An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger™)

Thierry Oggier*, Michael Lehmann, Rolf Kaufmann, Matthias Schweizer, Michael Richter, Peter Metzler, Graham Lang, Felix Lustenberger and Nicolas Blanc

CSEM SA, Badenerstrasse 569, CH - 8048 Zürich

ABSTRACT

A new miniaturized camera system that is capable of 3-dimensional imaging in real-time is presented. The compact imaging device is able to entirely capture its environment in all three spatial dimensions. It reliably and simultaneously delivers intensity data as well as range information on the objects and persons in the scene. The depth measurement is based on the time-of-flight (TOF) principle. A custom solid-state image sensor allows the parallel measurement of the phase, offset and amplitude of a radio frequency (RF) modulated light field that is emitted by the system and reflected back by the camera surroundings without requiring any mechanical scanning parts. In this paper, the theoretical background of the implemented TOF principle is presented, together with the technological requirements and detailed practical implementation issues of such a distance measuring system. Furthermore, the schematic overview of the complete 3D-camera system is provided. The experimental test results are presented and discussed. The present camera system can achieve sub-centimeter depth resolution for a wide range of operating conditions. A miniaturized version of such a 3D-solid-state camera, the SwissRanger™ 2, is presented as an example, illustrating the possibility of manufacturing compact, robust and cost effective ranging camera products for 3D imaging in real-time.

Keywords: Time-of-flight camera, TOF, 3D, ranging, CMOS/CCD, demodulation, lock-in pixel, SwissRanger

1. INTRODUCTION

The capability of human beings to perceive their environment in three dimensions is a key aspect of their visual system. However, this represents a major challenge to film-based and electronic camera systems. Since we are living in a three-dimensional world, it seems obvious that, generally speaking, sensors have to be able to capture the world in three dimensions as well. The technological requirements for such 3D systems are very stringent and obtaining reliable real-time distance information over an entire scene is still an unsolved problem to a large extent. In particular, there is a clear need for robust and cost effective solutions. So far, mainly sensors measuring distances in one dimension have found their way into the market. For a long time, capturing complete scenes remained limited to two-dimensional sensors, providing only monochrome or color images without depth information. The latest technological progress now promises for the first time to lead to sensors that can “see distances” within complete scenes. Numerous applications will directly benefit from such a 3D-technology.

In this paper, a brief overview of the theoretical background for the underlying Time-of-Flight (TOF) principle is first given. The technological requirements of the system performing the distance measurement and the corresponding algorithms as well as an example of a practical implementation are presented. The theoretical limits of such a system set by the photon shot noise are derived. Test results, in particular with regards to sensitivity and distance resolution, are discussed and, for the first time, the novel miniaturized 3D-camera system SwissRanger™ 2 with a USB2.0 interface is presented. Furthermore, the distance accuracy achieved is compared to the theoretical predictions. Finally, an outlook for three-dimensional sensing in real-time is provided.
2. “SEEING DISTANCES”

2.1 Theoretical background

The principle of the distance measurement considered relies on the physical propagation properties of light. The complete 3D camera system is based on two main components, namely a light emitter and a light detector. The emitted optical signal is modulated in amplitude. This signal is reflected by the objects within the illuminated scene and travels back to the camera system, where its precise time of arrival is measured. In the case of continuous-wave modulation, the phase delay between the emitted and the detected signal is recorded. This phase is measured by each individual pixel within an area array image sensor, meaning that each pixel is able to demodulate the incoming RF-modulated light field. Hence, complete phase maps and, finally, distance maps can be acquired with a lateral resolution defined by the number of pixels in the focal plane array.

The signal phase is detected by synchronously demodulating the incoming modulated light within the detector. This demodulation can be performed by the correlation with the original modulation signal, leading to the so-called cross correlation. The cross correlation between the demodulation signal \( g(t) \) and the incoming optical signal \( s(t) \) is computed according to (1).

\[
c(t) = s(t) \otimes g(t) = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{+T/2} s(t) \cdot g(t + \tau) \, dt
\]

The experimental results presented in this paper are all based on a sinusoidal modulation of the emitted optical signal with a frequency of 20 MHz.

