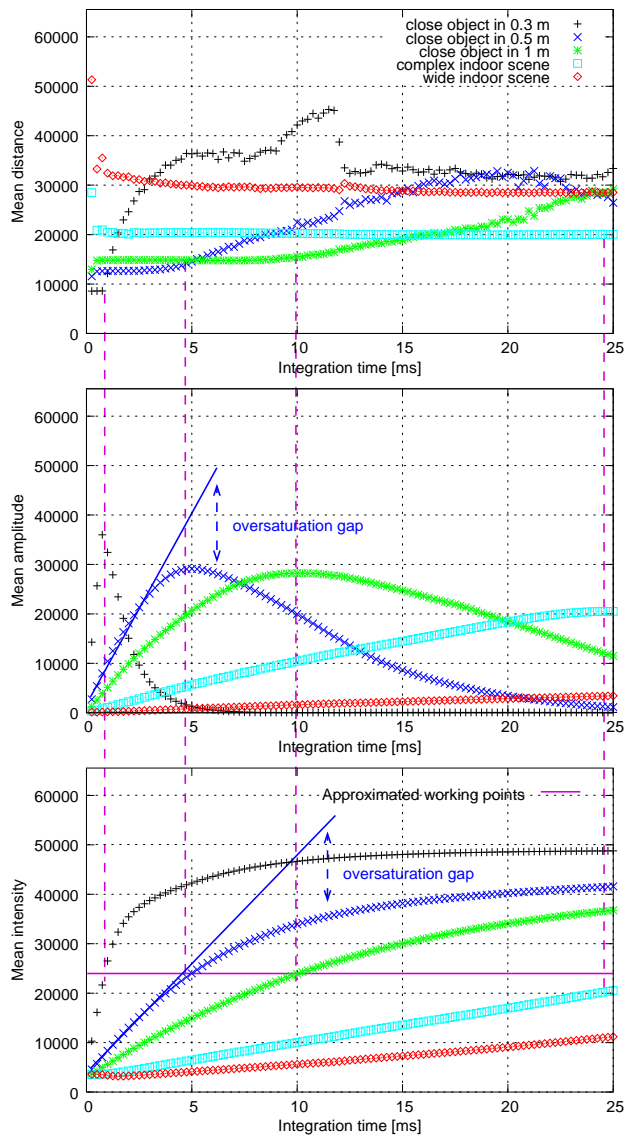


Fig. 3. Determining optimal integration time. Note: Data values represented here are register values with a range of 0 to 65532.

image acquisition [22]. The algorithm considers amplitude and intensity values of each pixel. The sum of both values are limited by a maximum value, which is technically determined. A heuristically found threshold signals oversaturation of single pixel values. Instead of adjusting the integration time to prohibit oversaturation of single values, we propose an algorithm to consider the mean accuracy. We found out, that mean accuracy correlates with mean intensity and mean amplitude in the linear segment. The most precise mean accuracy could be acquired with an integration time located near the amplitudes maxima.

The maxima determination is a complex task, because mean amplitude functions do not provide stable approximations of working points. A tricky way is to translate the maxima into the intensity diagram. The intersection points lie approximately on a straight line (see figure 3). Mean intensity values

can be calculated on the fly during capturing. We relocated the approximating value I_A a small interval below to fit equation 2 best. Note, that the image has a non-neglective saturation at the amplitudes maximum. Values for I_A between 15000 and 18000 provided reliable results. The integration time determination algorithm is based on the following scheme:

- 1) Calculate the mean intensity \bar{I}_t from the intensity dataset I_t at time t
- 2) Determine control deviation $D_t = \bar{I}_t - I_A$
- 3) Update control variable $c_{t+1} = -V_p * D_t + c_0$ for grabbing the next frame, where c_{t+1} is the integration time, V_p the proportional closed loop deviation parameter and c_0 a suitable initial value.

This approach assumes, that two images succeeding one another do mostly not differ significantly. We have chosen to use a proportional controller, because handling environment dynamics requires quick adaptations.

