MAXIMUM LIKELIHOOD RANGE ESTIMATION FOR 3D-IMAGING BASED ON PSEUDO-NOISE OPTICAL MODULATION

M'Hamed-Ali El Mechat[1] *and Hans-Andrea Loeliger*[2]

[1]CSEM Zurich/ETH Zurich, Dept. Inform. Techn. & Electr. Eng.
Technoparkstrasse 1, 8005, Zurich, Switzerland
phone: + (41) 44 497 1427, fax: + (41) 44 497 1400, email: mem@csem.ch
[2] ETH Zurich, Dept. Inform. Techn. & Electr. Eng.
CH-8092 Zurich, Switzerland
email: loeliger@isi.ee.ethz.ch

ABSTRACT

Acquiring real-time 3D images poses many challenges. Many 3D-imaging systems are based on estimating the time-of-flight (TOF) of photons between an emitter and a receiver, which in turn is based on transmitting a pseudo-noise optical signal and correlation measurements. The precision of such measurements is limited by photon shot noise. In this paper, we develop a maximum-likelihood estimator for the TOF (and thus for the range) that takes into account the Poisson distribution of the detected photons. The proposed estimator is shown to give better range estimates than a standard estimator that is used in current systems. The improvement is particularly significant when the period of the pseudo-noise sequence-based modulation signal is short.

1. INTRODUCTION

The time-of-flight (TOF) principle is a major technique, used by several systems [1, 2], for real-time optical acquisition of range images. This principle is depicted in Figure 1.
An emitter sends an optical signal $x(t)$, which is reflected by the target and travels back to a receiver. From the round-trip delay τ, the range can be computed according to

$$R = \frac{c \cdot \tau}{2} \tag{1}$$

where c is the speed of light.

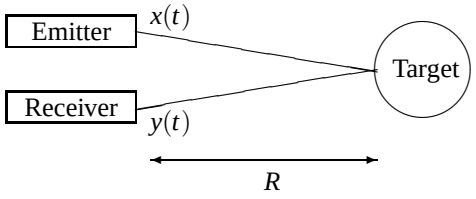

Figure 1: General scheme of a 3D-TOF optical system. A transmitter sends an optical signal $x(t)$. The receiver detects the optical signal $y(t)$. The latter is processed for extracting the TOF.

In this paper, we use the continuous-wave version of the TOF: the intensity of the optical signal is continuously modulated and the phase shift between the emitted signal and the detected one is exploited to compute the TOF [3].

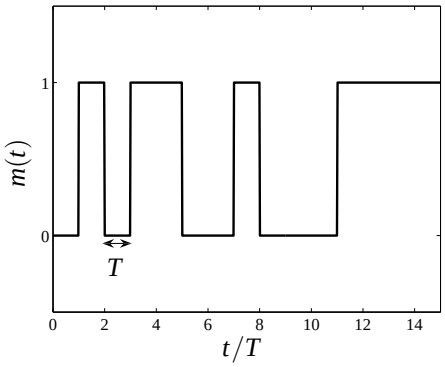

Figure 2: An example of a m-sequence of 15 chips.

The modulation signal $x(t)$ is defined in this paper as:

$$x(t) = \Gamma \cdot m(t) \tag{2}$$

where Γ is the optical peak-to-peak power and $m(t)$ belongs to a specific class of pseudo-noise (PN) sequences: the maximum-length sequences or m-sequences.
The m-sequences are generated by linear-feedback shift registers with irreducible polynomials [4]. Following the common terminology used in communication systems, each bit of the m-sequence is referred now as a chip. Each m-sequence has two versions: the unipolar version where the chips can take the value 0 or 1 and the bipolar one where the 0's of the unipolar version are converted into -1. Figure 2 shows one period of a m-sequence of 15 chips, where T is the chip duration.

In this paper, we will study the use of the maximum likelihood (ML) estimation of the TOF (and thus of R) based on a PN-modulated optical signal. The influence of the background (BG) light and of certain non-idealities at the receiver level are also considered. Even by taking into account these non-idealities, a manageable form of the maximum likelihood estimator (MLE) can still be achieved . We then assess the performance of the MLE and compare it to the standard method defined in [5].
The paper is structured as follows: section 2 presents the system model. In section 3, the range estimators are described before being compared in terms of performance in section 4.

