We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK
CITATION
INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

3D Cameras:
3D Computer Vision of wide Scope

Stefan May, Kai Pervoelz and Hartmut Surmann
Fraunhofer Institute for Intelligent Analysis and Information Systems
Germany

1. Introduction

The human visual sense is the one among all other senses that gathers most information we receive. Evolution has optimized our visual system to negotiate one's way in three dimensions even through cluttered environments. For perceiving 3D information, the human brain uses three important principles: stereo vision, motion parallax and a-priori knowledge about the perspective appearance of objects in dependency of their distance. These tasks pose a challenge to computer vision since decades. Today the most common techniques for 3D sensing are CCD- or CMOS-camera, laser scanner or 3D time-of-flight camera based. Even though evolution has shown predominance for passive stereo vision systems, three additional problems are remaining for 3D perception compared with the two mentioned active vision systems above. First, the computation needs a great deal of performance, since the correlation of two images from a different point of view has to be found. Second, distances to structureless surfaces cannot be measured, if the perspective projection of the object is larger than the camera's field of view. This problem is often called aperture problem. Finally, a passive visual sensor has to cope with shadowing effects and changes in illumination over time.

That is why for mapping purposes mostly active vision systems like laser scanners are used , e.g. [Thrun et al., 2000], [Wulf & Wagner, 2003], [Surmann et al., 2003]. But these approaches are usually not applicable to tasks considering environment dynamics.

Due to this restriction, 3D cameras [CSEM SA, 2007], [PMDTec, 2007] have attracted attention since their invention nearly a decade ago. Distance measurements are also based on a time-of-flight principle but with an important difference. Instead of sampling laser beams serially to acquire distance data point-wise, the entire scene is measured in parallel with a modulated surface. This principle allows for higher frame rates and thus enables the consideration of environment dynamics.

The first part of this chapter discusses the physical principles of 3D sensors, which are commonly used in the robotics community for typical problems like mapping and navigation. The second part concentrates on 3D cameras, their assets, drawbacks and perspectives. Based on these examining parts, some solutions are discussed that handle common problems occurring in dynamic environments with changing lighting conditions. Finally, it is shown in the last part of this chapter how 3D cameras can be applied to mapping, object localization and feature tracking tasks.

2. Range Sensing

Before focusing on 3D cameras and their applications, a short comparison of range sensors and their underlying principles is given. Since there are many different types of sensors for range sensing, the section focuses on those that are most common in the domain of robotics, i.e. stereo vision systems, 3D laser scanners and of course 3D cameras. The section first introduces into underlying measurement principles before it describes real sensor systems in more detail.

2.1 Range Measurement Principles

Different types of sensors are based on different measurement principles. The two main principles for technical systems are triangulation and time-of-flight. Both principles can further be separated into two subcategories: active and passive triangulation or respectively pulsed and phase shifted time-of-flight.

2.1.1 Triangulation

This technique is called triangulation since the object whose distance should be measured forms a triangle with two parts of the sensor (cf. Figure 1). If the sensor consists of one receiver part and one active transmitter part, the measurement principle is called **active triangulation**. If it consists only of two passive receivers, it is called **passive triangulation**.
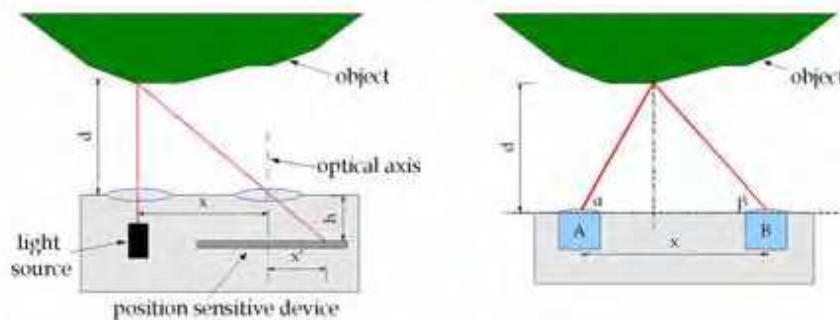

Figure 1. Left image: Working principle of active triangulation. Right image: Working principle of passive triangulation

Active triangulation. The configuration of a simple active triangulation sensor can be seen in figure 1. A light source projects a single point onto the object and the reflection of the light point is measured by the receiver part of the sensor. This receiver is a position sensitive device, which can determine the point where the light reflection has hit the receiver. By knowing the position of the sensor's optics, the distance between transmitter and receiver x, their distance to the optics h and the hitting point of the light reflection x', it is possible to calculate the distance d of the object by the formula:

$$d = h \cdot \frac{x}{x'} \tag{1}$$

Such a simple sensor configuration restricts the distance measurement capability to one single point. To determine the shape of an object, either the sensor or the object itself must be moved and several measurements have to be taken. Higher sophisticated triangulation sensor systems use two-dimensional light sources as well as two-dimensional receivers. They project a light pattern onto the object, which is received by, e.g. a 2D camera system. Such a system directly provides 3D shape information of the measured objects.

Passive triangulation. This principle is well known in nature and has been improved over millions of years since it is the basic principle of the human visual sense. Being more precise, it is the base of the human depth perception. It consists as well as a technical passive triangulation sensor of two receivers (the eyes or two cameras respectively) which are observing an overlapping area (cf. figure 1). If a specific point p is in the field of view of both receivers, it is possible to determine its distance d to the sensor. Therefore each receiver calculates the angle between the line from the sensor to point p and the optical axis of the receiver. In combination with the distance x between the two receivers the distance to the point p is calculated by

$$d = \frac{x}{\dfrac{1}{\tan\alpha} + \dfrac{1}{\tan\beta}}, \qquad (2)$$

where α is the angle from receiver A to the point p and β is the angle from receiver B to the point p. This formula assumes that the optical axes of receiver A and B are parallel. The most important task here is to find distinctive points in the images and assign correspondences between the points in the two images. Each point in the image of receiver A has to be correctly identified in the image of receiver B. Wrong assignments will result in wrong distance measurements.

2.1.2 Time-of-Flight

As the name already implies, this principle utilizes the time a specific signal needs to travel from the sensor to the object and back. For calculating this time, different methods can be used. In the following, two of them will be described in more detail, namely the **impulse time-of-flight** method and the **phase difference** method.

Impulse Time-of-Flight. This method is the most obvious one, since a timer is started when a signal is sent to the object and stopped when its reflection is received. By knowing the speed of the signal, the distance to the object can be calculated directly. In practice most often a short laser impulse is sent out to the object and the time until it is detected by an optical receiver is measured.

From that travel time t, the distance d can be calculated by the formula

$$d = t \cdot c, \qquad (3)$$

where c is the speed of light.