B. Accuracy determination

After integration time determination the distance accuracy has to be rated. Supposing that the photonic interference is the main reason for inaccuracy, its influence has to be minimized to get precise data. A physical law relating inaccuracy with photonic interference is defined formally as [23]:

$$\Delta L = \frac{L}{\sqrt{8}} * \frac{\sqrt{I}}{2 * A}, \quad (2)$$

where ΔL represents the inaccuracy, L the maximum distance, A the amplitude value and I the intensity value. The maximum distance is defined as

$$L = \frac{c}{2 * f_{mod}}, \quad (3)$$

where c is the speed of light and f_{mod} the modulation frequency. The Swiss Ranger device uses a frequency of 20 MHz. That's why it has an effective range of 7,5 meters.

Assuming linearity, equation 2 must have its minimum near the amplitude maximum.

Combining integration time determination and inaccuracy thresholding has to be proved for connectivity. We used the inaccuracy threshold as pre-filter, which rejects points with inaccuracies higher than a fixed level. Remaining points were considered for the mean inaccuracy. Figure 4 shows a scene taken in our robotic pavilion. This scene was captured in different distances to determine oversaturating effects. Two significant measurement examples are illustrated in figure 5. One of the measured scene runs into oversaturation due to near objects. The other measured scene does not include near objects, why it provides good linearity.

The measurement series included the assimilation of the mean inaccuracy. Figure 6 shows the mean inaccuracy in dependency of threshold values. Each minima is nearly located at the same position. Consequently, it will have no influence on the integration time determination.

There is also another fact indicating, that the right integration time is found. It is the number of data values fitting the



Fig. 4. Complex scene used for a measurement series to determine integration time. Note, that the poster has a glossy surface and quickly leads to oversaturation for short distances.

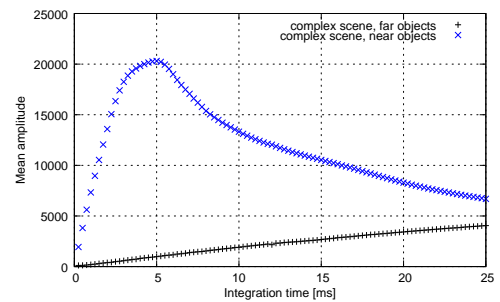


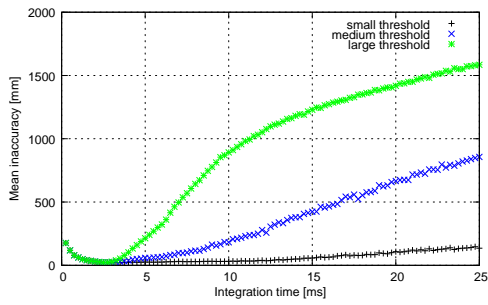
Fig. 5. Mean amplitude graph of two different scenes. The amplitude values of the upper curve run into oversaturation.

given threshold. Figure 7 illustrates, that not only the accuracy of valid points is growing (falling inaccuracy) in dependency of integration time, but also the number of points fitting the threshold constraint.

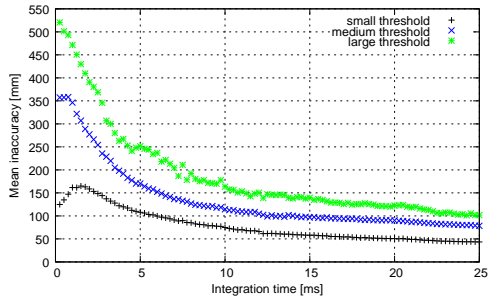
Although, the optimal integration time is found, there is still erroneous data remaining. Pixel information with less distance accuracy will be useless. The threshold experiment demonstrates a simple filter to identify some of these pixels. Some further experiments were done based on this filter with the purpose of proving the sensors reliability in serious environmental situations.