The incoming optical signal onto the sensor is thus composed of an offset signal, partially due to, e.g. background illumination, and a sinusoidally modulated signal due to the reflected optical wave front from the objects and persons in the scene. The demodulation function can be assumed to be sinusoidal as well. The incoming optical signal and the demodulation function can thus be expressed mathematically as given in (2).

\[
s(t) = 1 + a \cdot \cos(\omega \cdot t - \phi) \\
g(t) = \cos(\omega \cdot t)
\]

Evaluating equation (1) for selected phase delays of \( \tau_0 = 0^\circ, \tau_1 = 90^\circ, \tau_2 = 180^\circ \) and \( \tau_3 = 270^\circ \), the phase \( \phi \), the offset \( B \) and the amplitude \( A \) of the incoming signal are computed according to equations (3) to (5) [1], [2]

\[
\phi = \arctan \left( \frac{c(\tau_1) - c(\tau_3)}{c(\tau_0) - c(\tau_2)} \right)
\]

\[
B = \frac{c(\tau_0) + c(\tau_1) + c(\tau_2) + c(\tau_3)}{4}
\]

\[
A = \sqrt{\left[ C(\tau_3) - C(\tau_1) \right]^2 + \left[ C(\tau_0) - C(\tau_2) \right]^2} / 2
\]

The parameters \( A, B \) and \( \phi \) are graphically represented in Figure 1.
The phase delay $\phi$ represents the propagation delay and is directly proportional to the target distance. The offset $B$ can be used to provide a conventional 2D intensity image. At the same time, the offset indicates if saturation occurs within the solid-state image sensor. The amplitude $A$ can be interpreted as a direct measure of the depth resolution achieved.

The target distance $D$ to the camera can be computed according to equation (6).

$$D = L \cdot \frac{\phi}{2 \cdot \pi}, \text{ with}$$

$$L = \frac{c}{2 \cdot f_m},$$

where $L$ represents the non-ambiguity distance range, $c$ the speed of light and $f_m$ the RF modulation frequency. This algorithm is commonly known as the four buckets, or four steps method [2].

### 2.2 Power预算 and photon shot noise limited range accuracy

This subsection illustrates the ultimate theoretical distance accuracy limits set by photon shot noise. In our experimental setup, the target is assumed to have the properties of a Lambert reflector as described in Figure 2.

Figure 1: Input signal with amplitude $A$, offset $B$ and phase $\phi$

Figure 2: Lambert reflector.
First, let us consider the relation between the optical light intensity in the image plane $P_{\text{image}}$, i.e., the optical power per unit area on the detector and the optical light intensity in the object plane $P_{\text{object}}$, i.e., the optical power per unit area on the target. Assuming no losses in the optical path through the air and optics, $P_{\text{image}}$ and $P_{\text{object}}$ are simply related by the F-number F/# of the imaging optics and thus can be computed according to equation 8.

$$P_{\text{image}} = P_{\text{object}} \left( \frac{1}{2 \cdot F/\#} \right)^2$$

Using the rules of error propagation and considering the algorithmic properties of the four buckets algorithm with a sampling duration of half of the modulation period, the photon shot noise limit for the depth measurement accuracy is given by (9). A detailed derivation of equations (8) and (9) can be found in [7].

$$\Delta L = \frac{L}{\sqrt{8}} \cdot \frac{\sqrt{B}}{2 \cdot A}$$

where $B$ corresponds to the offset or the average number of photo-detected electrons and $A$ corresponds to the amplitude of the demodulated incoming optical signal. As illustrated in Figure 1, $B$ consists of two components: it is related to the background illumination as well as to the active RF-modulated illumination. On the other hand, $A$ depends on the emitter side on the total emitted optical power and the modulation depth, on the imaging optics and filters, in particular F/# and spectral transmission properties, on the demodulation performance of the pixel, i.e. parameters such as the quantum efficiency and the fill factor (see also paragraph 3.4) as well as on the target characteristics such as distance and reflectivity.

