Experimental Evaluation of State of the Art 3D-Sensors for Mobile Robot Navigation[1)]

Peter Einramhof, Sven Olufs and Markus Vincze
Technical University of Vienna, Automation and Control Institute
{einramhof, olufs, vincze}@acin.tuwien.ac.at

Abstract:
In this paper we discuss the suitability of three optical 3D range finders for mobile robot navigation in indoor environments. The range finders under consideration are the CSEM SwissRanger time-of-flight camera[1)] *and both, a horizontal and vertical baseline stereo-vision system. Modelling the world in 2D and using 2D sensors that scan parallel to the ground for the task of navigation is still popular in the robotics community. This approach works in adapted, but not in general indoor environments. We evaluate the usability of named 3D sensors in such general environments by practical experiments.*

1 INTRODUCTION

Autonomous navigation in indoor environments i.e. office and domestic environments is a challenging task in robotics. For collision-free motion of a mobile robot, the sensor system – in this case, the vision system – must reliably detect obstacles in the scene, while precision is secondary. For self-localization and mapping of the robot's environment, the precision and completeness of the acquired (3D-)data is an important point. In both cases, real-time capability is required. To keep the amount and complexity of the data and thus computation time low, making use of the 2D-world assumption for navigation is still very popular. That is, the shape of every object when viewed from above, is modelled as a set of lines in a 2D map. Representing the world this way is "compatible" to how 2D range finders (e.g. SICK laser scanner) perceive their surrounding. The generally regular interior design of office environments with straight walls, closet fronts and rectangular tables with drawer boxes fulfill the 2D-world assumption quite well. However, when introducing mobile robots i.e. service robots into domestic environments, the 2D-world assumption no longer holds.
The advantages of 2D laser range finders are the high accuracy and reliability of their measurements. The disadvantage is that they only scan at one plane. Thus, obstacles that do not intersect with the scan plane cannot be detected. Assuming that the scan plane is parallel to

[1)] This work was supported by the European Union project MOVEMENT #IST-2003-511670.
[1)]`http://www.swissranger.ch/`

the floor, typical obstacles of this type found in indoor environments are chairs and tables. Even if their legs are detected by the laser, the seat and tabletop are generally not. As in these cases the three-dimensional character of the real world is no longer negligible, the use of 3D sensors is a possible solution.
Today the most common techniques for 3D sensing are based on usual CMOS/CCD, laser scanner [7, 9] or time of flight sensors [4, 11]. A typical CMOS/CCD 3D system is based on the well known stereo vision [3] or structure from motion approach [7]. Both have difficulties providing reliable sensor data in environments with changing illumination conditions like all passive sensors. Due to the nature of our application we can not assume a moving camera or robot. Valid sensor data must be available anytime i.e. for security issues. Active CMOS/CCD based methods e.g. structured light [1] gives a better robustness, but the suitability for autonomous mobile robot applications is quite limited. Active structured light sensors were successfully used in the MAKRO [6] project for autonomous vision-based canal inspection. A typical 3D laser scanner approach in robotics e.g. Kurt3D [9] is based on 2D laser range finders, actually designed for instutrial applications, which are expanded by an additional rotation axis for acquiring 3D information. To receive consistent data 3D with this kind of approach it is assumed that the robot is not moving. Other approaches are based on time integration with multiple laser scanners [8, 10]. Analog to the structure from motion approaches a moving camrea is assumed. Industrial 3D laser scanner are available e.g. the CYRAX scanner, but the usability of those systems is limited due to their size, weight and power consumption.
The new time-of-flight cameras are based on the photon mixer device technology e.g. the SwissRanger SR-2 CSEM. This kind of sensor is capable of producing dense 3D-data in real-time. The camera is based on a 2-dimensional image sensor and a modulated light source (sinusoidal, 20MHz). For each pixel of the image sensor, the camera delivers the measured distance as well as the intensity. Due to the nature of the sensor, the pixel resolution of TOF (Time of Flight) cameras relative low e.g. 124x160 (SR-2). Prasad et al. [5] increased the resolution of the TOF sensor by sensor fusion with an additional CCD camera.

The paper is organised as follows. Section 2 introduces the principle of the SwissRanger time-of-flight camera followed by a discussion of the challenges for vision sensors when used in robot navigation in Section 3. Experimental results are shown in Section 4. Finally, we discuss the results and the approach in Section 5.