Phase difference. A more complex method is the calculation of the travel time by measuring the phase difference between the sent signal and its reflection from the object. In principle the modulation of light waves itself could be used but since their wavelengths are in the range of some nm it would be difficult to determine the phase difference. Therefore the light signal is modulated again with a much longer wavelength.
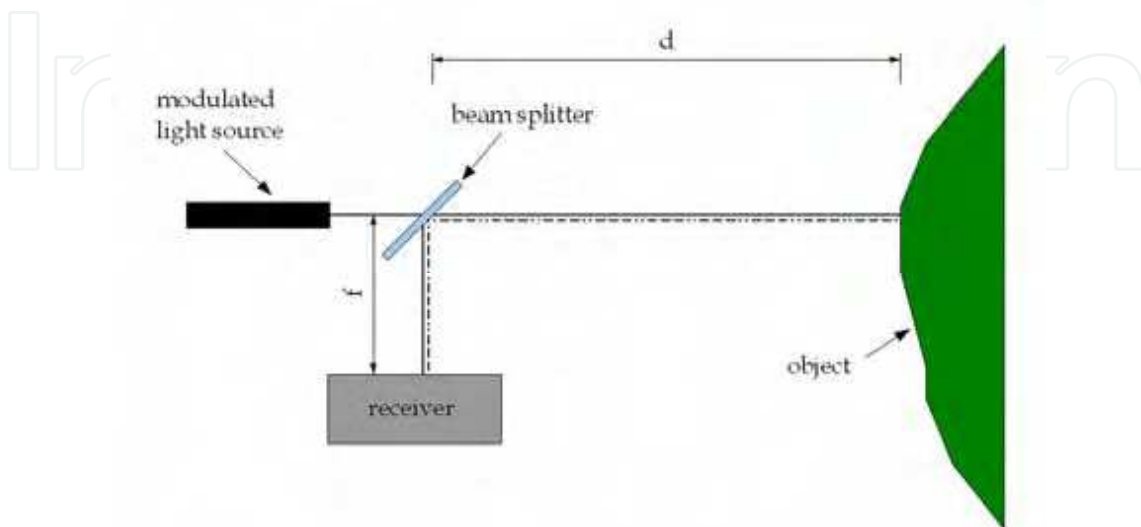
Figure 2. Drawing of the phase-difference time-of-flight measurement principle. A modulated light signal is split into a reference- and a measurement signal. The measured phase-difference gives the time-of-flight of the signal and thus the distance

As shown in figure 2 the modulated light signal is split into two signals by a semi permeable mirror, also called beam splitter. One of the signals, the reference signal, is sent directly to the internal receiver which has a distance f to the beam splitter. The other one, the measurement signal, is sent to the object which is located at a distance d. When the signal is reflected by the object and detected by the internal receiver of the range sensor, it has in total covered a distance d', which is defined by

$$d' = f + (2 \cdot d) . \tag{4}$$

Since the second signal has traveled a longer distance than the reference signal, the phase of the incoming signal is different. With this measured phase difference ϕ and the wavelength of the signal modulation λ, the distance d of the object can be calculated by

$$d = \frac{\phi}{360} \cdot \frac{\lambda}{2}, \tag{5}$$

where λ is the wavelength of the signal modulation.

Since the phase of a modulated signal is periodic with a cycle of 360° or 2π respectively, it is not possible to determine in which cycle of the modulation the measured phase is located. This is essential for distance measurement and a real sensor needs to deal with this problem. How this can be done is explained in the section 2.2.2.

2.2. 3D Laser Scanners

Laser scanners are very common sensors for range measurement in many fields of application and belong to the group of sensors based on the time-of-flight principle. 2D laser scanners are being very common in robotic applications since a long period of time. They are based on a laser source, a rotating mirror and a photosensitive sensor, whereat the laser source and the photosensitive sensor are building a one dimensional time-of-flight range sensor. For reaching the second dimension, the mirror deflects the laser beam continuously while it is rotating. Such a scanner could scan a circumferential area, but normally it is reduced to a smaller angle, e.g., 240°. For many applications where three-dimensional objects act in or interact with a three-dimensional environment, a 2D range sensor is not sufficient and a 3D sensor is required.

A 3D laser scanner can be realized in different ways, either the laser signal is deflected in two directions by a mirror instead of only one, or one can use more than one laser source deflected by a mirror [Ibeo, 2007]. A third option for developing 3D laser scanner was used by several groups, e.g. [Fraunhofer IAIS, 2007], [RTS, 2007]. Here commercially available 2D laser scanners were pivot-mounted and rotated while they are scanning, which gives three-dimensional data. Two of them will be described in more detail in the following subsections.

2.2.1 3DLS – A 3D Laser Scanner

The 3DLS is based on a SICK 2D laser scanner which is pivot-mounted in the horizontal axis [Fraunhofer IAIS, 2007]. This axis is driven by a servo motor to extend the standard scanner to a 3D laser scanner. The underlying measurement principle is the *Impulse Time-of-Flight*. The laser source is an infrared laser with a wavelength of λ = 905nm. A maximum field of view of 180° horizontal and 124° vertical can be scanned with an angular resolution of 1/4th degree and a precision of ±15mm. Depending on the chosen resolution, the scan time for a full 3D scan can vary from 3.2s for a resolution of 1° to 26.64s for the maximum resolution of 0.25°. With a size of 284 x 286 x 166mm (width x height x depth) and a weight of 7.4kg the 3DLS can be used for medium-sized or large mobile robotic systems as well as for stationary applications.
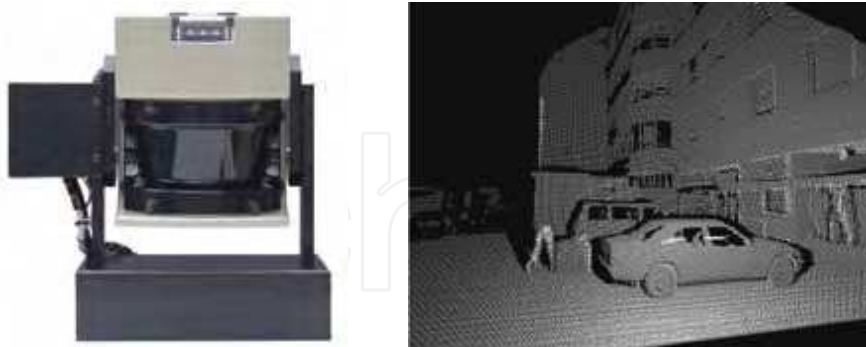

Figure 3. Left image: The 3D laser scanner system 3DLS with a SICK LMS291 laser scanner. Right image: 3D scan taken with the 3DLS

The 3DLS is available as indoor and outdoor version, which mainly differ in the operation temperature and the maximum range of the scanner. The maximum range is almost exclusively limited by the amount of light which is reflected by the measured object.

Theoretically, the maximum range of both versions is 80m but with an object reflectivity of only 10% (e.g. a black cardboard) the maximum range is specified with 10m for the indoor version and 30m for the outdoor version. Another property of the 3DLS which influences the quality of the resulting data is the diameter of the sent laser impulse, which signal increases over the traveled distance. For the 3DLS, the diameter of the laser impulse increased from around 1cm at the beginning to 15cm at a distance of 30m. This causes inaccuracies if the laser hits an edge and therefore is partially reflected from different distances.

The result of a 3D laser scan is a 3D point cloud, which can be seen in figure 3. The technical data of the 3DLS are summarized in the following table.

