Solid-State Time-of-Flight Range Camera

Robert Lange and Peter Seitz, *Member IEEE*

***Abstract*—The concept of a real-time range camera without moving parts is described, based on the time-of-flight (TOF) principle. It operates with modulated visible and near-infrared radiation, which is detected and demodulated simultaneously by a 2-D array of lock-in pixels employing the charge-coupled device principle. Each pixel individually measures the amplitude, offset and phase of the received radiation. The theoretical resolution limit of this TOF range camera is derived, which depends on the square root of the detected background radiation and the inverse of the modulation amplitude. Actual measurements of 3-D sequences acquired at 10 range images per second show excellent agreement between our theory and the observed results. A range resolution of a few centimeters over a range of 10 m, with an illumination power of a few hundreds of milliwatts is obtained in laboratory scenes for noncooperative, diffusely reflecting objects.**

***Index Terms*—Image sensor, time-of-flight, 3-D range camera.**

I. Introduction

SINCE we are living in a 3-D world, an adequate description of our environment for many applications includes the relative position and motion of the different objects in a scene. Nature has satisfied this need for spatial perception by providing most animals with at least two eyes. This stereo vision ability is the basis that allows the brain to calculate qualitative depth information of the observed scene [1].

Imaging 3-D measurement with useful distance resolution has mainly been realized so far with triangulation systems, either passive triangulation (stereo vision) [2] or active triangulation (e.g., projected fringe methods) [3]. These triangulation systems have to deal with shadowing effects and ambiguity problems (projected fringe), which often restrict the range of application areas. Moreover, stereo vision cannot be used to measure a contrastless scene. This is because the basic principle of stereo vision is the extraction of characteristic contrast-related features within the observed scene and the comparison of their position within the two images. Also, extracting the 3-D information from the measured data requires an enormous time-consuming computational effort. High depth resolution can only be achieved with a relatively large triangulation base and, hence, large camera systems.

A much more practical range measurement method is the time-of-flight (TOF) principle, an optical analogy to a bat's ultrasonic system rather than a human's stereo vision. So far, TOF systems are only available as 1-D systems (point measurement) [4], requiring laser scanners to acquire 3-D images [5]. Such TOF scanners are expensive, bulky, slow, vibration sensitive and, therefore, only suited for restricted application fields.

In this work, an imaging, i.e., nonscanning TOF-camera, is introduced, based on an array of demodulation pixels, where each pixel can measure both the background intensity and the individual arrival time of an RF-modulated (20 MHz) scene illumination with an accuracy of a few hundreds of picoseconds. The pixel's working concept is based on the charge-coupled device (CCD) principle [6]. We call these pixels *demodulation pixels* because they extract the target's distance and reflectivity from the received optical signal [7].

In Section II, we introduce the TOF measurement principle. We show that the accuracy of every TOF ranging system depends on the ratio of the optical signal power to the power of the background radiation. An estimation of the resolution limit is carried out in Section III, mainly considering the quantum noise as a final theoretical limitation. With the deduced equation, it is possible to predict a TOF camera's range resolution as a function of the received optical power.

The key element of the new 3-D-range camera, the 1-tap demodulation pixel, is introduced in Section IV. Characterizing measurements show the excellent performance of the pixel and the influence of the modulation frequency, wavelength, and optical power of the modulated light source on both measured contrast and time accuracy. An experimental range camera with 64×25 pixels is introduced in Section V, where also a typical range measurement is presented. With an overview in Section VI, we show which technology parameters would be required to realize our demodulation pixels in a standard CMOS rather than a special CMOS/CCD technology.

With this work, we demonstrate the first successful realization of an all-solid-state 3-D TOF range-camera without moving parts that is based on a dedicated customized PhotoASIC. The observed range measurement performance is very close to the theoretical limits.

II. TOF Measurment

We can measure an absolute distance if we manage to measure the absolute time that a light pulse needs to travel from a target to the detector. This indirect distance measurement is possible since the speed of light in air for different environmental conditions is known very precisely. In practice, the active light source and the receiver are located very closely to each other. This facilitates a compact setup and avoids shadowing effects.

Instead of light pulses, we use continuous-wave (CW) modulation in this work, where the phase difference between sent and received signals is measured, rather than directly measuring a light pulse's total trip time. As the modulation frequency is known, this measured phase directly corresponds to the time of flight, the quantity of interest.

Manuscript received August 28, 2000; revised November 16, 2000.

R. Lange was with the Industriestrasse 11, CH-8155 Niederhasli, Switzerland. He is now with GretagMacbeth AG, CH-8105 Regensdorf, Switzerland. (e-mail: lange@freesurf.ch).

P. Seitz is with the Centre Suisse d'Electronique et de Microtechnique SA, CSEM Zurich, CH-8048 Zurich, Switzerland (e-mail: peter.seitz@csem.ch).

Publisher Item Identifier S 0018-9197(01)01630-X.