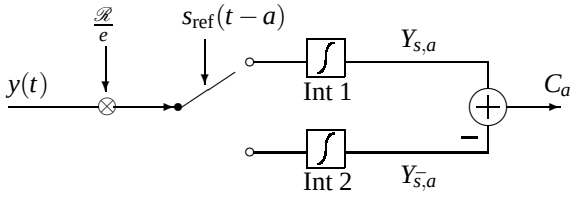

Figure 3: In-pixel demodulation process of the detected optical signal. The pixel output C_a results from the difference between the two integrators outputs.

2. SYSTEM MODEL

This section describes the system model specifications.

2.1 Ideal receiver

For an ideal system, the received signal can be expressed as:

$$y(t) \triangleq \alpha \cdot x(t-\tau) = \alpha \cdot \Gamma \cdot m(t-\tau) \tag{3}$$

where α is the attenuation factor due to the free-space loss effect [6].
This optical signal can be seen as a stream of photons. By assuming an efficient optoelectronic light source [7], the photon detection can be modeled as a collection of independent events. The counting process of these photons follows a Poisson distribution [8].
A certain amount of electrons are generated then by the photoelectric receiver. The number of these photo-electrons also follows a Poisson distribution. This effect is known as shot noise, which is the ultimate physical limitation of optical systems [1, 8].
We define K as the random variable that represents the number of generated photo-electrons during the integration time T_{int}. The probability distribution of K follows a Poisson law:

$$P(K=k|T_{\text{int}}) = \frac{e^{-\mu_K} \cdot (\mu_K)^k}{k!} \quad (k \in \mathbb{N})$$

where μ_K is the expected value of K for an integration time of T_{int}.
In this paper, we use the SwissRanger$^{\text{TM}}$ (SR) camera as a 3D-TOF system [9]. The camera emits the light in the near-infrared range. When the light travels back to the camera, it is demodulated through an array of pixels. Figure 3 depicts the in-pixel demodulation process [5].
The received optical signal $y(t)$ is converted into an electron current according to the pixel optical responsivity $\mathscr{R}$ [8]. The division by the elementary charge value e gives us the corresponding number of generated photo-electrons. The latter are accumulated on one of two integrator nodes determined by a switch, which is controlled by the reference signal $s_{\text{ref}}(t-a)$. The variable a represents a specific time-shift. The reference signal is the bipolar version of $m(t)$. When its value is 1 (resp -1), the electrons flow to the integrator Int 1 (resp. Int 2).
By accumulating the charges during T_{int} seconds, the integrator nodes act as electrons counters. Due to the shot noise effect, the integrator outputs, referred hereafter as charge packets, are Poisson distributed:

$$\begin{aligned} Y_{s,a} &\backsim \text{Poisson}(\mu_{s,a}) \\ Y_{\bar{s},a} &\backsim \text{Poisson}(\mu_{\bar{s},a}) \end{aligned}$$

where $\mu_{s,a}$ and $\mu_{\bar{s},a}$ correspond to the mean values. If we set T_{int} to one period of the m-sequences, the mean values can be expressed as [5]:

$$\begin{aligned} \mu_{s,a} \triangleq E[Y_{s,a}] &= E\left[\frac{\mathscr{R}}{e} \cdot \int_0^{T_{\text{int}}} y(t) \cdot \frac{(1+s_{\text{ref}}(t-a))}{2}\, dt\right] \\ &= \begin{cases} \mathscr{E}_x\left(2 - \frac{|a-\tau|}{T}\right) & \text{if } |a-\tau| \le T \\ \mathscr{E}_x \text{ otherwise} \end{cases} \end{aligned} \tag{4}$$

