Pseudo-direct ToF imaging using a multi-tap macro-pixel CMOS image sensor with oversampled reconstruction

Pham Ngoc Anh[1], Thoriq Ibrahim[1], Keita Yasutomi[2], Shoji Kawahito[2], Hajime Nagahara [3], Keiichiro Kagawa[2]
1 Graduate School of Integrated Science and Technology, Shizuoka University
3-5-1 Johoku, Hamamatsu, 432-8011 Japan
2 Research Institute of Electronics, Shizuoka University
3-5-1 Johoku, Hamamatsu, 432-8011 Japan
3 Institute for Datability Science, Osaka University
2-8 Yamadaoka, Suita, Osaka 565-0871 Japan
E-mail: kagawa@idl.rie.shizuoka.ac.jp

Abstract A pseudo-direct ToF (dToF) imaging method using an iToF image sensor composed of multi-tap macro-pixels in combination with compressed sensing is presented. This method can reproduce the received light waveforms as obtained in the dToF based on single-photon avalanche diodes (SPADs) while maintaining high spatial and temporal resolution. By applying the oversampling to the reconstruction, the received light waveforms are reconstructed with higher temporal resolution. Experimental results showed that distance precision was improved by the proposed method.

Keywords: ToF imaging, multi-tap CMOS image sensor, oversampled reconstruction, compressed sensing

1. Introduction

There are two methodologies for time-of-flight (ToF) depth imaging: direct (dToF) and indirect (iToF). The dToF sensor, based on the single electron avalanche diode (SPAD) [1], directly measures the reflected light waveform, which is immune to multipath interference (MPI), but requires a large area for time-to-digital converters and histogram builders. In contrast, the iToF sensor calculates the depth by the correlations between the received light waveform and the applied demodulation waveforms. iToF has a smaller circuit area and can estimate the depth with a higher spatial resolution, but is susceptible to MPI.

We have proposed a pseudo-direct ToF CMOS image sensor that can reproduce the received optical waveforms similar to those obtained by dToF. This method is based on the iToF image sensor [2]. The difference is that random temporal exposure codes are applied in demodulation. By using a multi-tap macro-pixel CMOS image sensor, which has multi-tap lateral electric field charge modulators (LEFM [3]), combined with compressed sensing, it is possible to efficiently sample a large amount of spatio-temporal data of the received optical waveforms in a single shot.

However, the precision of depth was still limited by the clock frequency of the coded exposure patterns. To improve this, we utilize the oversampling technique in image reconstruction by increasing the number of data points at a sampling frequency higher than the Nyquist rate. In this work, we measured the impulse response of the system for the temporal exposure codes at a sampling rate 10 times higher than that of the exposure codes. Experimental results show that the standard deviation of the depth is reduced with the oversampling method.

2. Pseudo-dToF CMOS image sensor with oversampling

The entire process has two major phases: sensing and reconstruction. Firstly, in the sensing phase, the image sensor measures the spatio-temporal optical signals emitted from a synchronized laser and reflected from objects. This sensor consists of a macro-pixel array, each of which has four subpixels (Fig. 1). Each subpixel contains a photodiode and four taps, consisting of a charge transfer gate and a charge storage diode. Photodiodes in the center of the subpixels convert photons into photoelectrons. Then, they are transferred to the tap where the transfer gate is open. The gates turn on and off multiple times during the exposure time, which is called coded exposure. Thus, the cross-correlation between the incident temporal optical signal and the electrical shutter signal is calculated in the charge domain without any signal processing circuits. Finally, the compressed pixel values for the four charge modulators of all subpixels are obtained (Fig. 2).

Secondly, in the reconstruction stage, the original optical waveforms for all subpixels are reproduced by solving the inverse problem. We utilize an iterative method based on the sparsity regularization. The image acquisition process, including imaging optics and our image sensor, is denoted by the following linear equation,

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{1}$$

where $\boldsymbol{x}$ is the optical input signal and $\boldsymbol{y}$ is the signal measured. Matrix $\boldsymbol{A}$ is a pre-measured spatio-temporal observation matrix. It includes the spatial point spread function of the imaging optics and the temporal impulse response of the sensor for the exposure codes. If the dimension of $\boldsymbol{y}$ is lower than $\boldsymbol{x}$, the signal is compressed and becomes an ill-posed problem. However, the input $\boldsymbol{x}$ can be reproduced by solving the inverse problem with sparsity constraints from the known $\boldsymbol{y}$ and $\boldsymbol{A}$. This process is performed by TVAL3, a compressive sensing solver, which minimizes the total variation as shown in Eq. 2. Here, $\boldsymbol{D_i}$ shows a spatio-temporal differential operator [4].

$$\hat{\boldsymbol{x}}^{(TV)} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \Sigma_i \parallel \boldsymbol{D}_i \boldsymbol{x} \parallel_1 \ \text{subject to}\ \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \tag{2}$$

To retrieve the depth for each subpixel, we find the peak temporal position in the reconstructed optical waveform. Quadratic fitting was performed to refine the depth.

The point of this study is to oversample the temporal impulse response. Having the matrix $\boldsymbol{A}$ temporally oversampled, we can increase the dimension of $\boldsymbol{x}$ after the reconstruction, i.e., we can obtain a denser optical waveform. The benefit of this method is that the temporal resolution can be improved without changing the image sensor hardware or the total exposure time, although the reconstruction takes longer.