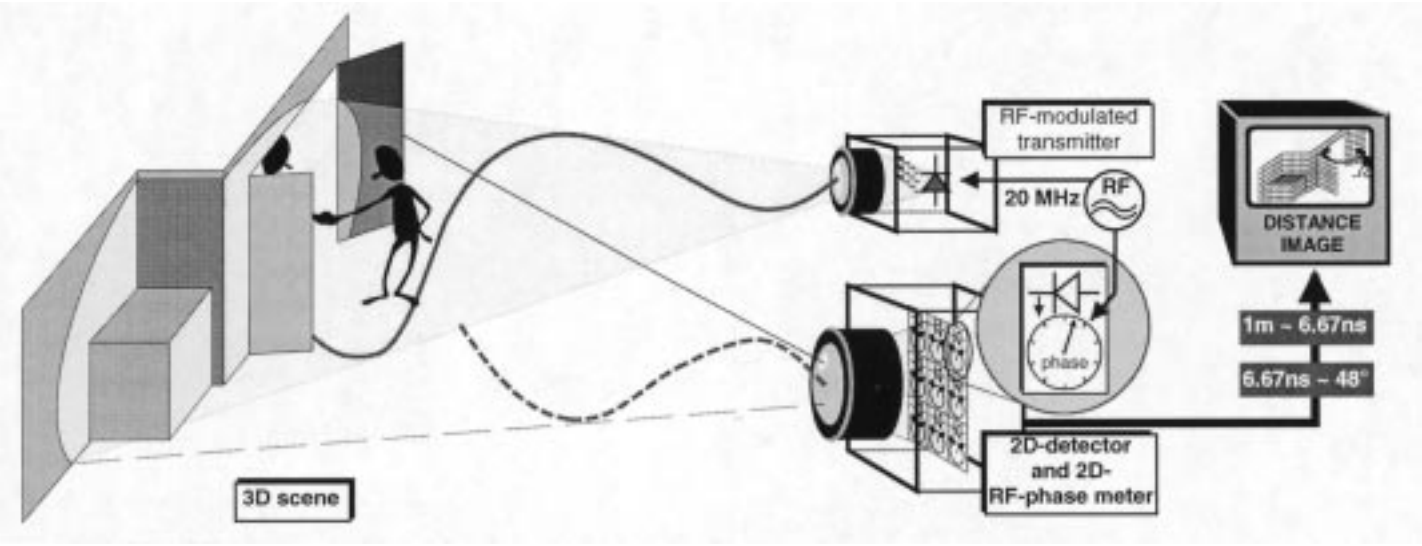


Fig. 1. Principle of a nonscanning 3-D TOF-camera.

Conventional TOF systems use a point detector and scan a modulated laser beam over the scene in order to acquire a 3-D image [5]. In contrast, the camera introduced here illuminates the entire scene simultaneously with a modulated light cone as illustrated in Fig. 1. This necessitates a 2-D-electrooptical demodulator and detector in the receiver path to measure the distances of some hundreds or thousands of points of the observed scene in parallel.

A. Demodulation and Sampling

We can retrieve the signal's amplitude and phase by synchronously demodulating the incoming modulated light within the detector. Demodulation of a received modulated signal can be performed by correlation with the original modulation signal (cross correlation). The measurement of the cross correlation function at selectively chosen temporal positions (phases) allows the phase of the investigated periodical signal to be determined. With the optical input signal $s(t)$ of modulation amplitude a and phase φ, and the demodulation or correlation signal $g(t)$ defined as follows:

$$s(t) = 1 + a \cdot \cos(\varpi t - \varphi) \quad \text{and} \quad g(t) = \cos(\varpi t). \tag{1}$$

The correlation function $c(\tau)$ can be calculated as

$$c(\tau) = s(t) \otimes g(t) = \frac{a}{2} \cdot \cos(\varphi + \tau). \tag{2}$$

We evaluate this function for different phases τ. Choosing $\tau_0 = 0°, \tau_1 = 90°, \tau_2 = 180°$, and $\tau_3 = 270°$ allows us to recalculate the phase φ and amplitude of the received optical signal $s(t)$. Considering that the received signal is often superimposed onto a background image, we must add an offset K to obtain the practically measured values

$$C(\tau_0) = c(\tau_0) + K = \frac{a}{2} \cdot \cos(\varphi) + K$$
$$C(\tau_1) = c(\tau_1) + K = -\frac{a}{2} \cdot \sin(\varphi) + K$$

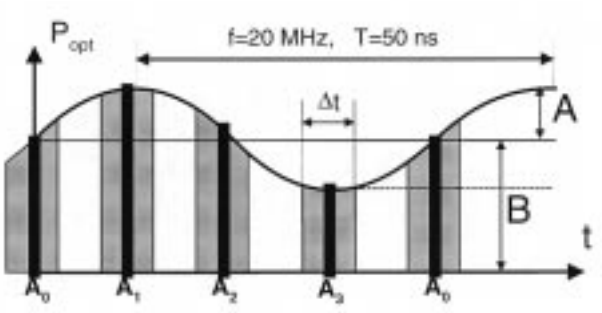


Fig. 2. Optical sinusoidally modulated input signal, sampled with four sampling points per modulation period. The signal frequency of 20 MHz defines the unambiguous distance range of 7.5 m.

$$C(\tau_2) = c(\tau_2) + K = -\frac{a}{2} \cdot \cos(\varphi) + K$$
$$C(\tau_3) = c(\tau_3) + K = \frac{a}{2} \cdot \sin(\varphi) + K. \tag{3}$$

Hence, we can determine the phase φ and amplitude a of $s(t)$

$$\varphi = a \tan\left(\frac{C(\tau_3) - C(\tau_1)}{C(\tau_0) - C(\tau_2)}\right) \tag{4}$$
$$a = \frac{\sqrt{[C(\tau_3) - C(\tau_1)]^2 + [C(\tau_0) - C(\tau_2)]^2}}{2}. \tag{5}$$

Another, slightly different approach is to sample the modulated signal synchronously. Since we sample a periodic signal, we can use the equations of the discrete Fourier transform (DFT) in order to calculate both the amplitude and phase of the base frequency and harmonics contained in the signal.

We use a sinusoidal wave (only base frequency, no harmonics) as the modulation signal and synchronously sample this wave with four equally spaced sampling points of duration Δt, as illustrated in Fig. 2. We can directly deduce the equations for phase φ, amplitude A and offset B from the definition of the DFT

$$\varphi = a \tan \frac{A_3 - A_1}{A_0 - A_2} \tag{6}$$

$$A = \frac{\delta}{\Delta t \cdot \sin\delta} \cdot \frac{\sqrt{(A_3 - A_1)^2 + (A_0 - A_2)^2}}{2} \quad (7)$$

$$B = \frac{A_0 + A_1 + A_2 + A_3}{4 \cdot \Delta t}. \quad (8)$$

Using the DFT, the finite number N of sampling points only allows the determination of a finite number of $(N/2) - 1$ discrete frequency components. For the illustrated case $(N = 4)$, the system is only sensitive to one discrete frequency. Since this frequency selectivity is a well-known property of lock-in amplifiers, we also call the demodulation pixels "lock-in pixels" [8].