C. Materials with high reflectivity and light emitting objects

Gut pointed out, that multipath reflexion and scattering light are types of environmental sensor influences. These effects could be confirmed during our tests. We took a small mirror and placed it somewhere in an indoor scene. Measurements towards the mirror were highly irritated by reflected light. It was not only noise, which superpose the measurement, but also erroneous data. Distance informations taken at several integration time settings were sometimes completely incorrect. Similar effects could be observed with bright moving objects close to the camera. The brightness of those objects also influences the surrounding pixels of their representation on chip. In animated distance data illustrations, the background seems to move. Figure 8(a) and 8(b) demonstrate this effect. Both figures show the same scene, except a very close hand



(a) Mean inaccuracy for oversaturated scenes



(b) Mean inaccuracy for non-saturated scenes

Fig. 6. Mean inaccuracy in dependency of threshold values

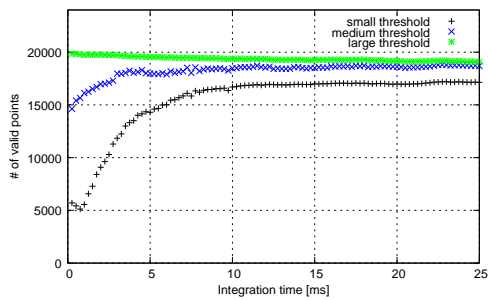
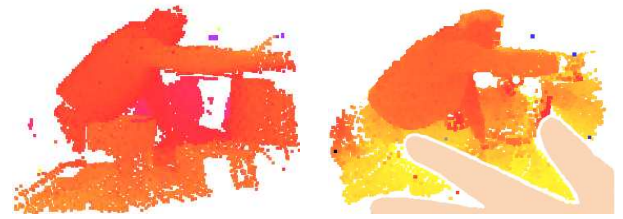


Fig. 7. Number of data values fitting a given threshold for the non-saturating scene

appearing at the bottom of the right picture. The hand was reconstructed afterwards in pink color to mark its position.

The threshold filter developed in the previous section could determine the erroneous data values reliably. Results of applying the threshold filter to similar scenes are shown in figure 9. All background informations, which are highly irritated could be filtered out. It is amazing, that the gaps between two fingers could also be determined. These gaps are located very close to their influencing areas, where scattering light dominates the scene.

The result of irritating the camera by its own emitted light induced us to carry out another experiment. We were using an ordinary light source instead of the mirror and came to the same results. The sensor was not only influenced by his own emitted light, but also by lamps or by sunlight, because ambient light has partly an infrared spectral component. The influence had been grown while lowering the reflection rate of the scene, e.g. if objects were moving away. In figure 10 you can see two drop lights, which were switched on and off during



(a)

(b)

Fig. 8. The Swiss Ranger is highly influenced by scattering light of near objects. These figures illustrate the problem that comes without applying filters. Areas with low intensity are being more influenced by scattering light. 8(a) Scene including no near objects. 8(b) Scene including a very close hand at the bottom.



(a)

(b)

Fig. 9. Applicability of threshold filter to light scattering scenes. 9(a) Usage of threshold filter to suppress erroneous data from scattering light of a very close hand. 9(b) Scattering light highly influences low amplitude areas.

capturing. The droplight area in figure 10(b) is very noisy, if the light is switched on. Applying the threshold filter provided stable results. The noisy area was filtered out completely. You can see the results in figure 11.

IV. DISCUSSION AND CONCLUSION

As conclusion, we state, that the Swiss Ranger time-of-flight camera SR-2 is a very interesting sensor for mobile robotics presumed that reliable calibration, lighting adaptation and accuracy filtering is applied. This paper presents a novell approach to solve these three main problems to use the camera in robotic applications.