Additional noise sources such as 1/f-noise, reset noise or thermal noise, can be integrated into (9) by adding a number of “pseudo-electrons” to the offset $B$ as shown in equation (10).

$$\Delta L = \frac{L}{\sqrt{8}} \cdot \frac{\sqrt{B + N_{\text{pseudo}}}}{2 \cdot A}$$

The range camera thus provides simultaneously for each single measurement and pixel within the 3D sensor array distance maps as given by equations (3) and (6), intensity maps as shown in equation (4) and, finally, the relative distance accuracy that has been achieved according to equation (10).

## 3. TECHNOLOGICAL REQUIREMENTS FOR THE 3D-CAMERA SYSTEM

This section describes the requirements for the main components that form the complete 3D-camera system in order to ensure the overall functionality and performance of this 3D imaging device. The performance of the 3D-camera system indeed depends critically on the performance of each of its components, namely the emitter, the optics, the sensor and the control electronics. Actually, the complete system needs to be optimized as a whole, meaning that the various components and their interplay must be optimized jointly.

### 3.1 Emitter

The emitter performance is mostly determined by its modulation frequency, the modulation contrast and the total emitted power. According to equations (8) and (9), a higher modulation frequency corresponds to a shorter non-ambiguity range and a higher depth resolution. In most applications the total emitted power is limited by eye safety considerations and hence cannot be increased arbitrarily. The wavelength of the optical emitter has to match the spectral sensitivity of the detector. Silicon being the base material for the present 3D imager, the wavelength is preferably in the visible or in the near-infrared range of the electromagnetic spectrum since, for most applications, invisible light is desired. The spectral bandwidth of the emitter as well as its temperature drift has to be minimized. These two parameters play a key role in applications where a strong background illumination, e.g. as found in outdoors applications, or an extended operating-temperature range are expected.
3.2 Optics

The optics must of course cover the desired field-of-view and match the specifications in terms of spatial resolution, shading and distortion. Generally, optics with a large aperture and thus a small F/# number are preferred, so as to maximize the light that finally reaches the image sensor. In order to reduce the impact of background illumination, a narrow band-pass filter centered on the peak wavelength of the emitter is typically used. For a compact and cost-effective solution, a miniaturized imaging optics is preferred. In the present case of a prototype series, commercially available elements were used for the optics, the optical mount and for the filters, whereas for large volume applications dedicated injection-molded optics that include also the optical mount and the band-pass filters will be designed and manufactured.

3.3 Electronics

The main control and readout electronics are implemented on a dedicated printed circuit board (PCB), insofar as they are not yet integrated on the sensor chip. The electronics presented in this paper have not yet been designed and optimized for one specific application. The current 3D solid-state optical range camera is meant as a general-purpose demonstrator and can be used to get first hands-on experience with this novel 3D imaging technology under real operating conditions. As such there is still much room for further improvement in terms of, e.g. performance and costs. The main task of the PCB consists in controlling and reading out the 3D sensor as well as driving the illumination unit. The PCB converts the outputs of the analog sensor to digital signals and performs all computations for the phase, offset and amplitude values as given in equations (3), (4), (5) and (6). Additionally, the PCB performs the data transmission to a personal computer using a standard USB2.0 serial interface.

3.4 Sensor

The sensor is one of the most critical components in the whole system. Basically it determines the performance of the receiver and thus, to a large extent, the overall system performance. The sensor has essentially to accomplish the four tasks schematically sketched in Figure 3; see [1].