We see that we obtain similar evaluation functions as with the correlation-approach. This comes as no surprise because the Fourier transform is defined as the correlation with harmonic base functions, as selected in (1).

Due to the fact that no ideal sampling can be performed, $A_0 \ldots A_3$ represent the integrated sampling values, i.e., they are proportional to the number of detected photons. For ideal sampling, the system would need an infinite bandwidth. The integrative nature of this practical sampling process can be interpreted as a convolution of the ideally sampled input signal with the sampling function, a $\mathrm{rect}(t/\Delta t)$ function, where Δt is the integration time.

Natural sampling has no influence on the measured phase as long as the integration time of each sampling point is shorter than the modulation period of the sampled signal, which is a reasonable claim. Only the measured amplitude is attenuated by a factor of $\delta/(\Delta t \cdot \sin\delta)$, depending on the integration interval Δt, with $\delta = \pi \cdot \Delta t/T$, where T is the modulation period. This factor is already considered in (7). For $\Delta t = T/2$ the measured amplitude is 64% of the real amplitude.

III. Optical Power Budget and Resolution Limits

A. Power Budget

For many practical reasons, especially for eye safety considerations, it is important to know the requirements for the optical power of the light source. For this purpose, two things have to be known: 1) how much optical power of modulated scene illumination is necessary to generate a given number of photoelectrons in one pixel and 2) how many photoelectrons per sampling point and pixel are necessary to achieve a certain range resolution? Both questions will be investigated and answered in this section. The number of generated electrons is related to the number of photons through the quantum efficiency of the imager. With the energy $h \cdot c/\lambda$ of one photon, the integration time T_{int}, the size of the light sensitive pixel area A_{pix}, and the size of the image in the sensor plane A_{image}, one can calculate the total amount of optical power that must arrive at the sensor. Consideration of optical losses k_{lens} of the lens and filters leads to the power in front of the lens over the aperture area. If the observed object is a Lambertian reflector, one can calculate the optical power on the observed scene, which has to be a factor of $(D/2R)^2$ higher than the power in front of the lens (D is the aperture of the lens and R is the distance to target). One obtains the above factor by taking the ratio of the cosine-weighted (Lambertian reflection characteristics) integral over the spherical cap of diameter D, (i.e., the share of optical power entering the lens), to the cosine-weighted integral over the complete hemisphere. Thus, we can calculate the required optical power of the illumination source $P_{\mathrm{light\_source}}$ as a function of the number of generated electrons N_e per pixel and the camera and object parameters:

$$P_{\mathrm{light\ source}} = \frac{N_e \cdot \frac{A_{\mathrm{image}}}{A_{\mathrm{pix}}} \cdot h \cdot c}{\rho \cdot \left(\frac{D}{2 \cdot R}\right)^2 \cdot k_{\mathrm{lens}} \cdot \mathrm{QE}(\lambda) \cdot \lambda \cdot T_{\mathrm{int}}} \quad (9)$$

where

N_e	number of electrons per pixel;
A_{image}	image size in sensor plane;
A_{pix}	light-sensitive area of pixel;
h	Planck's constant;
c	speed of light in vacuum;
ρ	reflectivity of object;
D	aperture of lens;
R	distance of object;
k_{lens}	losses of objective and filters;
$\mathrm{QE}(\lambda)$	quantum efficiency;
λ	wavelength of light;
T_{int}	integration time.

B. Quantum Noise as Final Range Resolution Limitation

The essential noise sources of CCD sensors and photodiode arrays are electronic and optical shot noise, thermal noise, reset noise, $1/f$ noise, and quantization noise. All of these noise sources can be reduced or eliminated by different signal processing techniques or cooling, except photon shot noise. Therefore, we start by investigating the influence of shot noise on the ranging accuracy.

Shot noise describes the statistical Poisson-distributed nature of the arrival process of photons and the generation process of electron-hole pairs. The standard deviation of shot noise is equal to the square root of the number of photons (optical shot noise) or photogenerated charge carriers (electronic shot noise). In the following, the required number of photoelectrons per sampling point and pixel to achieve a given range resolution is derived.

Applying the rules of error propagation to (6) and considering that each of the integrated sampling points $A_0 \ldots A_3$ shows a standard deviation of $\Delta A_i = \sqrt{A_i}$, one obtains the quantum noise limited phase error $\Delta\varphi$

$$\Delta\varphi = \sqrt{\sum_{i=0}^{3} \left(\frac{\partial\varphi}{\partial A_i}\right)^2 \cdot A_i}. \quad (10)$$

Solving for special phase values $(0°, 45°, 90°, 135°, 180°, \ldots)$, and considering that all sampling points are composed of an offset B (background plus dc-component of active illumination) and a phase-dependent component proportional to $A \cdot \cos(\varphi_0)$, where A and B are given in *number of electrons*, we obtain the range resolution ΔL

$$\Delta L = \frac{L}{360°} \cdot \Delta\varphi = \frac{L}{\sqrt{8}} \cdot \frac{\sqrt{B}}{2 \cdot A} \quad (11)$$

where

- L non-ambiguity distance range ($L = c/2f_{\mathrm{mod}}, f_{\mathrm{mod}} =$ modulation frequency);
- A (de)modulation amplitude, i.e., number of photoelectrons per pixel and sampling points generated by the modulated light source. A depends on the modulation depth of the modulated signal and the demodulation contrast of the pixel but also on the optical power of the modulated light source and the target's distance and reflectivity;
- B offset or acquired optical mean value, i.e., number of photoelectrons per pixel and sampling points generated by incoming light of the scene's background and the mean value of the received modulated light [cf., Fig. 2].

This range accuracy (11), which can only be improved by averaging, is the absolute limit of a lock-in range sensor working with four sampling points. One can see from the equation that a large background brightness ($B - A \approx B$ for $A \ll B$) not only restricts the number of available quantization levels but also drastically increases the quantum noise of the system. Background illumination can be reduced by measuring in the dark or by using spectral filters that only transmit the spectrum of the modulated light. Since, generally, the "active" optical power density on the illuminated scene increases with decreasing distance to the object, the ranging accuracy also increases for smaller distances. This is an important fact for navigation applications, where a high accuracy is often only needed close to the target.