Resolution	721 x 517 scan points (1/4th °)
Field of view	180° x 124° (hor. x vert.)
Range	80m (30m with 10% reflectance)
Frame Rate	Up to 1 fps (reduced resolution)
Dimensions (mm)	284 (W) x 286 (H) x 166 (D)
Weight	7.4 kg
Power supply	24V dc

Table 1. Property table of the Fraunhofer IAIS 3DLS

2.2.2 A 3D laser scanner based on the Hokuyo URG

Figure 4. Left image: Hokuyo 3D Scanner. Right image: Scan taken with the Hokuyo 3D Scanner. Note that a field of view of 248° is reached by only one rotating servo

Similar to the 3DLS, this scanner is based on a 2D laser scanner, the Hokuyo URG-04LX [Kawata et al., 2005], [Hokuyo Automatic, 2007]. Since it is very small and light weighted it is directly mounted on a servo drive to get the additional rotation axis. By using a pan-tilt-head (cf. figure 4), different scanning setups are possible. In difference to the 3DLS, this scanner measures the range by using the *phase difference* principle. For generating the modulated light signal, an infrared laser diode with a wavelength of λ = 785nm is used. As described in chapter 2.1.2 it is not possible to detect if the measured phase difference is more than one cycle period and therefore out of the maximum measurement range. To handle

that problem, two laser signals with different modulation frequencies are emitted alternately. Both phase differences are measured separately and used for determining the real distance of the measured object.

The maximum apex angle of this 3D laser scanner is 270° horizontally and 248° vertically with a resolution of 0.36°. Depending on the measured distance, the precision is at least ±2% of the distance. A full resolution scan takes 50 seconds. The technical data are summarized in the following table.

Resolution	1000 x 667 scan points (0.36°)
Field of view	270° x 248° (hor. x vert.)
Range	4,095m
Frame Rate	0,02 fps (full resolution)
Dimensions (mm)	80 (W) x 120 (H) x 75 (D)
Weight	350 g
Modulation frequencies	46.55 MHz and 53.2 MHz

Table 2. Property table of the Hokuyo URG based 3D laser scanner

2.3. 3D Cameras

These devices belong to the group of time-of-flight sensors. They use the phase-shift principle to determine distances. While the environment is being illuminated with infrared flashes, the reflected light is measured by a CCD- or CMOS-sensor or a combined technology. Amplitude data is represented by the incoming wave's amplitude, intensity by its offset (i.e. the background light) and distance by its phase shift. For the experiments in section 4, we have used a SwissRanger SR-2 device that can be seen in figure 5.

Figure 5. Left image: SwissRanger SR-2 device mounted on a pan-tilt unit. Right image: Sample image captured from a SwissRanger SR-2 device. The image is color coded (see color bar on the right side)

The SwissRanger SR-2 provides amplitude data, intensity data and distance data. All measurements are being organized by a FPGA, which provides an USB interface to access the data. The FPGA can be configured by setting one or more of its eleven registers. The most important register concerns the adjustment of the integration time, since the SR-2 does not provide an automatic integration time controller by itself (the follow-up model SR-3000 does). It ranges from 1 to 255, which are multiples of 255 μs. Finding the optimal value will be investigated in section 3.2.

Please pay some attention on table 3 and compare it with table 1. The comparison of the SwissRanger SR-2 and the Sick LMS device is important for the experiment in section 4.1.

Resolution	124 x 160
Field of view	43˚ x 46˚ (hor. x vert.)
Range	7.5 m
Frame Rate	Up to 30 fps
Dimensions (mm)	135 (W) x 45 (H) x 32 (D)
Weight	0.2 kg

Table 3. Property table of the SwissRanger SR-2 device

2.4. Stereo Cameras

Stereo vision is a mature technology in computer vision. Depth-measurements with stereo cameras have been investigated since decades, e.g. [Lucas & Kanade, 1981]. There are also lots of pre-calibrated systems available, but this technology still needs a great deal of performance since point correspondences from the left and the right image have to be found for enabling the calculation of depth information. For homogeneous regions it is difficult to find the correct correspondences. If these regions are bounded by unambiguous features, i.e. textured regions or edges and borders respectively an iteration scheme can be used to relax the correspondences of these features over the whole image. Otherwise there is no way to calculate any depth information. That is why related techniques 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 light conditions. Due to this drawback, this chapter discusses not any passive visual sensor system any further.

3. 3D Cameras – A Step forward in Computer Vision

This section discusses the technology of 3D cameras more detailed since it has the application potential for tackling dynamics in the field of 3D computer vision. For the investigations following below the SwissRanger SR-2 was used.

3.1. Challenges and Limitations

The adjustment of 3D cameras to dynamic scenes is still a difficult task. The accuracy is influenced by a couple of parameters. Some of them are predefined by the design of hardware and cannot be influenced by the user. Anyway, these parameters should be mentioned in this section to facilitate the understanding for the presented effects in the remainder of this chapter.

First of all, the accuracy is proportional to the modulation frequency. Doubling the frequency doubles the accuracy. But the frequency also determines the unambiguous range, which can be seen in equation (6).

$$R = \frac{c}{2 \cdot f_m},$$
(6)

where R is the unambiguity interval, c the speed of light and f_m the modulation frequency. A camera with a frequency of 20 MHz provides an unambiguous range of 7.5 m. A lower

frequency provides a higher range but less accuracy. To satisfy both criteria multiple frequencies can be used. For instance, this technology is currently used by the PMD[vision]® A2 from PMDTec.

Since the principle is based on integrating discharged electrons from incoming light, the optical power also influences the reachable accuracy. These electrons are collected within a conversion capacity, which can result in oversaturation, if the integration time is too high [Lange, 2000]. Both mentioned manufacturers in this chapter use a burst mode to increase the power output for short intervals at the same energy level over time.

For an application, the best measurement capability has to be adjusted by the integration time. This value has to be high enough to provide a high signal level, but low enough to avoid oversaturation. Oversaturation is indicated by both, the intensity and the amplitude data. Theoretically the relation between intensity and amplitude is constant as shown in figure 6, but unfortunately it shows a small deviation due to a non-ideal sinusoidal wave emitted by the sensor's LEDs [Lange, 2000].
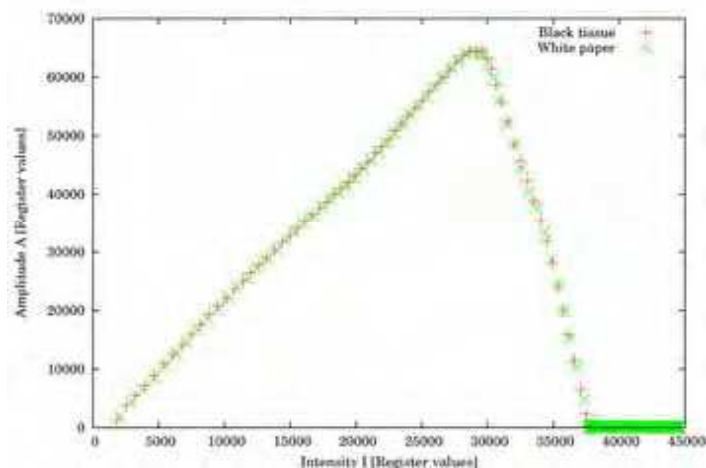

Figure 6. Relation between intensity and amplitude data of the SwissRanger SR-2 device in dependency of the integration time. Note that a higher integration time is also indicated by a higher intensity value. The amplitude is raising linearly until oversaturation occurs

There are also a number of other noise sources which theoretically influence the reachable accuracy. It is out of the scope of this chapter to explain all noise effects. A good theoretically work explaining them in detail can be found in [Lange, 2000]. This work also describes the dominance of shot noise that cannot be suppressed and, therefore, limits the theoretically reachable signal-to-noise ratio and the accuracy involved. Hence, the standard deviation ΔR is approximately given as:

$$\Delta R = \frac{c}{4 \cdot \pi \cdot f_m} \cdot \frac{\sqrt{I_l + I_b}}{\sqrt{2} \cdot A}, \tag{7}$$

where A is the amplitude and I the intensity. The intensity value is composed of the reflected constant component I_l of the LED illumination and the background illumination I_b.