- 1) First, we presented a simple approach to calibrate the camera, which is an essential precondition to grab precise 3D data.
- 2) Second, we implemented an online algorithm to determine approximately the optimal integration time for the next sequence. Due to the measurement principle, a 3D camera is susceptible to oversaturation, if close bright objects or light sources are present in the scene. It is important to determine the correct adaptation parameters for each scene.
- 3) Third, we developed accuracy filter based on the physical principle of photonic interference to determine in-

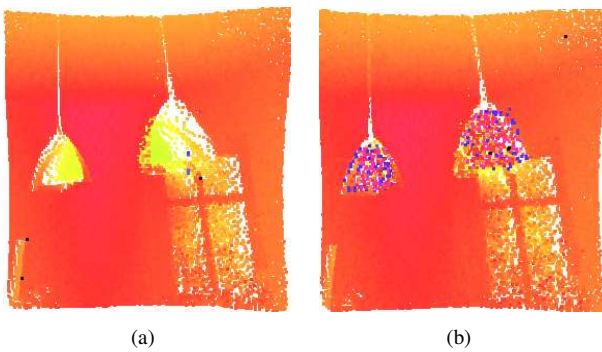


Fig. 10. Determination of influences of light emitting sources. 10(a) Image taken from Swiss Ranger device of two drop lights, which are switched off. 10(b) Image taken from Swiss Ranger device of two drop lights, which are switched on.

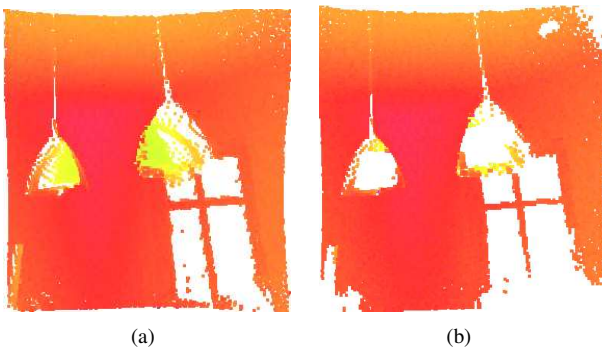


Fig. 11. Determination of erroneous data. 11(a) Filtered image taken from Swiss Ranger device of two drop lights, which are switched off. 11(b) Filtered image taken from Swiss Ranger device of two drop lights, which are switched on.

accurate distance values.

These three approaches enable the usage of this camera for robotic navigation and mapping tasks. A video showing the performance can be found at <http://www.ais.fraunhofer.de/~surmann/videos/iros2006.mpg>. The advantages of 3D cameras over previous visual sensing techniques, like laser range finders, are its small size, its small weight and mainly its high performance with up to 30 frames per second. Especially for navigation tasks, the absence of panning tilts is also worth mentioning.

Needless to say, a lot of work remains to be done. Future work will concentrate on 3D object segmentation and recognition as well as our SLAM algorithms implemented for 3D laser range finders. Furthermore, we are looking forward to the next generation of 3D cameras.

REFERENCES

[1] C. C. S. d'Electronique et de Microtechnique SA, "Swiss ranger sr-2, <http://www.swissranger.ch>," 2005.
 [2] P. GmbH, "Pmd-cameras, <http://www.pmdtec.com/einhalt/produkte/kamera.htm>."
 [3] C. C. S. d'Electronique et de Microtechnique SA, "Csem microtechnology, <http://www.csem.ch>," 2005.
 [4] S. Se, D. Lowe, and J. Little, "Local and Global Localization for Mobile Robots using Visual Landmarks," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '01)*, Hawaii, USA, October 2001.