C. Influence of Thermal Noise Sources

The additional photodetector noise sources, $1/f$-, reset-, and thermal noise, which can be summarized as *dark noise*, can be included in (11) by adding an additional number of pseudo-background-electrons N_{pseudo} to B. They are not correlated to the modulation signal and thus contribute to B rather than A

$$\Delta L = \frac{L}{\sqrt{8}} \cdot \frac{\sqrt{B + N_{\mathrm{pseudo}}}}{2 \cdot A.} \tag{12}$$

One obtains this number of pseudo-electrons N_{pseudo} simply by squaring the noise-equivalent number of noise electrons

$$N_{\mathrm{pseudo}} = (N_{\mathrm{dark}})^2 = \left(\frac{V_{\mathrm{dark\ noise}} \cdot C_{\mathrm{conv}}}{q \cdot A_{\mathrm{sf}}} \right)^2. \tag{13}$$

As an example, if we assume a measured dark noise of $V_{\mathrm{dark\ noise}}$ =0.63 mV rms, containing all noise sources except the photoelectron shot noise. With a typical output amplification of A_{sf} =0.9 and a conversion capacitance of C_{conv} =40 fF, this corresponds to an equivalent number of 175 dark noise electrons, leading to N_{pseudo} =30,000.

The received optical signal of the modulated light source contributes to both offset and demodulation amplitude. Integrated over a certain integration time T_{int}, this signal directly adds a number of photoelectrons $\mathrm{PE}_{\mathrm{opt}}$ to the effective offset B_{eff}

$$B_{\mathrm{eff}} = N_{\mathrm{backgr.}} + N_{\mathrm{pseudo}} + \mathrm{PE}_{\mathrm{opt}}. \tag{14}$$

Also, the number of demodulation-photoelectrons A can be expressed as a function of the optical mean power or the total number of photoelectrons per pixel generated by the modulated light source $\mathrm{PE}_{\mathrm{opt}}$. Only the modulation contrast C_{mod}, a parameter of the modulated light source, (usually 100%), and the demodulation contrast C_{demod}, which depends on the pixel performance (usually less than 100%), have to be considered

$$A = C_{\mathrm{mod}} \cdot C_{\mathrm{demod}} \cdot \mathrm{PE}_{\mathrm{opt}}. \tag{15}$$

Thus, we can rewrite (12)

$$\Delta L = \frac{L}{\sqrt{8}} \cdot \frac{\sqrt{N_{\mathrm{backgr.}} + N_{\mathrm{pseudo}} + \mathrm{PE}_{\mathrm{opt}}}}{2 \cdot C_{\mathrm{mod}} \cdot C_{\mathrm{demod}} \cdot \mathrm{PE}_{\mathrm{opt}}}. \tag{16}$$

IV. 1-Tap Demodulation Pixel

A. Pixel Concept

Demodulation pixels for TOF ranging can be realized by giving them the ability to sample the incoming light in the time domain synchronously with its modulation frequency. In order to enlarge the SNR, the sampling process must be repeated a large number of times, integrating the small signal amounts of each sampling point to a larger signal value. This successive integration of the sampling points: 1) increases the system's sensitivity to discrete frequencies only, i.e., spectral selectivity; 2) lowers the (unwanted) sensitivity to neighboring frequency components; and 3) improves the SNR by increasing the signal strength. In order to realize such a pixel capable of sampling a modulated electrooptical wave, required are: 1) detection of light (conversion of optical signal in electrical signal); 2) a fast shutter mechanism; 3) repetitive noise-free signal addition; and 4) in-pixel signal storage. All of these requirements can be realized with the CCD principle. Better than any other electronic principle realized in integrated circuits, the CCD principle allows nearly noise-free addition and directed charge transport of electrical signals in the charge domain.

Our demodulation pixels have been fabricated in a 2.0-μm CMOS/CCD process, a slightly modified CMOS process which is available as an inexpensive prototyping service [multi project wafer (MPW)]. This process offers the freedom to implement CCD's with sufficiently good performance for our application, although the performance is inferior to dedicated CCD technologies.

Fig. 3 shows the layout and a cross-sectional view (including the potential distribution in the semiconductor) of the 1-tap pixel. The pixel has a size of 65 μm by 21 μm and an optical fill factor of more than 20%. By applying proper gate voltages to the photogates, the potential gradient in the semiconductor can be influenced. If the control voltages of the photogates are changed synchronously with the modulated incoming light, optically generated charge carriers can either be integrated under the integration gate (IG) (if they belong to the sampling interval) or are dumped to the dump diffusion (if they do not belong to the sampling interval). This process can be repeated until the integration gate has accumulated a sufficiently large signal. In contrast to the lock-in pixel introduced in [8], this

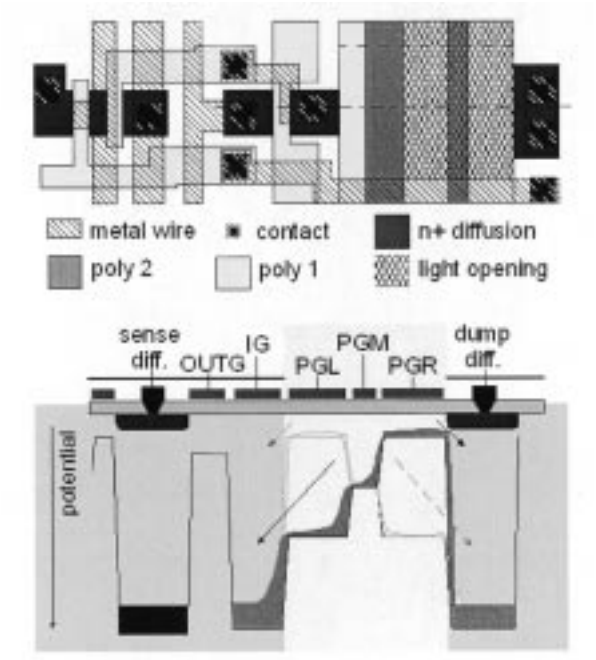


Fig. 3. One-tap lock-in pixel: pixel layout and cross sectional view of the CCD part. (IG: integration gate. PGL/PGM/PGR: left/middle/right photogate) Dimensions: photogate: $21 \times 14.5\ \mu\text{m}^2$, pixel: $21 \times 65\ \mu\text{m}^2$, aspect ratio: 1:3, fill factor: 22%.