![Figure 3: Main functions of the sensor](image)
First, the sensor transforms the incoming optical signal into electron-hole pairs. The performance of this task is essentially given by the inherent quantum efficiency of the silicon process and the fill factor of the sensor. In order to demodulate the incoming 20 MHz signal, a fast charge separation and transport has to take place within each pixel. The sensor’s ability to separate and transfer the charges to the corresponding output node can be expressed as a demodulation contrast, which is defined as the ratio of the demodulated amplitude to the acquired offset signal according to equation (11), with the amplitude and the offset defined in equations (4) and (5).

\[
C_{\text{demodulation}} = \frac{A}{B}
\]  

(11)

It is worthwhile remembering that, within one single modulation period of 50 ns corresponding to the modulation frequency of 20 MHz, typically only a few photons impinge on each individual pixel and hence only a few photoelectrons are generated per pixel. For a broad range of operating conditions, even less than 1 electron is created per modulation period. The repeated addition of the electrons generated over numerous modulation periods is thus necessary and represents one very important feature of the current embodiment. The approach of adding charge at the pixel level almost noise-free is tightly linked to the CCD technology/principle that indeed enables the efficient transfer and addition of electronic charges and hence is a key element to the success of the present technique. Moreover, the in-pixel storage and processing of the different signal samples enables a high degree of flexibility in the readout process. The overall sensor architecture is similar to a CMOS active pixels sensor (APS) enabling, for example, the definition and readout of regions-of-interest (ROI) that can be set rather arbitrarily.

4. PRACTICAL IMPLEMENTATION OF SWISSRANGER™ 2 (SR-2)

The complete 3D-camera system is composed of the sensor board including the sensor itself and its control/readout electronics, the illumination board with drivers and light emitting diodes (LEDs), the aluminum case and the imaging optics with a band-pass filter. Other than the 3D sensor, only commercially available electronic and optical components have been used to manufacture the SR-2 camera demonstrator. The objective for the present 3D-camera implementation and, in particular, its electronics was not only to provide a reliable operation of the sensor but also to keep ample freedom in the demonstration and testing of several new ideas and functions implemented in the sensor. This means that some options and features were incorporated in the electronics mainly as support during the test phase, hence options that later on might not be necessary at all for a specific application.

4.1 Illumination, Optics, Electronics

In order to keep the current system costs down, the illumination board has so far been based on commercially available LEDs. An optical band-pass filter is placed in front of the sensor in order to reduce the impact of background illumination. Other illumination sources with a smaller spectral bandwidth such as lasers, vertical cavity surface emitting lasers (VCSELs), or resonant cavity LEDs (RCLEDs) would permit the use of narrower band-pass filters and thus a further decrease in the background illumination, finally improving the overall system performance.

The electronic sensor board is the master of the 3D-camera system. To ensure that the camera meets the specifications, a multi-layer PCB is used in conjunction with fine pitch connections and ball-grid array components. The PCB size could thus be kept to a minimal size with all necessary hardware fitted on a surface of 27.5 x 130 mm². A field programmable gate array (FPGA XC2S200) controls the overall board and timing, including the control and readout of the sensor, the digital-to-analog converter (DAC), the analog-to-digital converter (ADC), the memory, as well as the USB 2.0 interface chip. All calculations for the distance maps are implemented within the FPGA. The depth accuracy achieved is also computed for every single pixel in the sensor array in real-time on-board. An 8-bit AD5308 ADC is used to deliver the required voltage levels for the sensor. As for the 12-bit A/D conversion of the analog sensor signals, it is accomplished using a commercial AD9238 ADC. A schematic overview of the sensor board is represented in Figure 4.
4.2 3D Sensor

The custom designed 3D sensor represents the core element of the 3D-camera. The sensor is manufactured in the 0.8 μm CMOS/CCD technology of ZMD (Zentrum Mikroelektronik, Dresden) [15]. The specific process is based on a CMOS process that offers, in addition, a CCD option [3]. This unique combination permits the optimal use of the strengths of both CMOS and CCD technologies in terms of sensor performance while at the same time keeping the flexibility of the co-integration of electronic control and readout circuitry. The pixel architecture is based on the so-called “lock-in pixel” principle described in [4], [5] and [6], and an APS-readout structure [8]. This offers the possibility to address every single pixel in the imager individually, allowing the definition of specific ROIs. A key factor determining the performance of the system is related to the efficient and fast separation and transport of the photo-generated charges within the pixels. Analytical expressions for charge transport mechanisms in CCD structures can be found in [9], [10], [11], [12] and [13].