2 THE SWISSRANGER TIME-OF-FLIGHT CAMERA

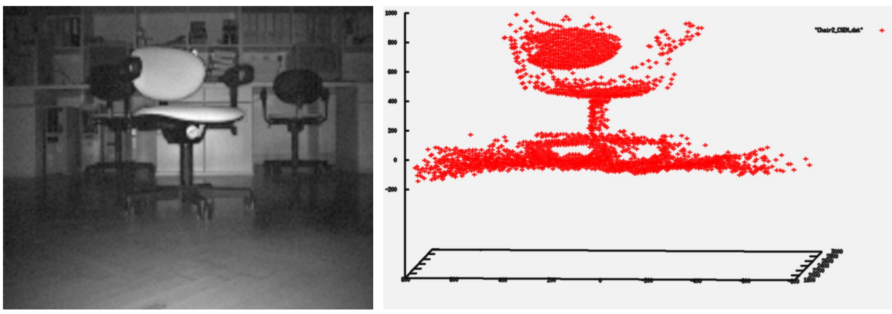
The following section briefly describes the measurement principle of the SwissRanger in order to provide a base for the subsequent discussion of the vision challenges arising in its use in mobile robot navigation. An in-depth discussions of the SwissRanger's implementation can

(a) Swissranger 2 (b) SwissRanger 3000

Figure 1: CSEM's time-of-flight cameras SwissRanger SR-2 and SR-3000

be found in [11] and [2].
The SwissRanger (SR-2 and SR-3000, Fig. 1) is a 3D time-of-flight camera, manufactured by the Swiss company CSEM that is capable of producing dense 3D-data in real-time. The camera is based on a 2-dimensional image sensor (SR-2: 160 x 124 pixel, SR-3000: 176 x 144 pixel) and a modulated light source (sinusoidal, 20MHz). Every pixel of the sensor samples the amount of modulated light that is reflected by objects in the scene. This is done four times per measurement period at equidistant intervals. The phase shift between the emitted and the returning light and subsequently the distance are derived. The intensity of the objects in the image is recovered from the average light reflected. By default, the SwissRanger uses a modulation frequency of 20MHz, resulting in a *non-ambiguity range* of 7.5m.
As the camera actively illuminates the scene, it also works under poor lighting conditions and even in complete darkness. The infrared part of (direct) sunlight can lead to over saturation of the pixels. Other than a stereo camera system, the SwissRanger does not require textured surfaces for generating dense 3D-data.

With each measurement, the raw data provided by the SwissRanger consists of two arrays



(a) Intensities (b) 3D-data

Figure 2: Examples for an intensity and 3D-image calculated from the SwissRanger's raw data

of 16bit-values. The first array contains the measured distances for each pixel, the second the associated intensities. Square-root scaling of the intensity values yields a common 8bit gray-level image (Fig. 2(a)). The camera's driver comes with a set of configuration files containing model-specific calibration information. Based on this information, the driver can calculate 3D-data (Fig. 2(b)). As an alternative, the user can calibrate the camera as described in [11] and use the calibration data for calculating 3D-data independent of the driver.

3 CHALLENGES FOR VISION SENSORS IN NAVIGATION

3.1 Field of View (FOV)

The SR-3000 (SR-2) has a horizontal FOV of 47.5 deg (43 deg) and a vertical FOV of 39.6 deg (46 deg). The FOV of a stereo camera is tradeoff between precision and view range. A larger view range will usually result in a poor precision or the chance that small objects are not detected. Experiments have shown that a FOV of more than 60 deg is suitable to only a limited extent. Compared to a laser range finder, the cameras' narrow FOV generally results in the detection of far less features (e.g. line segments) usable for map-based self-localization. With respect to obstacle avoidance, the narrow horizontal FOV is problematic, especially when obstacles are close to the mobile robot but not directly in front of the camera. Other than the laser ranger, the vertical FOV of the cameras allows to detect low obstacles on the ground (e.g. a doorstep) and protruding obstacles above the ground (e.g. a tabletop). However, this advantage is relativised when considering that the cameras (and therefore the robot) must have sufficient distance to the obstacles for them to be within the vertical FOV.

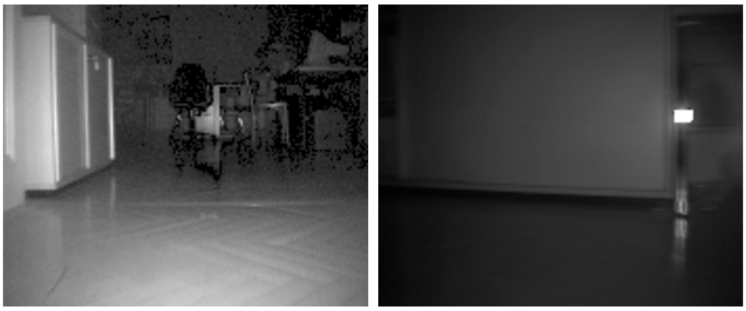


Figure 3: The left image shows the noise introduced to a measurement (intensity data, normalized) taken with the SwissRanger as result of a too short integration time. The right image shows over saturation and motion blur due to a too long integration time.