1-tap pixel has only one storage site (IG). Therefore, the four sampling points have to be acquired sequentially rather than in parallel. Also, as can be seen in Fig. 3, each pixel has its own amplifier stage, which can be randomly accessed and reset over on-chip address decoders [active pixel concept (APS)]. With this APS readout, only a few CCD-transport steps are necessary for the acquisition and readout process, thus requiring only relatively poor charge transfer efficiency of the CCD structures. Since all sampling points of one pixel are acquired with the same structure and stored at the same storage site, this pixel has a highly uniform response for all sampling points.

B. Characterization of 1-Tap Pixel Performance

The demodulation contrast C_{demod} is a measure for the efficiency of the charge separation process in the CCD. In loose terms, C_{demod} describes the shutter efficiency. It should be as high as possible in order to achieve large signal amplitude and hence a good signal-to-noise ratio for a high phase accuracy [cf., (16)]. In a manner similar to the contrast definition in optics, C_{demod} can be defined as

$$C_{\text{demod}} = \frac{\text{measured amplitude}}{\text{measured offset}}. \tag{17}$$

A very simple model, introduced in Fig. 4, gives a first estimate of the demodulation contrast and numerical values we can expect. Even under ideal stationary conditions, we will never obtain a demodulation contrast of 100% with the current pixels. Assuming a planar potential underneath each CCD gate, 50% of the charge generated under the left photo gate cannot be separated and will be collected under the integration gate, independent of the photogates' bias conditions. This is because there

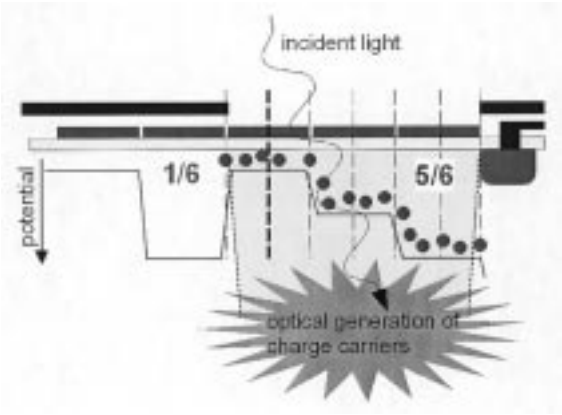


Fig. 4. Simple model of shutter inefficiency. For the dc case, we can expect a demodulation contrast of 67% (amplitude: $1/6 - 5/6$; offset: $1/6 + 5/6$).

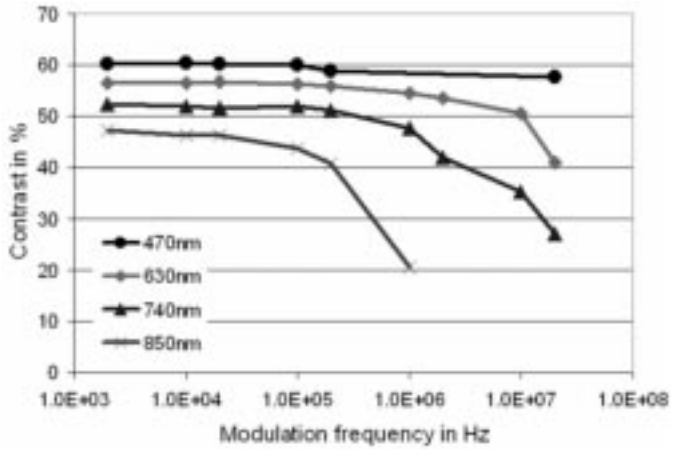


Fig. 5. Demodulation contrast versus modulation frequency and wavelength.

is no potential barrier between this photogate and the integration gate. The same is the case for the right photogate, where a fraction of charge will always travel directly into the dump diffusion and will never be detected. Hence, we can expect to measure a relative amplitude of $4/6$ $(5/6 - 1/6)$ and an offset of $6/6$ $(5/6 + 1/6)$, corresponding to a demodulation contrast of 67%. This means that, for this simple model, we cannot expect the demodulation contrast to be better than 67%, even for idealized dc conditions. So far, however, we have neglected two important parameters: 1) the demodulation frequency and 2) the influence of the light's wavelength.

1) Influence of the Demodulation Frequency: The higher the demodulation frequency (or sampling frequency), the shorter is the time available for the single charge carriers to travel from their site of optical generation to the storage area. If they do not reach the right storage site in time, they will contribute to a wrong sampling point and lower the demodulation amplitude. To optimize the pixel's electrical bandwidth, photogates of small length are required. The smaller the gate length, the more quickly the photoelectrons arrive at the storage sites and the higher becomes the influence of fringing fields.

2) Influence of the Wavelength: Since, generally, electrical fields within a CMOS-processed semiconductor exist only near the semiconductor surface, substrate-regions more than about 2–5 μm away from the surface are essentially field-free. Charge

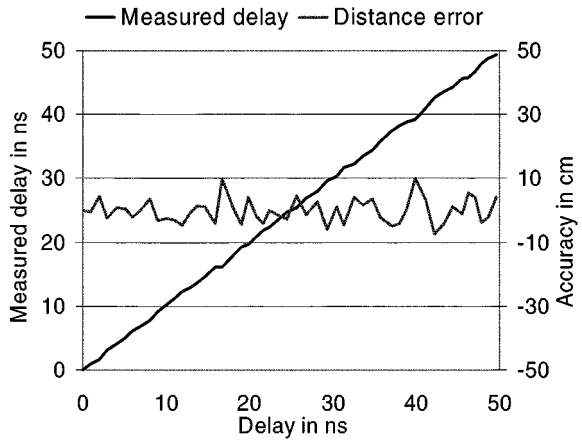


Fig. 6. Phase delay for 736-fW optical power on one pixel. Time error: 260-ps rms → distance accuracy: 3.84 cm rms.