As a rule, it can be said that the proper saturation of a pixel's capacitance provides its best accuracy. The emitted light is uniformly distributed (only approximately, see [Gut, 2004]) on a surface proportional to the quadratic distance. Therefore, the reflected intensity is also proportional to the quadratic distance, whereas the received background light (caused by sunlight) is independent of it [Schneider, 2003]. For both constituent parts the standard deviation yields a different dependency on the object's distance. If one of these components is dominant the standard deviation has the following characteristic[1]:

$$\Delta R \sim \begin{cases} r, if\ I_l \gg I_b \\ r^2, if\ I_l \ll I_b \end{cases} \tag{8}$$

For indoor applications with less background illumination case one can be assumed. This relation will be relevant for the experiments mentioned below.

3.2. Tackling Environment Dynamics

Figure 7 shows the same scene taken twice with the SwissRanger SR-2 device at two different integration times. The integration time for the measurement shown in the left figure was not adjusted properly with respect to the near object. The bothered area of the hand indicates that effect. But this measurement might provide better values in the background area.
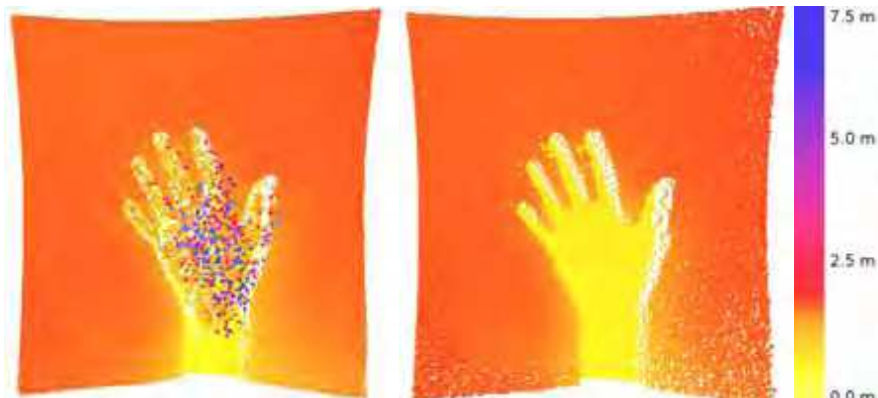

Figure 7. Two sample distance measurements of a close hand in false color representation. The left measurement was taken with an integration time of 15 ms, which is definitely too high for the near object. The right measurement was taken with an integration time of 4 ms and fits better for that scene

3.2.1 Setting up the integration time

The diffuse reflectivity of objects is an important parameter for precise measurements. Typically, a scene comprises objects with different reflectivity. It can vary from 2% for black rubber tire up to 100% for white paper at a wavelength of 900nm [Lange, 2000]. Let us assume scenarios containing only a small single object with high reflectance close to the sensor to explain the compromise that has to be met. Since the integration time can only be

---

[1] Note that also the amplitude has a quadratic dependency on the distance.

adjusted for all pixel elements together, one might guess that it is the best strategy to avoid each pixel from oversaturation. Focusing the small object will most likely decrease the accuracy for the remaining scene. This also means, that the signal level for objects with low diffuse reflectivity will be low if objects with high reflectivity are in the same range of vision during measurement.

One suitable method is to merge multiple captures at different integration times. It reduces the frame rate but increases the dynamic range.

In [May, 2006] we have presented an alternative integration time controller based on mean intensity measurements. This solution was empirically found and showed a suitable dynamic range for our experiments without affecting the frame rate. It also alleviates the effects of small bothering areas. The averaged amplitude in dependency of intensity can be seen in figure 8.
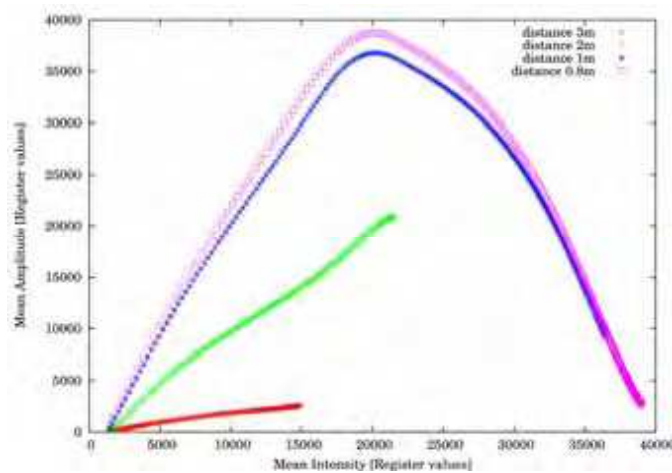

Figure 8. Relation between mean amplitude and mean intensity. Note that the characteristic is now a mixture of the characteristic of each single pixel (cf. figure 6)

We used a proportional closed-loop controller to adjust the integration time from one frame to the next as shown in the following itemization.

The control deviation variable I_a was assigned with a value of 15000 for the illustrations in this chapter. It has been chosen conservatively with respect to the characteristic shown in figure 8.

1. Calculate the mean intensity $\overline{I_t}$ from the intensity dataset I_t at time t.

2. Determine control deviation $D_t = \overline{I_t} - I_a$.

3. Update control variable $c_{t+1} = -V_p \cdot D_t + c_t$ for grabbing the next frame, where c_t and c_{t+1} are the integration times for two frames following one another, V_p the proportional closed loop deviation parameter and c_0 a suitable initial value.

Independent of the chosen control method, the integration time has always to be adjusted with respect to the application. A change of integration time causes an apparent motion considering the distance measurement values. Therefore, it is necessary for the application to take the presence of control deviation into account while using an automatic integration time controller.

The newest model from Mesa Imaging, the SwissRanger SR-3000 provides an automatic integration time exposure based on the amplitude values. For most scenes it works properly. In some cases of fast scene change it could occur that a proper integration time cannot be found. This is up to the missing intensity information due to the backlight suppression on chip. The amplitude diagram does not provide a non-ambiguous working point. A short discussion on the backlight suppression will be given in section 3.3.