[5] C. Jennings and D. Murray, "Stereo vision based mapping and navigation for mobile robots," 1997. [Online]. Available: citeseer.ist.psu.edu/murray97stereo.html
 [6] P. Biber, H. Andreasson, T. Duckett, and A. Schilling, "3D Modeling of Indoor Environments by a Mobile Robot with a Laser Scanner and Panoramic Camera," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '04)*, Sendai, Japan, September 2004.
 [7] P. Allen, I. Stamos, A. Gueorguiev, E. Gold, and P. Blaer, "AVENUE: Automated Site Modelling in Urban Environments," in *Proceedings of the Third International Conference on 3D Digital Imaging and Modeling (3DIM '01)*, Quebec City, Canada, May 2001.
 [8] A. Georgiev and P. K. Allen, "Localization Methods for a Mobile Robot in Urban Environments," *IEEE Transaction on Robotics and Automation (TRO)*, vol. 20, no. 5, pp. 851 – 864, October 2004.
 [9] V. Sequeira, K. Ng, E. Wolfart, J. Goncalves, and D. Hogg, "Automated 3D reconstruction of interiors with multiple scan-views," in *Proceedings of SPIE, Electronic Imaging '99, The Society for Imaging Science and Technology /SPIE's 11th Annual Symposium*, San Jose, CA, USA, January 1999.
 [10] P. Besl and N. McKay, "A method for Registration of 3-D Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239 – 256, February 1992.
 [11] M. Hebert, M. Deans, D. Huber, B. Nabbe, and N. Vandapel, "Progress in 3-D Mapping and Localization," in *Proceedings of the 9th International Symposium on Intelligent Robotic Systems, (SIRS '01)*, Toulouse, France, July 2001.
 [12] O. Wulf and B. Wagner, "Fast 3D-Scanning Methods for Laser Measurement Systems," in *Proceedings of the International Conference on Control Systems and Computer Science (CSCS14)*, Bucharest, Romania, July 2-5 2003.
 [13] H. Surmann, A. Nüchter, and J. Hertzberg, "An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments," *Journal Robotics and Autonomous Systems*, vol. 45, no. 3 – 4, pp. 181 – 198, December 2003.
 [14] P. Kohlhepp, M. Walther, and P. Steinhaus, "Schritthalte 3D-Kartierung und Lokalisierung für mobile Inspektionsroboter," in *Proceedings of the Autonome Mobile Systeme 2003, 18. Fachgespräche*, R. Dillmann, H. Wörn, and T. Gockel, Eds., December 2003.
 [15] O. Wulf, K. O. Arras, H. I. Christensen, and B. A. Wagner, "2D Mapping of Cluttered Indoor Environments by Means of 3D Perception," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '04)*, New Orleans, USA, April 2004, pp. 4204 – 4209.
 [16] H. Automatic, "Hokuyo, small size 2d laser scanner, <http://www.hokuyo-aut.jp/>," 2006.
 [17] S. Thrun, D. Fox, and W. Burgard, "A real-time algorithm for mobile robot mapping with application to multi robot and 3D mapping," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '00)*, San Francisco, CA, USA, April 2000.
 [18] C. Früh and A. Zakhor, "3D Model Generation for Cities Using Aerial Photographs and Ground Level Laser Scans," in *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR '01)*, Kauai, Hawaii, USA, December 2001.
 [19] H. Zhao and R. Shibasaki, "Reconstructing Textured CAD Model of Urban Environment Using Vehicle-Borne Laser Range Scanners and Line Cameras," in *Second International Workshop on Computer Vision System (ICVS '01)*, Vancouver, Canada, July 2001, pp. 284 – 295.
 [20] J. W. Weingarten, G. Gruener, and R. Siegwart, "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, September 2004.
 [21] O. Gut, "Untersuchungen des 3d-sensors swissranger," Master's thesis, Swiss Federal Institute of Technology Zurich, 2004.
 [22] Thierry Oggier, Bernhard Büttgen, Felix Lustenberger, Guido Becker, Björn Rüegg and Agathe Hodac, "Swissranger sr3000 and first experiences based on miniaturized 3d-tof cameras, http://www.swissranger.ch/pdf/application_sr3000_v1.1.pdf," CSEM, IEE, Fachhochschule Rapperswil Switzerland, Tech. Rep., 2005.
 [23] Peter Seitz, Thierry Oggier and Nicolas Blanch, "Optische 3d-halbleiter-kameras nach dem flugzeitprinzip, <http://www.extenza-eps.com/extenza/loadhtml?objectIdvalue=49132&type=abstract>," CSEM, Tech. Rep., 2004.