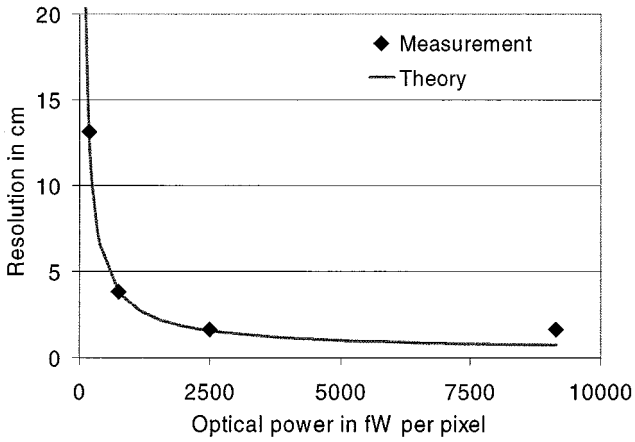


Fig. 7. Measured and estimated distance resolution versus received optical power per pixel. Integration time: 12.5 ms per sampling point, total 50 ms.

carriers that are optically generated in these regions move by thermal diffusion in random directions. Either they recombine within the substrate, or they are collected by an electrical field near the surface. However, they can travel several tens of micrometers before being captured by an electrical field. Since long-wavelength photoelectrons are generated deeper in the semiconductor than photoelectrons generated by light of shorter wavelengths, their tendency to travel before being detected (crosstalk) is much higher. Additionally, photoelectrons generated far from the electrically active zone need time to reach the demodulating electrical field. This temporal delay in arrival time will lead to an additional wavelength-dependent decrease in demodulation contrast for high frequencies.

The measurements in Fig. 5 show the obtained contrast decrease for high frequencies and longer wavelengths. For 20-MHz modulation and 630-nm wavelength, we achieve a modulation contrast of better than 40%, which is still quite good compared to the dc performance of about 60%. Furthermore, we can state that the simple model described in Fig. 4 predicts the dc performance very well.

3) Phase Accuracy Measurements: By varying the time delay of the modulated light source in steps of 1 ns, we can

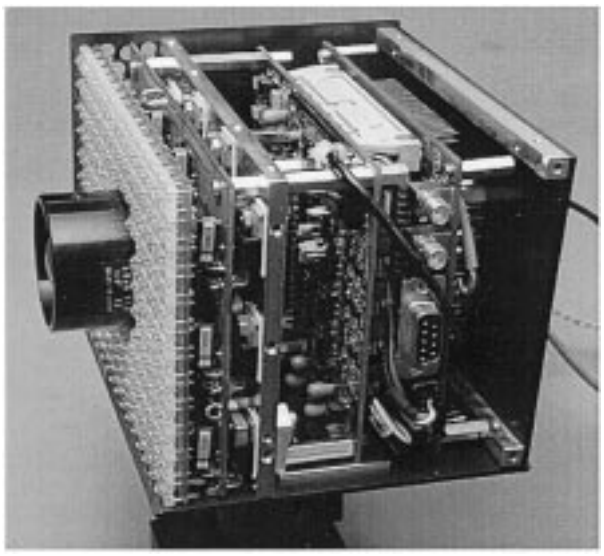


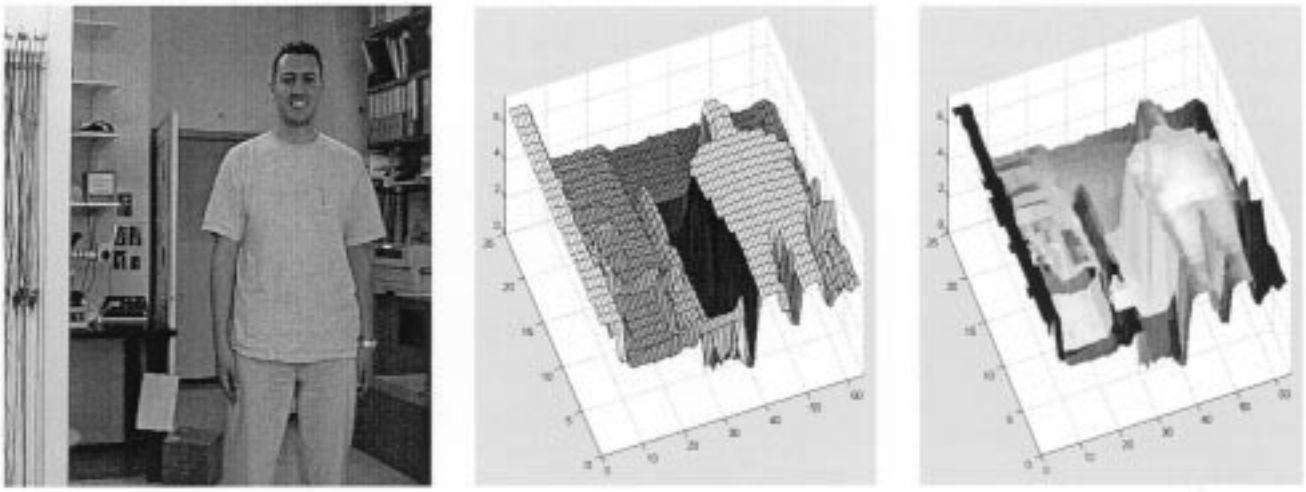
Fig. 8. Imaging TOF range camera (3-D). From left to right: LED board, LED driver board, heat sink for LED drivers, camera driver board, sequencer board. (Size: 95 mm (H) × 125 mm (W) × 105 mm (D)).

compare the measured time delay to the real time delay (or phase delay) over one complete modulation period while maintaining the same optical input power. Fig. 6 illustrates such a measurement for an optical input power of 736 fW per pixel, resulting in an accuracy of 260 ps. The measured time error can be converted into an equivalent distance error (50 ns =7.5 m).