3.2.2 Consideration of accuracy

It is not possible to guarantee certain accuracies for measurements of unknown scenes, since they are affected by the influences mentioned above. However, the possibility to evolve the accuracy information for each pixel eases that circumstance. In section 4 two examples using this information will be explained. For determining the accuracy equation (7) is used. Assuming that the parameters of the camera (in general this is the integration time for users) are optimally adjusted, the accuracy only depends on the object's distance and its reflectivity. For indoor applications with less background illumination, the accuracy is linearly decreasing (see equation (8)). Applying a simple threshold is one option for filtering out inaccurate parts of an image. Setting a suitable threshold primarily depends on the application. Lange stated with respect to the dependency between accuracy and distance [Lange, 2000]: "This is an important fact for navigation applications, where a high accuracy is often only needed close to the target". This statement does not hold for every other application like mapping, where unambiguousness is essential for registration. Unambiguous tokens are often distributed over the entire scene. Higher distances between these tokens provide geometrically higher accuracies for the alignment of two scans. After this consideration, increasing the threshold linearly with the distance for indoor applications suggests itself. This approach enlarges the information gain from the background and can be seen in figure 9.

A light source in the scene decreases the reachable accuracy. The influence of the accuracy threshold can be seen in figure 10. Bothered areas are reliably removed. The figure shows also that the small bothering area of the lamp does not much influence the integration time controller based on mean intensity values, even so that the surrounding area is determined precisely.

Figure 9. Two images taken with a SwissRanger SR-2 device of the same scene. Left image: without filtering. Right image: with accuracy filter. Only data points with an accuracy better than 50mm are remaining

Figure 10. Influence of light emitting sources. Top row: The light source is switched off. Lower Row: The light source is switched on. Note that the bothered area could reliably be detected

3.3. Latest Improvements and expected Innovations in Future

Considering equation (7) a large background illumination ($I_b \gg I_l$) highly affects the sensor's accuracy by increasing the shot noise and lowering its dynamics. Some sensors nowadays are equipped with some background light suppression functionalities, e.g. spectral filters or circuits for constant component suppression, which are increasing the signal-to-noise ratio [Moeller et al., 2005], [Buettgen et al., 2006].

Suppressing the background signal has one drawback. The amplitude represents the infrared reflectivity and not the reflectivity we sense as human-beings. This might take effects on computer vision systems inspired by our human visual sense, e.g. [Frintrop, 2006]. Some works in the past had also proposed a circuit structure for a pixel-wise-integration capability [Schneider, 2003], [Lehmann, 2004]. Unfortunately, this technology did not become widely accepted due to a lower fill-factor. Lange explained the importance of the optical fill factor as follows [Lange, 2000]: "The optical power of the modulated illumination source is both expensive and limited by eye-safety regulations. This requires the best possible optical fill factor for an efficient use of the optical power and hence a high measurement resolution."

4. 3D Vision Applications

This section investigates the practical influence of upper mentioned thoughts by presenting some typical applications in the domain of autonomous robotics currently investigated by us. Since 3D cameras are comparatively new to other 3D sensors like laser scanners or stereo cameras, the porting of algorithms defines a novelty per se; e.g. one of the first 3D maps

created with registration approaches mostly applied to laser scanner systems up to now was presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems in 2006 [Ohno, 2006]. The difficulties to come across with these sensors are discussed in this section. Furthermore, a first examination on the capabilities for tackling environment dynamics will follow.

4.1. Registration of 3D Measurements

One suitable registration method for range data sets is called the Iterative Closest Points (ICP) algorithm and was introduced by Besl and McKay in 1992 [Besl & McKay, 1992]. For the readers convenience a brief description of this algorithm is repeated in this section. Given two independently acquired sets of 3D points, M (model set) and D (data set), which correspond to a single shape, we aim to find the transformation consisting of a rotation R and a translation t which minimizes the following cost function:

$$E(R,t) = \sum_{i=1}^{|M|} \sum_{j=i}^{|D|} \omega_{i,j} \left\| m_i - (Rd_j + t) \right\|^2. \tag{9}$$

$\omega_{i,j}$ is assigned *1* if the *i-th* point of M describes the same point in space as the *j-th* point of D. Otherwise $\omega_{i,j}$ is *0*. Two things have to be calculated: First, the corresponding points, and second, the transformation *(R,t)* that minimizes *E(R,t)* on the base of the corresponding points. The ICP algorithm calculates iteratively the point correspondences. In each iteration step, the algorithm selects the closest points as correspondences and calculates the transformation *(R,t)* for minimizing equation (9). The assumption is that in the last iteration step the point correspondences are correct. Besl and McKay prove that the method terminates in a minimum [Besl & McKay, 1992]. However, this theorem does not hold in our case, since we use a maximum tolerable distance d_{max} for associating the scan data. Such a threshold is required though, given that 3D scans overlap only partially. The distance and the degree of overlapping have a non-neglective influence of the registration accuracy.

4.2. 3D Mapping – Invading the Domain of Laser Scanners

The ICP approach is one upon the standard registration approaches used for data from 3D laser scanners. Since the degree of overlapping is important for the registration accuracy, the huge field of view and the high range of laser scanners are advantages over 3D cameras (compare table 1 with table 3). The following section describes our mapping experiments with the SwissRanger SR-2 device.

The image in figure 11 shows a single scan taken with the IAIS 3D laser scanner. The scan provides a 180 degree field of view. Getting the entire scene into range of vision can be done by taking only two scans in this example. Nevertheless, a sufficient overlap can be guaranteed to register both scans. Of course there are some uncovered areas due to shadowing effects, but that is not the important fact for comparing the quality of registration. A smaller field of view makes it necessary to take more scans for the coverage of the same area within the range of vision. The image in figure 12 shows an identical scene taken with a SwissRanger SR-2 device. There were 18 3D images necessary for a circumferential view with sufficient overlap. Each 3D image was registered with its previous 3D image using the ICP approach.
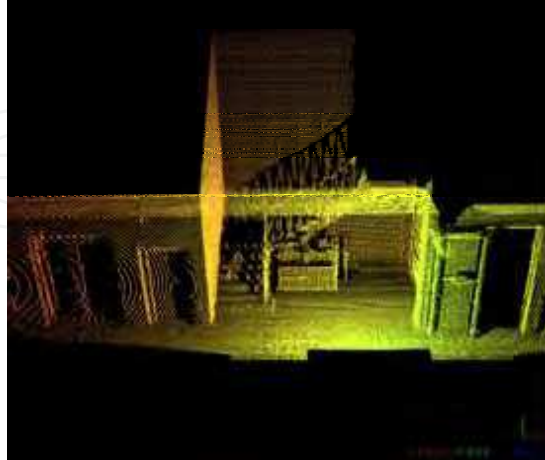
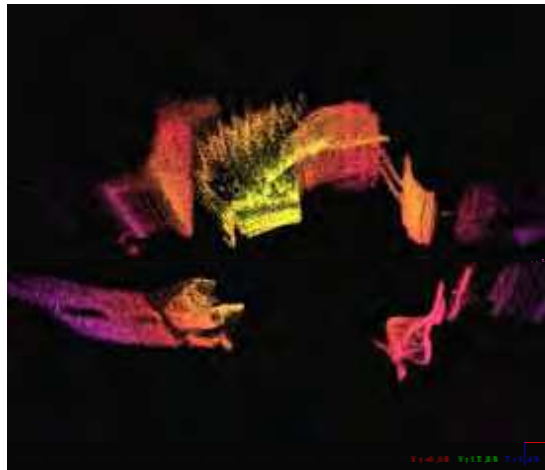Figure 11. 3D scan taken with an IAIS 3D laser scanner

Figure 12. 3D map created from multiple SwissRanger SR-2 3D images. The map was registered with the ICP approach. Note the gap at the bottom of the image, that indicates the accumulating error

4.2.1. "Closing the Loop"

The registration of 3D image sequences causes a non-neglective accumulation error. This effect is represented by the large gap at the bottom of the image in figure 12. These effects have also been investigated in detail for large 3D maps taken with 3D laser scanners, e.g. in [Surmann et al., 2004], [Cole & Newman, 2006]. For a smaller field of view these effects occur faster, because of the smaller size of integration steps. Determining the closure of a loop can be used in these cases to expand the overall error on each 3D image. This implies that the present captured scene has to be recognized to be already one of the previous captured scenes.