3.2 Integration Time

Due to the measurement principle of the Swissranger, setting the correct integration time is a key point for getting reliable measurements. Too short integration times lead to strong measurement noise, and too high integration times can cause over saturation of the SwissRanger's pixels, thus falsifying the measurement, especially with near objects. Furthermore, long integration times lead to motion blur when the mobile robot is moving (Fig. 3).
May et al. [4] proposed an algorithm for the online adaptation of the integration time, based on the intensity values of the current measurement. However, in [2] O. Gut showed that each

integration time requires a separate calibration of the SwissRanger. Furthermore, he found that subsequent measurements of a static scene deliver too long distances that tend towards final (lower) distance values after about 20-40 measurements. The longer the integration time the more noticeable is this effect.
Due to the nature of the stereo camera as a passive sensor, we face the problem that the quality of the 3D-data depends on the quality of the corresponding raw images. This means that "good" illumination is required. Office environments usually fulfill this requirement much better than domestic environments.

3.3 Aliasing

With the SwissRanger measuring the phase shift between emitted and detected light, we face the problem of *aliasing* in the image. In case objects further away than the non-ambiguity range (7.5m) reflect enough light, this results in an *aliasing effect*. This means that distances greater than this range are mapped back to distances, again beginning at 0m. Such erroneous measurements can be eliminated: "Real" short distances are associated with high measured intensity, while the intensity of alias pixels is very low. Using a thresholding operation of the measurements' intensities depending on the associated distance value is sufficient.

4 EXPERIMENTAL RESULTS

4.1 Test setup

For our experiments we used a mobile robot with a SICK LMS 200 laser range finder mounted on its front. The sensor has a horizontal FOV of 180 deg and an angular resolution of 0.5 deg. The scanning plane is 11cm above the ground plane. On top of the SICK, we mounted a SwissRanger 3000, with its optical axis 33cm above and parallel to the ground plane. We used the integration time preset by the camera's driver. Furthermore, a stereo-vision system[2)] was mounted on the robot's front. All named sensors are oriented into the robot's driving direction. The stereo-vision system is used in two setups: the typical stereo camera with a left-right camera "Stereo (H)" and a 90 degree rotated top-down camera "Stereo (V)". The stereo system was calibrated using the well-known method proposed by Zhang [12].

[2)] custom built system with Videre-Design SRI Stereo Engine, 10cm baseline, 2x ImagingSource DFK 21F04, VGA resolution, Pentax 6mm

4.2 Results

For evaluating the suitability for obstacle avoidance, the detectability of obstacles that are problematic for 2D laser range finders was tested: various types chairs and tables. We used simple occupancy grids with a resolution of 10cm. With our test setup, up to minumum distances depending on the height of the obstacle and the vertical FOV of the cameras, chairs and tables were reliably detected. However, especially with tables we encounter the problem that the vision systems look under the tableplate and don't detect the obstacle anymore.
The precision of the sensors with respect to detecting vertical surfaces (e.g. wall or closet front) for map-based self-localization was evaluated with three typical "robot senses wall" scenes.That is, the robot is 1) aligned orthogonally to a wall, 2) aligned skewed to a wall and 3) facing a corner. The distance of the robot to the wall is approximately 3m. All the scenes chosen contained weak structure and texture in order to obtain more representative measurements. Please note that all used sensors have a different field of view.
The method used for detecting the wall in the data of the SwissRanger and stereo-vision system is straightforward, similar to Weingarten et al. [11]: For both sensors, we used the sensor readings whose rays are parallel to the floor plane. In order to remove noise we used the average of all readings within a vertical 4degree window to the ground-parallel axis. With stereo vision, we face the problem that we can have a lack of such sensor readings, i.e. a sparse depth image. In this case we use two kernel functions to seek up- and downward the parallel axis. The function stops at the first found distance. If both functions yield a value, the nearest distance is used.
Table 1 shows the results of all three sensors. The robot's laser ranger was used as ground truth. The rotated top-down stereo camera "Stereo (V)" gives here the lowest error while the standard deviation is high. This results from various false positives that are observed "inside" the wall i.e. the observed distance is further away than the wall itself (Fig. 4). Similar results are also observed with "Stereo (H)": most of the observed distances are also "inside" the wall. The vertical stereo setup performed better than the horizontal setup as in the test scenes horizontal features (e.g. transition between the wall and the floor) were dominant. Only few vertical features were present.

The SwissRanger shows adequate results with the best standard deviation ratio values.

	average	Std. deviation	min	Max
Csem	10.841	3.333	1.079	29.580
Stereo (H)	31.100	24.753	1.763	167.778
Stereo (V)	6.752	50.654	1.843	55.080

Table 1: Measured error of the sensors, Stereo (H) is horizontal baseline, Stereo (V) vertical baseline. All units in cm.