By attenuating the LED light (with calibrated neutral density filters), the optical input power can be varied. For higher optical power, the distance accuracy increases, and for lower power, it decreases. This gives the possibility of measuring the range accuracy versus the optical power received in each pixel. The results of these measurements are shown Fig. 7, where also the theoretical resolution based on (16) is plotted.

Excellent agreement between theory—summarized in (11)—and the measured distance resolution is obtained. The optical input power of about 750 fW corresponds to a number of only 0.08 electrons generated in a modulation period time of 50 ns, i.e., we obtain a distance resolution of 4 cm if only one electron is generated, statistically speaking, every 13th modulation period. This is equivalent to a photocurrent of 250 fA (1 electron in $13 \cdot 50$ ns). Since the sampling interval is only half the modulation period, this means that one electron will be integrated to the sampling point only every 26th modulation period for this range resolution. For generation of a photoelectron every fourth modulation cycle, the range accuracy is 1.6 cm, i.e., a time resolution of 100 ps or a phase resolution of 0.75°.

A drawback of the 1-tap pixel structure is the fact that the four sampling points have to be acquired serially, since the pixel only contains one storage site. For long integration times, this "serial tap-integration" reduces the application area of the pixel in 3-D-measurement to scenes with relatively slow dynamics. We are aware of this problem and suggest an alternative structure with four storage sites per pixel in reference [9]. In order to avoid a nonhomogeneous sensitivity of the different sampling points, as could be observed for the 4-tap lock-in pixel [8]–[10], this new



(a) (b) (c)

Fig. 9. 3-D indoor scene. (a) High-resolution b/w photograph. (b) Measured 3-D data. The distance information is coded both in the z-axis-direction and in grey-level. (c) Measured 3-D and 2-D b/w data. The distance information is coded in the z-axis direction; the color of each pixel is the classical b/w intensity information, also measured with the TOF camera. ($T_{\rm int}$: 100 ms, 10-Hz frame rate)

pixel approach will use two separate optically identical light sensitive areas.

V. 3-D TOF Range Camera

Fig. 8 shows our all-solid-state 3-D time-of-flight camera, which is based on a customized photoASIC offering 64×25 1-Tap pixels. The LED illumination module can be exchanged; currently, we operate the camera either with 810-nm LEDs or with 630-nm LEDs. The LED illumination provides an optical output of about 900 mW. We can perform distance measurements with an integration time of 100 μs to 100 ms. The modulation frequency is 20 MHz, resulting in an unambiguous distance range of 7.5 m for the homodyne (phase-shifting) modulation method chosen. Range ambiguity can be resolved by heterodyne modulation modes (frequency shifting or multiple frequency modulation) in future work. With both illumination units, we achieve a ranging accuracy of better than 5 cm over the full distance range for noncooperative targets. The camera reaches a demodulation depth of 35% for the 630-nm module and 25% for the 810-nm module. The measurement of a 3-D scene is illustrated in Fig. 9.

Optical TOF-range measurement has several application-dependent advantages over the well-established microwave radar. The most important one is the higher lateral resolution. While the number of pixels for an optical range camera is practically unlimited, a 10-GHz microwave-radar would require for each pixel an antenna of 3.5-m diameter to resolve an area of 10-cm diameter in a 10-m distance only because of the diffraction limit ($D = 1.22 \cdot \lambda \cdot z/\Delta x$, D is the aperture of antenna, λ is the wavelength, z is the distance to the target, and Δx is the lateral resolution on the target).

On the other hand, microwave radars are barely influenced by rain and fog, because—in contrast to optical systems—their wavelength is longer than the typical diameter of the water drops. The performance of optical TOF-systems is therefore inferior under bad weather conditions.

VI. Sub-Micron Standard CMOS Technology for Lock-In Pixels

The underlying principle of the demodulation pixels is the CCD principle, which allows for transport of free charge carriers into defined directions within the semiconductor and also offers the possibility of adding charge packets essentially free of noise, which is not easily possible with any other CMOS circuitry. Our current sensors are realized with a special CMOS/CCD process. Although the CCD principle is essential for the pixel concept, in contrast to standard CCD's, good CCD performance is not required. This is because we do not need to transport the charge signal over hundreds of CCD gates but only over a few. The next generation of our TOF-sensor arrays will, therefore, be realized in a standard CMOS technology. However, care has to be taken to choose the right process. In contrast to traditional CCD processes, most CMOS processes do not offer a second gate mask, which would allow to realize overlapping CCD gates. Instead one has to realize CCD structures with gaps between the gates. To make possible a charge transfer from one gate to the other, no potential barrier must appear between the gates; thus, the diffusion of any active region between the gates has to be avoided. Today's developments in CMOS technology lead to a continuing decrease in minimum feature size, which allows one to realize gaps in the range of fractions of m icrons between two "CCD" gates. However, at the same time, the inter-gate-gaps become smaller the substrate doping increases. This leads to a smaller extension of the depletion region into the semiconductor.

To realize CCD's in a standard CMOS process, the appropriate combination of minimum gate-gate distance (CCD gap) and surface substrate doping concentration is important. The condition that charge carriers can move from one CCD gate to the neighboring gate is that the space charge regions of the gates overlap.

Fig. 10 shows the depletion width as a function of substrate doping concentration and effective gate voltage $V_G - V_{\rm FB}$ (gate voltage minus flatband voltage).

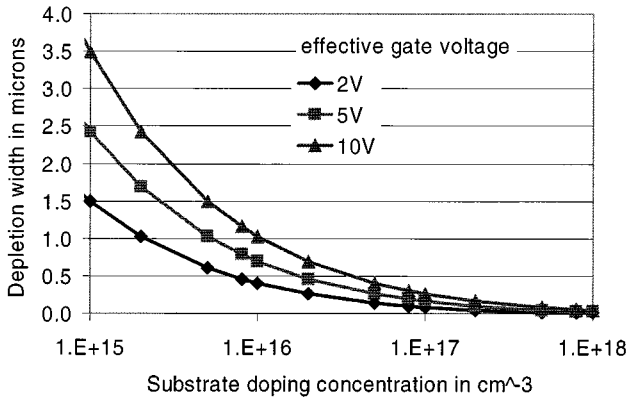


Fig. 10. Width of depletion region of a MOS gate versus substrate doping concentration and effective gate voltage [9].