4.2.2. "Bridging the Gap"

The second difficulty for the registration approach is that a limited field of view makes it more unlikely to measure enough unambiguous geometric tokens in the space of distance data or even sufficient structure in the space of grayscale data (i.e. amplitude or intensity). This issue is called the aperture problem in computer vision. It occurs for instance for images taken towards a huge homogeneous wall (see [Spies et al., 2002] for an illustration). In the image of figure 12 the largest errors occurred for the images taken along the corridor. Although points with a decreasing accuracy depending on the distance (see section 3.2.2) were considered, only the small areas at the left and the right border contained some fairly accurate points, which made it difficult to determine the precise pose. This inaccuracy is mostly indicated in this figure by the non-parallel arrangement of the corridor walls. The only feasible solution to this problem is a utilization of different perspectives.

4.3. 3D Object Localization

Object detection is a highly investigated field of research since a very long period of time. A very challenging task here is to determine the exact pose of the detected objects. Either this information is just implicitly available since the algorithm is not very stable against object transformations or the pose information is explicit but not very precise and therefore not very reliable. For reasoning about the environment it may be enough to know which objects are present and where they are located but especially for manipulation tasks it is essential to know the object pose as precise as possible. Examples for such applications are ranging from "pick and place" tasks of disordered components in industrial applications to handling task of household articles in service-robotic applications.

In comparison to color camera based systems the use of 3D range sensors for object localization provide much better results regarding the object pose. For example Nuechter et al. [Nuechter et al., 2005] presented a system for localizing objects in 3D laser scans. They used a 3D laser scanner for the detection and localization of objects in office environments. Depending on the application one drawback of this approach is the time consuming 3D laser scan which needs at least 3.2 seconds for a single scan (cf. table 1). Using a faster 3D range sensor would increase the timing performance of such a system essentially and thus open a much broader field of applications.

Therefore Fraunhofer IAIS is developing an object localization system which uses range data from a 3D camera. The development of this system is part of the DESIRE research project which is founded by the German Federal Ministry of Education and Research (BMBF) under grant no. 01IME01B. It will be integrated into a complex perception system of a mobile service-robot. In difference to the work of Nuechter et al. the object detection in the DESIRE perception system is mainly based on information from a stereo vision system since many objects are providing many distinguishable features in their texture. With the resulting hypothesis of the object and it's estimated pose a 3D image of the object is taken and together with the hypothesis it is used as input for the object localization.

The localization itself is based on an ICP based scan matching algorithm (cf. section 4.1). Therefore each object is registered in a database with a point cloud model. This model is used for matching with the real object data. For determining the pose, the model is moved into the estimated object pose and the ICP algorithm starts to match the object model and the object data. The real object pose is given by a homogeneous transformation. Using this

object localization system in real world applications brings some challenges, which are discussed in the next subsection.

4.3.1 Challenges

The first challenge is the pose ambiguities of many objects. Figure 13 shows a typical object for a home service-robot application, a box of instant mashed potatoes. The cuboid shape of the box has three plains of symmetry which results in the ambiguities of the pose. Considering only the shape of the object, very often the result of the object localization is not a single pose but a set of possible poses, depending on the number of symmetry planes. For determining the real pose of an object other information than only range data are required, for example the texture. Most 3D cameras additionally providing gray scale images which give information about the texture but with the provided resolution of around 26.000 pixels and an aperture angle of around 45° the resolution is not sufficient enough for stable texture identification. Instead, e.g., a color camera system can be used to solve this ambiguity. This requires a close cooperation between the object localization system and another classification system which uses color camera images and a calibration between the two sensor systems. As soon as future 3D cameras are providing higher resolutions and maybe also color images, object identification and localization can be done by using only data from a 3D camera.

Figure 13. An instant mashed potatoes box. Because of the symmetry plains of the cuboid shape the pose determination gives a set of possible poses. Left: Colour image from a digital camera. Right: 3D range image from the Swissranger SR-2

Another challenge is close related to the properties of 3D cameras and the resulting ability to provide precise range images of the objects. It was shown that the ICP based scan matching algorithm is very reliable and precise with data from a 3D laser scanner, which are always providing a full point cloud of the scanned scene [Nuechter, 2006], [Mueller, 2006]. The accuracy is static or at least proportional to the distance. As described in section 3.2.2 the accuracy of 3D camera data is influenced by several factors. One of these factors for example is the reflectivity of the measured objects. The camera is designed for measuring diffuse light reflections but many objects are made of a mixture of specular and diffuse reflecting materials. Figure 14 shows color images from a digital camera and range images from the Swissrange SR-2 of a tin from different viewpoints. The front view gives reliable range data of the tin since the cover of the tin is made of paper which gives diffuse reflections. In the second image the cameras are located a little bit above and the paper cover as well as high reflecting metal top is visible in the color image. The range image does not show the top

since the calculated accuracy of these data points is less than 30 mm. This is a loss of information which highly influences the result of the ICP matching algorithm.
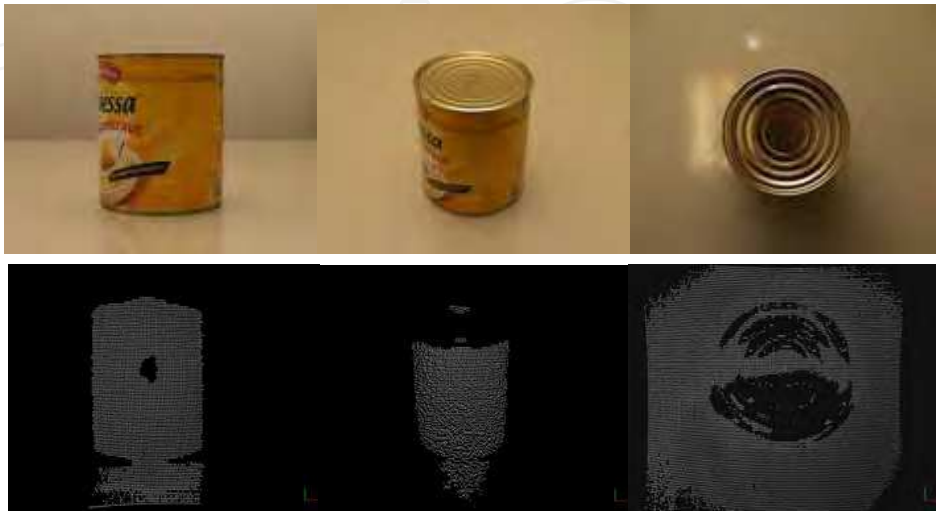