Depending on the reflection angle and surface material the distances are observed "inside"

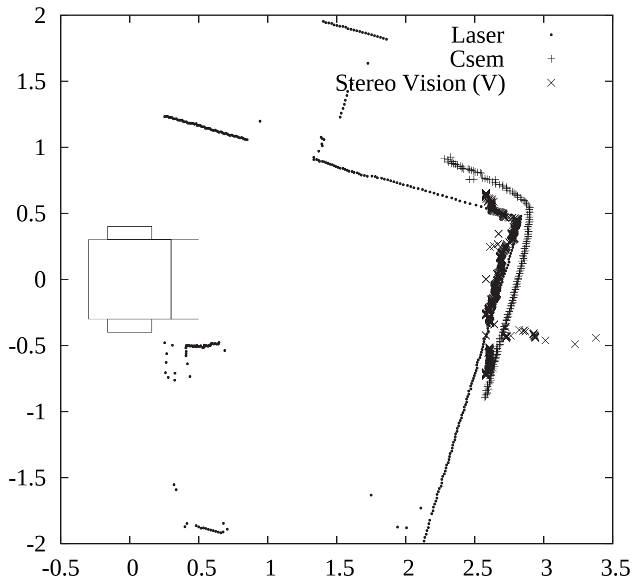


Figure 4: Sensor Readings from Scene 3, the robot is placed at (0,0). Units in meter.

the wall (Fig. 4). In total, all measurements of the CSEM seem to be "scaled". Note that measuring the exactly same scene using different integration times has a noticeable impact on the precision (5-10cm).

5 DISCUSSION AND FUTURE WORK

In this paper we discussed the suitability of optical 3D range finders for mobile robot navigation applications. The CSEM and vertical stereo vision showed both an adequate results within the test scenes. We showed the practical issues in robot navigation tasks within the sensors i.e. FOV, aliasing and integration time. Compared to the 2D laser range finder, the clear advantage of both sensors is that they provide three-dimensional data of the environment. This is especially an advantage when navigating in domestic and office environments featuring obstacles such as chairs and tables that are hard to detect with 2D-sensors only. One clear advantage of these sensors is their small size and small weight.

Our next steps will be to aim for a fusion of the data acquired by a SwissRanger and a stereo camera system that are used in parallel, in order to get a high-resolution 3D sensor. Furthermore, another interesting topic is the extraction of features in the intensity images of the SwissRanger by using 2D-vision methods, and subsequently examining the associated 3D-data of "interesting" regions only. By this approach, the detection of small object (i.e. obstacles) in the noisy 3D-data might become more robust.

References

[1] D. Foli, J. Salvi, and El M. Mouaddib. Uncalibrated vision based on structured light. In

International Conference on Robotics & Automation, Seoul, Korea, 2001.

[2] O. Gut. "untersuchungen des 3d-sensors swissranger", master's thesis. Technical report, Swiss Federal Institute of Technology Zurich, 2004.

[3] C. Jennings and D. Murray. Stereo vision based mapping and navigation for mobile robots. In *IEEE International Conference on Robotics and Automation*, pages 1694–1699, 1997.

[4] Stefan May, Bjn Werner, Hartmut Surmann, and Kai Pervz. 3d time-of-flight cameras for mobile robotics. In *International Conference on Intelligent Robots and Systems IROS 2006*, Beijing, China, 2006.

[5] T.D.Arun Prasad, Klaus Hartmann, Wolfgang Weihs, Seyed Eghbal Ghobadi, and Arnd Sluiter. First steps in enhancing 3d vision technique using 2d/3d sensors. In *Computer Vision Winter Workshop 2006*, Telc, Czech Republic, 2006.

[6] E. Rome, J. Hertzberg, F. Kirchner, U. Licht, H. Streich, and Th. Christaller. Towards autonomous sewer robots: the makro project. *Urban Water*, (1):57–70, 1999.

[7] Stephen Se, David Lowe, and Jim 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)*, pages 414–420, Maui, Hawaii, October 2001.

[8] D. Stavens and S. Thrun. A self-supervised terrain roughness estimator for off-road autonomous driving. In *Conference on Uncertainty in AI (UAI)*, 2006.

[9] Hartmut Surmann, Kai Lingemann, Andreas Nchter, and Joachim Hertzberg. A 3d laser range finder for autonomous mobile robots. In *32nd International Symposium on Robotics*, pages 153–158, Seoul, Korea, 2001.

[10] S. Thrun, D. Fox, and W. Burgard. A real-time algorithm for mobile robot mapping with application to multi robot and 3d mapping. In *International Conference on Robotics & Automation*, San Francisco, CA, USA, 2000.

[11] J. W. Weingarten, G. Gruener, and R. Siegwart. A state-of-the-art 3d sensor for robot navigation. In *International Conference on Intelligent Robots and Systems IROS 2004*, Sendai, Japan, 2004.

[12] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.