According to Fig. 10 for a process with a substrate doping concentration of 10^{17} cm^{-3}, the depletion regions of two neighboring gates with a gap of 0.5 μm would not touch, even for gate voltages of 5 V. For a process with a 10^{15} cm^{-3} substrate, however, 2 V would be sufficient for the depletion regions to touch for a gap width of 3 μm. To summarize, a standard CMOS process can be used to realize the CCD charge transport, and hence lock-in pixels, if the following conditions are fulfilled.

1) Substrate doping concentration and minimum gate-to-gate distance (both as low as possible) have to lead to lateral overlapping space charge regions for moderate gate voltages ($\leq 2\,\mathrm{V} \ldots 5\,\mathrm{V}$).
2) The gate material has to be transparent, i.e., not siliconized.

Successful realizations of CCDs in single-gate-layer CMOS processes have been reported in [11] and [12].

VII. Discussion

We have presented pixel structures that are capable of quickly sampling modulated optical waves. This makes them ideally suited for the realization of imaging 3-D TOF cameras, where every pixel can measure a distance in addition to the scene intensity. For such cameras, no mechanically moving parts are necessary and, with the single exception of the photoASIC, only standard components are required, resulting in relatively low costs of realization. The performance of the introduced 1-tap pixel is close to the physical limitations. Even for single photons arriving only every few modulation cycles, the arrival time of the light wave can be measured with an accuracy of a few hundred picoseconds, resulting in a range accuracy of a few centimeters. Also, we derived an important connection between received optical power, background radiation, and the resulting distance resolution. This equation allows one to precisely predict the range accuracy for known scene reflectivity, illumination power and integration time.

We have introduced a complete 3-D range camera which proves the validity of our theoretical estimations and shows that it is possible to perform imaging TOF range measurements with our new pixel arrays for reasonable modulated illumination power.

References

[1] S. Coren, L. Ward, and J. Enns, *Sensation and Perception*, 4th ed. Orlando, FL: Harcourt Brace, 1994.
[2] P. Besl, "Active optical range imaging sensors," *Machine Vis. and Applic.*, vol. 1, pp. 127–152, 1988.
[3] R. G. Dorsch, G. Hausler, and J. M. Herrmann, "Laser triangulation: Fundamental uncertainty in distance measurement," *Appl. Opt.*, vol. 33, pp. 1306–1314, 1994.
[4] G. Beheim and K. Fritsch, "Range finding using frequency-modulated laser diode," *Appl. Opt.*, vol. 25, pp. 1439–1442, 1986.
[5] I. Moring, T. Heikkinen, R. Myllyla, and A. Kilpela, "Acquisition of three-dimensional image data by a scanning laser rangefinder," *Opt.Eng.*, vol. 28, pp. 897–905, 1989.
[6] W. Boyle and G. Smith, "Charge coupled semiconductor devices," *Bell Syst. Tech. J.*, vol. 49, pp. 587–593, 1970.
[7] R. Lange, P. Seitz, A. Biber, and S. Lauxtermann, "Demodulation pixels in CCD and CMOS technologies for time-of-flight ranging," in *Proc. SPIE*, vol. 3965A, San Jose, CA, 2000, pp. 177–188.
[8] T. Spirig, P. Seitz, O. Vietze, and F. Heitger, "The lock-in CCD—Two-dimensional synchronous detection of light," *IEEE J. Quantum Electron.*, vol. 31, pp. 1705–1708, Sept. 1995.
[9] R. Lange, "3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology," Ph.D. dissertation, Univ. Siegen, Dept. Elect. Eng. Comput. Sci., Germany, 2000.
[10] T. Spirig, "Smart CCD/CMOS Based Image Sensors with Programmable, Real-Time, Temporal and Spatial Convolution Capabilities for Applications in Machine Vision and Optical Metrology," Ph.D. dissertation, ETH, Zurich, Switzerland, 1997.
[11] M. Furumiya, K. Hatano, Y. Nakashiba, I. Murakami, T. Yamada, T. Nakano, Y. Kawakami, T. Kawasaki, and Y. Hokari, "A $1/2$-inch 1.3 N-pixel progressive scan CCD image sensor employing 0.25 mm gap single-layer Poly-Si electrodes," in *Proc. IEEE Int. Solid-State Circuits Conf.*, San Francisco, CA, 1999.
[12] J. Kramer, "Photo-ASICs: Integrated Optical Metrology Systems with Industrial CMOS Technology," Ph.D. dissertation, ETH, Zurich, Switzerland, 1993.

Robert Lange received the Ph.D. degree in technical sciences from the University of Siegen, Siegen, Germany, in 2000.

He was with the Image Sensing Group, Centre Suisse d'Electronique et de Microtechnique, Zurich, Switzerland, during 1997–2000, where his research activities were mainly focused on the new development and operation of application-specific 2-D and 3-D range image sensors in CMOS and CCD technology. He is currently with GretagMacbeth, Regensdorf, Switzerland, where he works in the field of color measurement.

Peter Seitz (M'88) received the Ph.D. degree in physics in 1984 from ETH Zurich, Switzerland.

From 1984 to 1987, he was a staff member of the RCA Research Laboratories, in both Princeton, NJ, and Zurich, Switzerland. In 1987, he joined the Paul Scherrer Institute, Zurich, Switzerland, where he formed the Image Sensing Research Group. Since 1997, he has been with the Swiss Center for Electronics and Microtechnology (CSEM), in both Neuchâtel and Zurich, Switzerland, heading the Image Sensing Section in the Photonics Division. He is also an Associate Professor at the Institute for Microtechnology, University of Neuchâtel, Neuchâtel, Switzerland. He is the author of over 100 publications in the fields of applied optics, image sensing, machine vision, and optical microsystems engineering, and holds 10 patents.

Dr. Seitz is the Secretary General of the European Optical Society EOS and a member of the Swiss Physical Society.