Figure 14. Images of a tin from different view points. Depending on the reflectivity of the objects material the range data accuracy is different. In the range images all data points with a calculated accuracy less than 30mm are rejected. Left: The front view gives good 3D data since the tin cover reflects diffuse. Middle: From a view point above the tin, the cover as well as the metal top is visible. The high reflectivity of the top results in bad accuracy so that only the cover part is visible in the range image. Right: From this point of view, only the high metal top is visible. In the range image only some small parts of the tin are visible

4.4. 3D Feature Tracking

Using 3D cameras to full capacity necessitates taking advantage of their high frame rate. This enables the consideration of environment dynamics. In this subsection a feature tracking application is presented to give an example of applications that demand high frame rates. Most existing approaches are based on 2D grayscale images from 2D cameras since they were the only affordable sensor type with a high update rate and resolution in the past. An important assumption for the calculation of features in grayscale images is called intensity constancy assumption. Changes in intensity are therefore only caused by motion. The displacement of two images is also called optical flow. An extension to 3D can be found in [Vedula et al., 1999] and [Spies et al., 2002]. The intensity constancy assumption is being combined with a depth constancy assumption. The displacement of two images can be calculated more robustly. This section will not handle scene flow. However the depth value of features in the amplitude space should be examined so that the following two questions are answered:

- Is the resolution and quality of the amplitude images from 3D cameras good enough to apply feature tracking kernels?
- How stable is the depth value of features gathered in the amplitude space?

To answer these questions a Kanade-Lucas-Tomasi (KLT) feature tracker is applied [Shi, 1994]. This approach locates features considering the minimum eigenvalue of each 2x2

gradient matrix. Tracking features frame by frame is done by an extension of previous Newton-Raphson style search methods. The entire approach also considers multi-resolution to enlarge possible displacements between the two frames. Figure 15 shows the result of calculating features in two frames following one another. Features in the present frame (left feature) are connected with features from the previous frame (right feature) with a thin line. The images in figure 15 show that many edges in the depth space are associated with edges in the amplitude space. The experimental standard deviation for that scene was determined by taking the feature's mean depth value of 100 images. The standard deviation was then calculated from 100 images of the same scene. These experiments have been performed two times, first without a threshold and second with an accuracy threshold of 50mm (cf. formula 7). The results are shown in table 4 and 5.

Experimental standard deviation σ = 0.053m, Threshold ΔR = ∞				
Feature #	Considered	Mean Dist [m]	Min Dev [m]	Max Dev [m]
1	Yes	-2.594	-0.112	0.068
2	Yes	-2.686	-0.027	0.028
3	Yes	-2.882	-0.029	0.030
4	Yes	-2.895	**-0.178**	**0.169**
5	Yes	-2.731	**-0.141**	**0.158**
6	Yes	-2.750	-0.037	0.037
7	Yes	-2.702	**-0.174**	**0.196**
8	Yes	-2.855	**-0.146**	**0.119**
9	Yes	-2.761	-0.018	0.018
10	Yes	-2.711	-0.021	0.025

Table 4. Distance values and deviation of the first ten features calculated from the scene shown in the left image of figure 15 with no threshold applied

Experimental standard deviation σ = 0.017m, Threshold ΔR = 50mm				
Feature #	Considered	Mean Dist [m]	Min Dev [m]	Max Dev [m]
1	Yes	-2.592	-0.110	0.056
2	Yes	-2.684	-0.017	0.029
3	Yes	-2.881	-0.031	0.017
4	No	-2.901	**-0.158**	**0.125**
5	Yes	-2.733	**-0.176**	**0.118**
6	Yes	-2.751	-0.025	0.030
7	No	-2.863	**-0.185**	**0.146**
8	No	-2.697	**-0.169**	**0.134**
9	Yes	-2.760	-0.019	0.015
10	Yes	-2.711	-0.017	0.020

Table 5. Distance values and deviation of the first ten features calculated from the scene shown in the left image of figure 15 with a threshold of *50mm*

The reason for the high standard deviation is the noise criterion for edges. The signal reflected by an edge is a mixture of the background and object signal. A description of this

effect is given in [Gut, 2004]. Applying an accuracy threshold alleviates this effect. The standard deviation is decreased significantly. This approach has to be balanced with the number of features found in an image. Applying a more restrictive threshold might decrease the number of features too much. For the example described in this section an accuracy threshold of $\Delta R = 10mm$ decreases the number of features to *2* and the experimental standard deviation σ to *0.01m*.
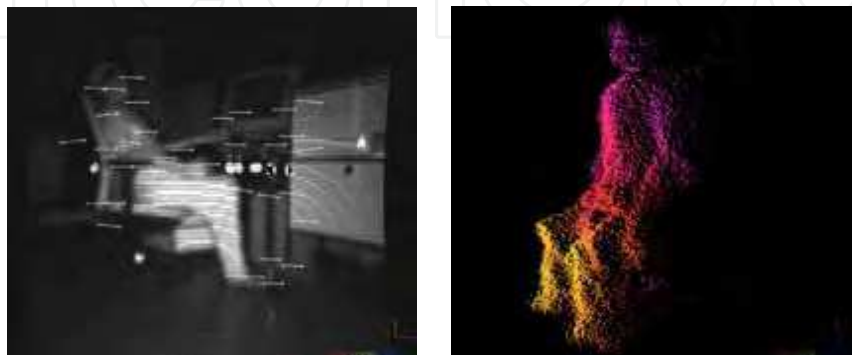

Figure 15. Left image: Amplitude image showing the tracking of KLT-features from two frames following one another. Right image: Side view of a 3D point cloud. Note the appearance of jump edges at the border area

5. Summary and Future work

First of all, a short comparison of range sensors and their underlying principles was given. The chapter further focused on 3D cameras. The latest innovations have given a significant improvement for the measurement accuracy, wherefore this technology has attracted attention in the robotics community. This was also the motivation for the examination in this chapter. On this account, several applications were presented, which represents common problems in the domain of autonomous robotics.

For the mapping example of static scenes, some difficulties have been shown. The low range, low apex angle and low dynamic range compared with 3D laser scanners, raised a lot of problems. Therefore, laser scanning is still the preferred technology for this use case.

Based on the first experiences with the Swissranger SR-2 and the ICP based object localization, we will further develop the system and concentrate on the reliability and the robustness against inaccuracies in the initial pose estimation. Important for the reliability is knowledge about the accuracy of the determined pose. Indicators for this accuracy are, e.g., the number of matched points of the object data or the mean distance between found model-scene point correspondences.

The feature tracking example highlights the potential for dynamic environments. Use cases with requirements of dynamic sensing are predestinated for 3D cameras. Whatever, these are the application areas 3D cameras were once developed.

Our ongoing research in this field will concentrate on dynamic sensing in future. We are looking forward to new sensor innovations!