and

$$\begin{aligned} \mu_{\bar{s},a} \triangleq E[Y_{\bar{s},a}] &= E\left[\frac{\mathscr{R}}{e} \cdot \int_0^{T_{\text{int}}} y(t) \cdot \frac{(1-s_{\text{ref}}(t-a))}{2}\, dt\right] \\ &= \begin{cases} \mathscr{E}_x \cdot \frac{|a-\tau|}{T} & \text{if } |a-\tau| \le T \\ \mathscr{E}_x \text{ otherwise} \end{cases} \end{aligned} \tag{5}$$

where $\mathscr{E}_x$ is defined as the mean power generated by the signal component:

$$\mathscr{E}_x \triangleq E\left[\frac{\mathscr{R}}{2e}\int_0^{T_{\text{int}}} y(t)\, dt\right] = E\left[\frac{\mathscr{R}}{2e}\int_0^{T_{\text{int}}} \alpha \cdot x(t-\tau)\, dt\right]$$

Usually, several periods of the optical signal are integrated in order to increase the signal-to-noise ratio [8].
The difference between the charge packets, $Y_{s,a}$ and $Y_{\bar{s},a}$, gives us the pixel output C_a. The latter is also a random variable with the following mean value:

$$\begin{aligned} \mu_a \triangleq E[C_a] &= \mu_{s,a} - \mu_{\bar{s},a} \\ &= E\left[\frac{\mathscr{R}}{e} \cdot \int_0^{T_{\text{int}}} y(t) \cdot s_{\text{ref}}(t-a)\, dt\right] \qquad (6) \\ &= \begin{cases} 2\mathscr{E}_x\left(1 - \frac{|a-\tau|}{T}\right) & \text{if } |a-\tau| \le T \\ 0 \text{ otherwise} \end{cases} \qquad (7) \end{aligned}$$

For an ideal system, only the shot noise determines the performance limitations. In practice, additional non-idealities also have an impact on the performance.

2.2 Non-idealities

Two non-idealities are discussed in this section: the BG light and the demodulation contrast.

2.2.1 Background light

In some cases, additional light sources are emitting in the vicinity of the camera. They produce an extra amount of optical power received by the camera. If we denote this additional optical power by Γ_{BG}, equation (3) becomes:

$$y(t) = \alpha \cdot \Gamma \cdot m(t-\tau) + \Gamma_{\text{BG}}$$

We assume a constant Γ_{BG} during the integration time.

2.2.2 Demodulation contrast

During the demodulation process, errors can occur because some generated photo-electrons might flow to the wrong integrator node due to a non-ideal switch. These errors can be estimated by a parameter called the demodulation contrast c_d. The latter is included in the interval $[0;1]$ and indicates the inherent pixel demodulation quality. The closer the value c_d is to 1, the more the demodulation process becomes reliable [10].
By taking into account the BG light and the demodulation contrast, we can rewrite equations (4), (5) and (7) as [5]:

$$\mu_{s,a} = \begin{cases} \mathscr{E}_x\left(1 + c_d - c_d\frac{|a-\tau|}{T}\right) + \mathscr{E}_{\mathrm{BG}} \cdot \frac{n+c_d}{n} & \text{if } |a-\tau| \leq T \\ \mathscr{E}_x + \mathscr{E}_{\mathrm{BG}} \cdot \frac{n+c_d}{n} & \text{otherwise} \end{cases}$$

$$\mu_{\bar{s},a} = \begin{cases} \mathscr{E}_x\left(1 - c_d + c_d\frac{|a-\tau|}{T}\right) + \mathscr{E}_{\mathrm{BG}} \cdot \frac{n-c_d}{n} & \text{if } |a-\tau| \leq T \\ \mathscr{E}_x + \mathscr{E}_{\mathrm{BG}} \cdot \frac{n-c_d}{n} & \text{otherwise} \end{cases}$$

and

$$\mu_a = \begin{cases} 2\mathscr{E}_x c_d\left(1 - \frac{|a-\tau|}{T}\right) + \mathscr{E}_{\mathrm{BG}} \cdot \frac{2c_d}{n} & \text{if } |a-\tau| \leq T \\ \mathscr{E}_{\mathrm{BG}} \cdot \frac{2c_d}{n} & \text{otherwise} \end{cases} \quad (8)$$

where n is the sequence length and $\mathscr{E}_{\mathrm{BG}}$ is defined as the mean power generated by the BG light:

$$\mathscr{E}_{\mathrm{BG}} \triangleq E\left[\frac{\mathscr{R}}{2e}\Gamma_{\mathrm{BG}} \cdot T_{\mathrm{int}}\right] \quad (9)$$

The next section shows how the TOF and hence the range R are computed.