The various architectures and implementations of the 3D-sensors, as developed at CSEM, differ not only in terms of spatial resolution and functionality but also down to the very specific layout of the pixels. In the simplest implementation, every pixel provides one single detection site and one single storage element, the optically sensitive element here being a MOS capacitor, as in most CCD imagers as well as photo-gate CMOS imagers. Within each pixel, these so-called 1-tap sensors can thus only store one sample out of the four samples that are required to compute the phase, offset and amplitude of the incoming signal, as described by equations (3) to (5). Therefore, the four samples have to be acquired sequentially and hence create difficulties in imaging moving targets. More recently, 3D-sensors based on 2-tap as well as 4-tap structures have also been designed, manufactured and tested. In the latter case all four sampling events can be performed directly in parallel, leading to improved results in terms of sensitivity/speed of the system. The sensor implemented in the SwissRanger™ 2 camera is based on the 2-tap pixel architecture.

While, for a 1-tap sensor, half of the optical signal is discarded immediately, since the integration is limited to half of the modulation period, 2-tap pixels allow the detection of all incoming light. The present 3D sensor has a lateral resolution of 160 x 124 pixels with a fill factor of 17.3 %. In order to improve the sensor performance in terms of quantum efficiency in the NIR range, speed and demodulation contrast, it is designed with a buried-channel CCD structure. The buried channel indeed increases the ability to separate and detect the photo-generated electrons, while improving the
performance of the sensor in terms of noise and sensitivity [10]. The buried channel implant as well as the substrate doping levels were determined and optimized by extensive simulations utilizing the ISE-TCAD tools [14].

Figure 5 illustrates a simulation result of a CCD-gate structure as implemented in the SwissRanger™ 3D sensor. In this specific example, the electrostatic potential in the silicon is displayed for selected voltages applied to the CCD gates. The potential distribution within the silicon is shown using a grayscale representation.

5. TEST RESULTS

The most important parameter that finally determines the quality and performance of the 3D sensor can be summarized by the question: what is the minimum incoming optical energy on a single pixel that needs to be detected to achieve a given distance resolution? The number of photo-generated electrons accumulated in a single pixel during one exposure period is directly proportional to the optical energy impinging on one pixel. The relation between the optical energy on a pixel and the photo-generated electrons is given by (12).

\[ E_{opt} = \frac{N_{el}}{QE \cdot FF} \cdot \frac{hc}{\lambda}, \]  

with \( N_{el} \) the number of photo-generated electrons, \( QE \) the quantum efficiency, \( FF \) the fill factor, \( c \) the speed of light, \( h \) the Planck constant and \( \lambda \) the wavelength of the emitted signal.

A series of measurements was performed on a general-purpose test bench with an exposure time of 1 ms for different illumination conditions. The test bench is not yet fully optimized and improvements are expected once the tests can be performed and repeated on a dedicated electronics board. The emitter consists of an illumination source with a peak wavelength at 810 nm. The images are acquired on the test bench and then further processed on a PC. The distance resolution is determined by the standard deviation computed from a set of more than 100 distance measurements for different optical energies impinging on the pixel surface. The measured standard deviation as a function of the impinging optical energy on the pixel is plotted in figure 6. The ultimate theoretical distance resolution limit, as given by the photon
shot noise, was determined accordingly, using the acquired amplitude $A$ and offset $B$ and inserting them into equation (9). For the selected modulation frequency of 20 MHz, the non-ambiguity range $L$ equals 7.5 m.

Figure 6 reveals that the standard deviation of the measured distances as a function of the incoming optical energy on the pixel area approaches the theoretical photon shot noise limit.