6. References

Besl, P. & 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, ISSN: 0162-8828

Buettgen, B.; Oggier, T.; Lehmann, M.; Kaufmann, R.; Neukom, S.; Richter, M.; Schweizer, M.; Beyeler, D.; Cook, R.; Gimkiewicz, C.; Urban, C.; Metzler, P.; Seitz, P.; Lustenberger, F. (2006). High-speed and high-sensitive demodulation pixel for 3D imaging, In: Three-Dimensional Image Capture and Applications VII. *Proceedings of SPIE*, Vol. 6056, (January 2006) pp. 22-33, DOI: 10.1117/12.642305

Cole, M. D. & Newman P. M. (2006). Using Laser Range Data for 3D SLAM in Outdoor Environments, In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1556-1563, Orlando, Florida, USA, May 2006

CSEM SA (2007), *SwissRanger SR-3000 - miniature 3D time-of-flight range camera*, Retrieved January 31, 2007, from http://www.swissranger.ch

Frintrop, S. (2006). *A Visual Attention System for Object Detection and Goal-Directed Search*, Springer-Verlag, ISBN: 3540327592, Berlin/Heidelberg

Fraunhofer IAIS (2007). *3D-Laser-Scanner,* Fraunhofer Institute for Intelligent Analysis and Information Systems, Retrieved January 31, 2007, from http://www.3d-scanner.net

Gut, O. (2004). *Untersuchungen des 3D-Sensors SwissRanger*, Eidgenössische Technische Hochschule Zürich, Retrieved January 21, 2007, from http://www.geometh.ethz.ch/publicat/diploma/gut2004/Fehlereinfluesse/index_fe.html

Hokuyo Automatic (2007), *Scanning laser range finder for robotics URG-04LX*, Retrieved January 31, 2007, from http://www.hokuyo-aut.jp/products/urg/urg.htm

Ibeo Automobile Sensor GmbH (2007), *Ibeo ALASCA XT Educational System*, Retrieved January 31, 2007, from http://www.ibeo-as.com/deutsch/products_alascaxtsingle_educational.asp

Kawata, H.; Ohya, A.; Yuta, S.; Santosh, W. & Mori, T. (2005). Development of ultra-small lightweight optical range sensor system, *International Conference on Intelligent Robots and Systems 2005,* Edmonton, Alberta, Canada, August 2005.

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

Lehmann, M.; Buettgen, B.; Kaufmann, R.; Oggier, T.; Stamm, M.; Richter, M.; Schweizer, M.; Metzler, P.; Lustenberger, F.; Blanc, N. (2004). *CSEM Scientific & technical Report 2004*, CSEM Centre Suisse d'Electronique et de Microtechnique SA, Retrieved January 20, 2007, from http://www.csem.ch/corporate/Report2004/pdf/SR04-photonics.pdf

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints, *International Journal of Computer Vision,* Vol. 60, No. 2, (November 2004) pp. 91-110, ISSN: 0920-5691

Lucas, B. D. & Kanade, T. (1981). An Interative Image Registration Technique with an Application to Stereo Vision, In *Proceedings of the 7th International Conference on Artificial Intelligence* (IJCAI), pp. 674-679, Vancouver, British Columbia, August 1981

May, S.; Werner, B.; Surmann, H.; Pervoelz, K. (2006). 3D time-of-flight cameras for mobile robotics, *In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS), pp. 790-795, Beijing, China, October 2006

Moeller, T.; Kraft H.; Frey, J.; Albrecht, M.; Lange, R. (2005). *Robust 3D Measurement with PMD Sensors*, PMDTechnologies GmbH. Retrieved January 20, 2007, from http://www.pmdtec.com/inhalt/download/documents/RIM2005-PMDTech-Robust3DMeasurements.pdf.

Mueller, M.; Surmann, H.; Pervoelz, K. & May, S. (2006). The Accuracy of 6D SLAM using the AIS 3D Laser Scanner, In *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems* (MFI), Heidelberg, Germany, September 3-6, 2006

Nuechter A., Lingemann K., Hertzberg J. & Surmann, H. (2005). Accurate Object Localization in 3D Laser Range Scans, In *Proceedings of the 12th International Conference on Advanced Robotics* (ICAR '05), ISBN 0-7803-9178-0, pages 665 - 672, Seattle, USA, July 2005.

Nuechter A. (2006). Semantische dreidimensionale Karten für autonome mobile Roboter, *Dissertation*, Akademische Verlagsgesellschaft Aka, ISBN: 3-89838-303-2, Berlin

Ohno, K.; Nomura, T.; 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), pp. 5279-5285, Beijing, China, October 2006

PMD Technologies (2007), "*PMD Cameras*", Retrieved January 31, 2007, from http://www.pmdtec.com/e_inhalt/produkte/kamera.htm

RTS Echtzeitsysteme (2007), *Mobile Serviceroboter*, Retrieved January 31, 2007, from http://www.rts.uni-hannover.de/index.php/Mobile_Serviceroboter

Schneider, B. (2003). Der Photomischdetektor zur schnellen 3D-Vermessung für Sicherheitssysteme und zur Informationsübertragung im Automobil, *Dissertation*, University of Siegen, 2003

Shi, J. & Tomasi, C. (1994). Good Features to Track, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), pp. 595-600, Seattle, June 1994

Spies, H.; Jaehne, B.; Barron, J. L. (2002). Range Flow Estimation, *Computer Vision Image Understanding* (CVIU2002) 85:3, pp.209-231, March, 2002

Surmann, H.; Nuechter, A.; Lingemann K. & Hertzberg, J. (2003). An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments, *Robotics and Autonomous Systems, 45, (December 2003)* pp. 181-198

Surmann, H.; Nuechter, A.; Lingemann, K. & Hertzberg, J. (2004). 6D SLAM A Preliminary Report on Closing the Loop in Six Dimensions, In *Proceedings of the 5th IFAC Symposium on Intelligent Autonomous Vehicles* (IAV), Lisabon, Portugal, July 2004

Thrun, S.; Fox, D. & Burgard, W. (2000). 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), pp. 321-328, ISBN: 0-7803-5886-4, San Francisco, February 1992

Vedula, S.; Baker, S.; Rander, P.; Collins, R. & Kanade, T. (1999). Three-Dimensional Scene Flow, *In Proceedings of the 7th International Conference on Computer Vision* (ICCV), pp. 722-729, Corfu, Greece, September 1999

Wulf, O. & Wagner, B. (2003). Fast 3d-scanning methods for laser measurement systems, *In Proceedings of International Conference on Control Systems and Computer Science* (CSCS14), Bucharest, Romania, February 2003

Vision Systems: Applications

Edited by Goro Obinata and Ashish Dutta

Computer Vision is the most important key in developing autonomous navigation systems for interaction with the environment. It also leads us to marvel at the functioning of our own vision system. In this book we have collected the latest applications of vision research from around the world. It contains both the conventional research areas like mobile robot navigation and map building, and more recent applications such as, micro vision, etc.The fist seven chapters contain the newer applications of vision like micro vision, grasping using vision, behavior based perception, inspection of railways and humanitarian demining. The later chapters deal with applications of vision in mobile robot navigation, camera calibration, object detection in vision search, map building, etc.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Stefan May, Kai Pervoelz and Hartmut Surmann (2007). 3D Cameras: 3D Computer Vision of Wide Scope, Vision Systems: Applications, Goro Obinata and Ashish Dutta (Ed.), ISBN: 978-3-902613-01-1, InTech, Available from:
http://www.intechopen.com/books/vision_systems_applications/3d_cameras__3d_computer_vision_of_wide_scope

INTECH
open science | open minds