3. RANGE ESTIMATORS

In this section, we first describe the linear correlation estimator (LCE), which is currently implemented, and we then derive the maximum-likelihood range estimator.

3.1 Linear correlation estimator

Based on equation (6), it is obvious that the pixel output is a correlation between $y(t)$ and $s_{\mathrm{ref}}(t-a)$. By setting successively a to two different values: 0 and T and assuming $|\tau| \leq T$, we obtain two correlation samples (i.e. pixel outputs), C_0 and C_T, with the following mean values:

$$\begin{aligned} \mu_0 &\triangleq \mu_a\Big|_{a=0} = 2\mathscr{E}_x c_d\left(1 - \tilde{\tau} + \frac{\rho_{\mathrm{BG}}}{n}\right) \\ \mu_T &\triangleq \mu_a\Big|_{a=T} = 2\mathscr{E}_x c_d\left(\tilde{\tau} + \frac{\rho_{\mathrm{BG}}}{n}\right) \end{aligned} \quad (10)$$

where $\tilde{\tau} = \tau/T$ and $\rho_{\mathrm{BG}} = \mathscr{E}_{\mathrm{BG}}/\mathscr{E}_x$.
Figure 4 shows the behavior of C_0 and C_T according to the TOF value for a noise-free channel (i.e. $C_0 = \mu_0$, $C_T = \mu_T$, $c_d = 1$ and $\mathscr{E}_{\mathrm{BG}} = 0$).
Based on this figure and equations (10), we can notice that the more the TOF increases, the more C_0 increases and C_T decreases. When the TOF is larger than T, one or both samples are equal to 0 and the range can no longer be computed. Hence, the maximum TOF value, which can be detected, is equal to T, which gives us a maximum detectable range of $R_{\mathrm{Max}} = cT/2$.
From these restrictions and the behavior of the correlation

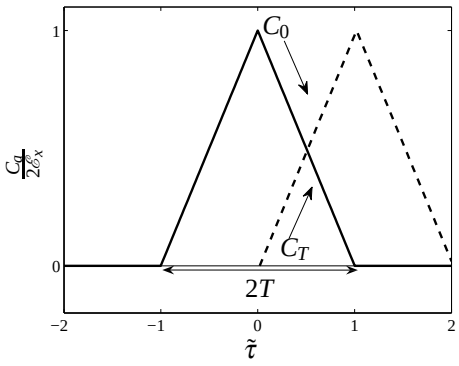

Figure 4: Pixel outputs behavior in a noise-free channel.

samples, we derived the formulas for the normalized TOF and the corresponding range [5]:

$$\tilde{\tau}_{\mathrm{LCE}} = \frac{C_T}{C_0 + C_T}$$

and

$$R = R_{\mathrm{Max}} \cdot \tilde{\tau}_{\mathrm{LCE}} \quad (11)$$

3.2 Maximum likelihood range estimator

According to the previous section, two correlation values are needed by the LCE for computing the range. Each correlation value is a difference between two charge packets. Thus, we collect in total four charge packets: $Y_{s,0}$, $Y_{\bar{s},0}$, $Y_{s,T}$ and $Y_{\bar{s},T}$. The latter are statistically independent and Poisson distributed, which means that their joint probability distribution (JPD) can be written as:

$$\begin{aligned} f_\theta &= \prod_{\mathrm{i,j}} P(Y_{\mathrm{i,j}} = y_{\mathrm{i,j}} | \mu_{\mathrm{i,j}}) \\ &= \frac{e^{-\sum_{\mathrm{i,j}} \mu_{\mathrm{i,j}}}}{\prod_{\mathrm{i,j}} y_{\mathrm{i,j}}!} \cdot \prod_{\mathrm{i,j}} (\mu_{\mathrm{i,j}})^{y_{\mathrm{i,j}}} \end{aligned} \quad (12)$$