Figure 6: Distance resolution (standard deviation) as a function of the optical energy falling on a pixel compared to the theoretical limit given by the photon shot noise.

A distance resolution in the sub-centimeter range can clearly be achieved on a broad range of illumination levels. Please note that the measurements were performed without background illumination. In this test series, a demodulation contrast, such as defined by (11), of 40 % has been achieved. The full well capacity, i.e. the maximum number of electrons that can be accumulated before saturation sets in, is larger than 1’000’000 electrons.

6. 3D CAMERA DEMONSTRATOR

The novel miniature SwissRanger™ SR-2 camera demonstrator allows optical range imaging of complete scenes in real-time. An example of a raw depth image is given in Figure 7. In Figure 7a, the conventional gray level image of the measured scene acquired by the SR-2 is depicted. Figure 7b represents the distance map, coded in black and white. Objects close to the camera are coded darker than those far away. Finally, Figure 7c combines the distance information and the gray level information, resulting in a three-dimensional black and white picture.
The complete 3D camera SR-2 includes the illumination unit, the sensor board with its control electronics and a USB2.0 interface. The latter allows the 3D camera to be easily interfaced to any recent laptop or desktop computer. A simple graphical user interface enables the read out of the 3D image data and the configuration of the camera settings. The complete 3D-camera system is represented in Figure 8.
7. OUTLOOK

The optical range imaging system presented in this work offers the possibility of entering completely new application fields. This technique is inherently non-contact and presents good resolution in all three spatial directions at frame rates appropriate to many real-time applications. It is expected that, in numerous application fields, the added value of the depth information will allow the replacement of traditional sensors and cameras by novel camera systems providing complete 3D information. Optimized sensors with increased robustness against high background illumination and even larger spatial resolutions (e.g. VGA resolution) will clearly ensure an even larger commercial impact of such a 3D technology. The continuous progress in the field of microelectronics, micro optics and micro technology will definitely enable the performance as well as the degree of integration of such 3D-camera systems to be further extended. Advances in semiconductor materials and processing, e.g. in CMOS/CCD technologies as well as components such as LEDs and semiconductor lasers that are crucial to this technology, notably for the emitter and the detector, will directly lead to improved 3D-camera systems. Appropriate algorithms and image processing techniques that fully take advantage of this additional depth information still need to be invented and it is expected to be a very vivid field of research and development in the near future.

8. CONCLUSION

A novel and miniaturized 3D-camera system for real-time range imaging has been presented. The system is based on the time-of-flight principle with continuous wave modulation and a custom smart CMOS/CCD sensor with 124x160 so-called “lock-in pixels”. Each individual pixel in the sensor is able to perform a demodulation of the incoming signal and hence a phase or distance measurement. All other components are commercially available electronics and optical components, permitting the manufacture of new compact and cost effective 3D TOF cameras.

In this paper, the measurement principle, as well as analytical expressions for the distance resolution, have been presented. The distance resolution depends, to a first order, on the optical energy that is reflected back onto the detector, implying that many parameters such as the total emitted optical power, the field of view, the distance range, the object reflectivity and the optical aperture play a significant role. The optimization of the complete system performance requires that every single component be jointly addressed, to ensure their optimal interplay. The requirements of the main system components, including the dedicated 3D image sensor, have been accordingly examined and discussed in more detail. A practical implementation of a 3D TOF-camera demonstrator has been presented, including its various key components and electronics. The 3D sensor, manufactured in a 0.8 \( \mu \)m CMOS/CCD technology, exhibits good performance, being close to the photon shot noise limit. Distance resolution in the sub-centimeter range has been achieved under realistic operating conditions.

ACKNOWLEDGMENT

The authors would like to thank all the members of the Image Sensing Group of CSEM SA for their continuous support in the various phases of this research and development project.

REFERENCES


*thierry.oggier@csem.ch; phone +41 1 497 14 15, fax +41 1 497 14 00, http://www.csem.ch/imaging