where $i \in \{s, \bar{s}\}$ and $j \in \{0, T\}$.
The variable θ denotes the vector that contains the unknown parameters of the JPD. In our case, the vector θ has four unknown components:

$$\theta = (\mathscr{E}_x, c_d, \rho_{\mathrm{BG}}, \tilde{\tau})$$

According to the ML principle [11], we want to maximize equation (12) with respect to θ. In other words, we must find the particular value of θ, denoted by $\widehat{\theta}$, that fulfills the following equation system:

$$\begin{aligned} \nabla(f_\theta) &= 0 \\ \mathbf{z}^T(\Delta(f_\theta))\mathbf{z} &\leq 0 \end{aligned} \quad (13)$$

where $\mathbf{z} \in \mathbb{R}^4$, T is the transpose operator, ∇ is the differential operator and Δ the Laplacian operator.
The first equation means that $\widehat{\theta}$ is an extremum of f_θ (the gradient vanishes at $\widehat{\theta}$). The second equation ensures that this extremum is a maximum (The Hessian matrix evaluated

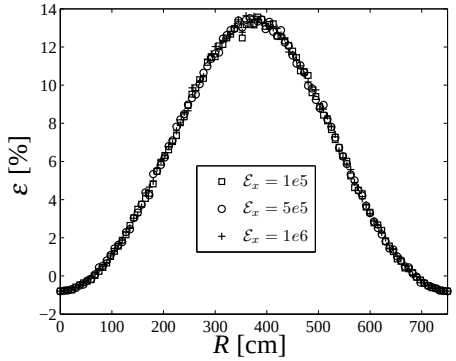

Figure 5: Behavior of the comparison parameter ε in a basic channel. The mean power $\mathscr{E}_x$ and the sequence length n do not have any impact on ε.

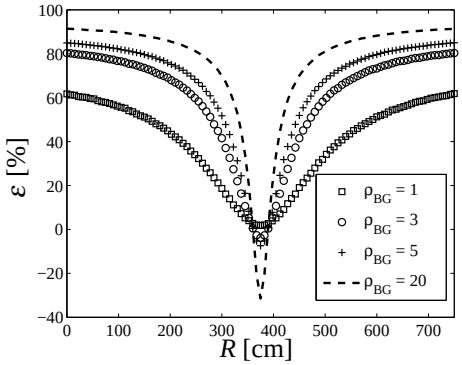

Figure 6: Behavior of the comparison parameter ε in a BG light channel. Case 1: we test different values of ρ_{BG} but we keep the sequence length constant ($n = 127$ chips).

at $\widehat{\theta}$ is negative definite) [12]. The sought TOF corresponds to the fourth component of $\widehat{\theta}$.
Since no explicit solution is obtained for the TOF when we solve the equation system (13), we repeated the computations assuming a demodulation contrast c_d equal to 1. This strong assumption relies on the recent progress in the camera technology, which tend to increase the demodulation quality. Based on this assumption, an explicit solution is found for the normalized TOF:

$$\tilde{\tau}_{\text{MLE}} = \frac{n(Y_{s,0}+Y_{\bar{s},0})(Y_{\bar{s},T}-Y_{s,T})}{2(n(Y_{\bar{s},0}Y_{\bar{s},T}-Y_{s,0}Y_{\bar{s},0})+(Y_{s,0}+Y_{\bar{s},0})(Y_{s,T}+Y_{\bar{s},T}))} + \frac{Y_{\bar{s},T}(Y_{\bar{s},0}-Y_{s,0})+Y_{s,T}(Y_{s,0}+3Y_{\bar{s},0})}{2(n(Y_{\bar{s},0}Y_{\bar{s},T}-Y_{s,0}Y_{\bar{s},0})+(Y_{s,0}+Y_{\bar{s},0})(Y_{s,T}+Y_{\bar{s},T}))}$$

The range is computed by substituting $\tilde{\tau}_{\text{LCE}}$ in (11) by $\tilde{\tau}_{\text{MLE}}$.

4. COMPARISON

In order to evaluate the performance of both range estimators, we estimate the root mean squared error (RMSE):

$$\text{RMSE} = \sqrt{E[(R-R_{\text{tr}})^2]}$$

where R is the computed range and R_{tr} is the true one.
From the RMSE, we define the following comparison parameter:

$$\varepsilon = \frac{\text{RMSE}_{\text{LCE}} - \text{RMSE}_{\text{MLE}}}{\text{RMSE}_{\text{LCE}}} \qquad (14)$$

If ε is positive (resp. negative), it means that the MLE is more (resp. less) accurate than the LCE. We estimated ε for three channels: the basic channel, the BG light channel and the demodulation contrast channel. For all simulations performed, the chip duration T is set to 50ns, which leads to a maximum distance R_{max} of 750cm.

4.1 Basic channel

In this channel, there is no BG light (i.e $\rho_{\text{BG}} = 0$) and the demodulation process is perfect (i.e. $c_d = 1$). The range accuracy is only limited by the shot noise effect. Figure 5 depicts the behavior of ε according to the range and for three different values of $\mathscr{E}_x$.

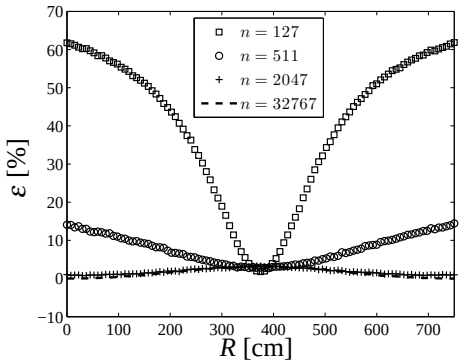

Figure 7: Behavior of the comparison parameter ε in a BG light channel. Case 2: we keep the BG light ratio constant ($\rho_{\text{BG}} = 1$) but we test different values of n.

The MLE offers a better accuracy than the LCE over the whole range interval. The improvement depends on the range, with a maximum gain of approximately 14%, achieved at a range of 375cm. These results are independent on the sequence length n and the mean power $\mathscr{E}_x$.

4.2 Background light channel

In this channel, a certain amount of BG light is added to the system but we assume a perfect demodulation process, i.e. $c_d = 1$. In the following example, the BG light ratio ρ_{BG} is set to different values and the sequence length is fixed to 127 chips.
Figure 6 shows that the MLE is globally more accurate than the LCE. The improvement is even more important when the BG light increases in the system. It means that the MLE is more resistant to the BG light effect than the LCE. For a BG light ratio ρ_{BG} of 20, a gain of up to 90% can be achieved. However, for a small range interval around 375cm, the LCE is more accurate and can offer a gain of up to 30%.
Another test was performed for different sequence lengths and the BG light ratio ρ_{BG} set to 1. The results are displayed in Figure 7. We can see that as the sequence length increases then the difference between the estimator decreases. In other words, for long sequences, the two estimators offer the same performance.

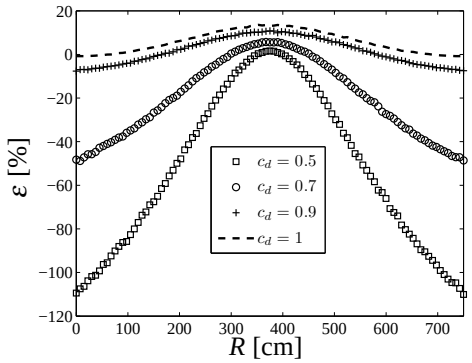

Figure 8: Behavior of the comparison parameter ε in a demodulation contrast channel. Case 1: we test different values of c_d but we keep the sequence length constant ($n = 127$).

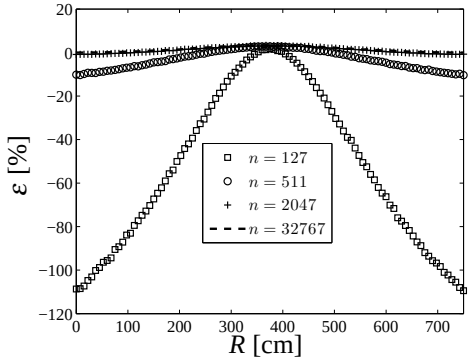

Figure 9: Behavior of the comparison parameter ε in a demodulation contrast channel. Case 2: we keep the demodulation contrast constant ($c_d = 0.5$) but we test different values of n.

4.3 Demodulation contrast channel

In this channel, there is no BG light (i.e. $\rho_{BG} = 0$), but the demodulation process is no longer perfect. Figure 8 displays the results for different values of c_d and a sequence length n set to 127 chips.

Clearly, the LCE offers better performance globally than the MLE. The gain reaches 100% at the range interval boundaries. We expected this performance loss for the MLE because the range formula of this estimator is derived assuming a demodulation contrast c_d of 1. This bias induces errors when the system has a non-perfect demodulation process. However, the more c_d increases then the closer the MLE performances are to the LCE performances. When c_d is equal to 1, we come back to the basic channel case.

We also performed a test for different sequences length and c_d fixed to 0.5. The results are displayed in Figure 9.

Similarly to the BG light channel, long sequences tend to make the two estimators comparable in terms of performance.

5. CONCLUSION

We presented in this paper a new range estimator based on the maximum likelihood principle, which uses the statistical properties of the collected charge packets. In channels dominated by the shot noise, the MLE can offer a maximum gain accuracy of 14% over the LCE. The MLE is also adapted to take into account the nonidealities present in a real system. For BG light channels, the MLE provides gains of up to 90%. However, when the demodulation process is not perfect, the MLE is worst than the LCE. The reason is due to a bias introduced in the range formula. The bias effect tends to be attenuated with recent technological progress. All these improvements are clearly visible when short length m-sequences are used. These results show us that the MLE gives an improvement over the LCE and an implementation of the MLE in a real camera system can be considered.

6. ACKNOWLEDGEMENTS

The authors would like to thank the CSEM image sensing group, especially Nicolas Blanc and Andrew Perkins. The help of Bernhard Büttgen of Mesa Imaging AG was also greatly appreciated.

REFERENCES

[1] P. Seitz, “Unified analysis and physical limitations of optical range-imaging techniques ” in *Proc. 1st Range Imaging research Day*, Zurich, Switzerland, September 8-9. 2005, pp. 9–19.

[2] F. Blais, “Review of 20 Years of Range Sensor Development ” in *Journal of Electronic Imaging*, vol. 13, pp. 231–240, Jan. 2004.

[3] R. Lange and P. Seitz, ”Solid-state time-of-flight range camera,” *IEEE Journal of Quantum Electronics*, vol. 37, pp. 390–397, Mar. 2001.

[4] K. Kettunen, “Code Selection for CDMA Systems ” Department of Information Studies, University of Tampere, Finland, 1997.

[5] B .Büttgen, M.E. Mechat, F. Lustenberger and P.Seitz, “Pseudonoise Optical Modulation for Real-Time 3-D Imaging With Minimum Interference ” in *IEEE Trans. on Circuits and Systems I*, vol. 54, pp. 2109–2119, Oct. 2007.

[6] J. Schiller, *Mobile Communications*. Address: Pearson Education, Harlow, 2003.

[7] P. Seitz, “Quantum-Noise Limited Distance Resolution of Optical Range Imaging Techniques ” in *IEEE Trans. on Circuits and Systems I*, vol. 55, pp. 2368–2377, Sep. 2008.

[8] B.E.A. Saleh and M.C. Teich, *Fundamental of Photonics*. Address: Wiley, New York, 1991.

[9] www.swissranger.ch

[10] B .Büttgen and P. Seitz, “Robust Optical Time-of-Flight Range Imaging Based on Smart Pixel Structures ” in *IEEE Trans. on Circuits and Systems I*, vol. 55, pp. 1512–1525, Jul. 2008.

[11] D.S.R. Eliason, *Maximum Likelihood Estimation: Logic and Practice*. Address: U.K:SAGE, London, 1993.

[12] I. N. Bronstein, K.A. Semendjajew, *Taschenbuch der Mathematik*. Address: Verlag Harri Deutsch, Frankfurt, 1997.