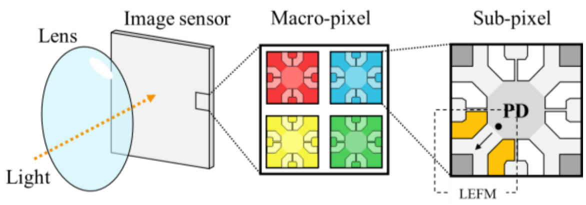

Fig. 1. Image sensor configuration. One macro-pixel is composed of 2×2 four-tap subpixels. The photoelectrons are transferred to one of the taps (through the gate shown in yellow).

3. Experiment results

ToF imaging experiments were conducted in the depth range of 0 - 16 m to verify the effectiveness of oversampling at a 10× ratio. The clock frequency of the 32-bit exposure codes was 303 MHz, i.e., a 1-bit code corresponds to 3.3 ns or 0.5 m in depth. A pulsed semiconductor laser with a wavelength of 660 nm and a pulse width of 7 ns was used. Fig. 3 shows the configuration of the targets. The depth maps for normal and 10× oversampling, are shown in Fig. 4. Dark frame subtraction and 100-image averaging were applied to improve the SNR. Fig. 5 compares the reproduced optical waveforms and their fitting curves with a quadratic approximation. The fitting curve is more accurate due to the increase in the number of points. The mean and standard deviation of the depth are quantitatively compared in Table 1.

4. Conclusion

In this paper, we demonstrated pseudo-dToF depth imaging with the oversampling of the impulse response for the exposure codes and showed that the standard deviation of the estimated depth was improved.

5. Acknowledgments

This work was supported by JST, CREST, JPMJCR22C1, and in part by Grants-in-Aid for Scientific Research (S), numbers 17H06102 and 18H05240. This work was also supported by VLSI Design and Education Center (VDEC), The University of Tokyo, with the collaboration with Cadence Corporation, Synopsys Corporation, and Mentor Graphics Corporation.

References

[1] C. Niclass, A. Rochas, P-A. Besse, E. Charbon: "Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes", IEEE J. Solid-State Circuits, Vol. 40, Issue 9 (2005).

[2] S. Kawahito, G. Baek, Z. Li, S. Han, M. Seo, K. Yasutomi, K. Kagawa: "CMOS lock-in pixel image sensors with lateral electric field control for time-resolved imaging", Int'l Image Sensor Workshop (2013).

[3] K. Kagawa, M. Horio, A-N. Pham, T. Ibrahim, S. Okihara, T. Furuhashi, T. Takasawa, K. Yasutomi, S. Kawahito, and H. Nagahara: "A dual mode 303 Megaframes per second charge domain time compressive computational CMOS image sensor", MDPI Sensors, Vol. 22, Article 1953 (2022).

[4] C. Li, W. Yin, Y. Zhang: "An efficient augmented Lagrangian method with applications to total variation minimization," Computational Optimization and Applications, Vol. 56 (2013).

Table. 1. Mean and standard deviation of objects' depth.

	Normal sampling		10x over sampling	
	Mean [m]	Standard deviation [cm]	Mean [m]	Standard deviation [cm]
Pillar	3.93	4.86	3.93	4.44
Pylon	5.94	4.12	5.97	4.79
Doll	7.91	28.59	7.94	2.80
Panel	12.96	18.45	12.89	4.37

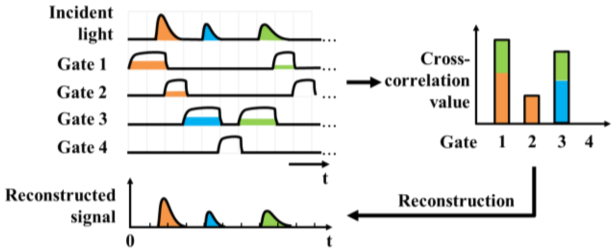

Fig. 2. Signal compression and reconstruction flow.

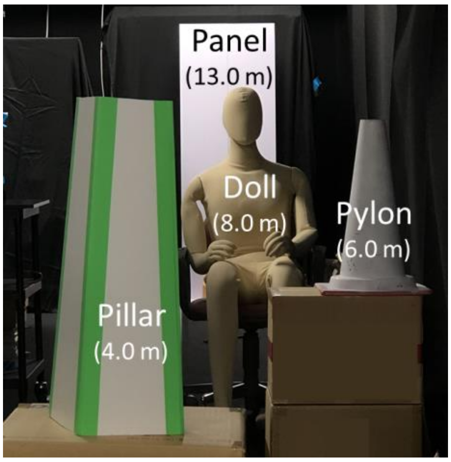

Fig. 3. Setup for ToF imaging experiment.

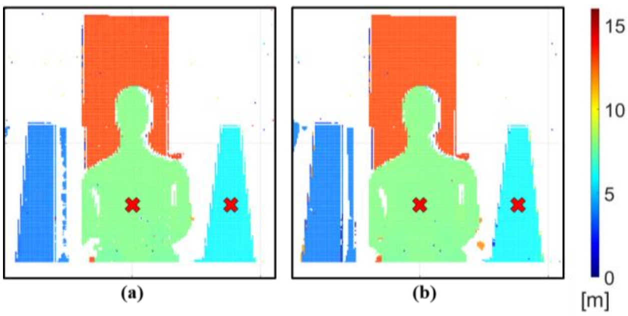

Fig. 4. Depth maps: (a) normal sampling, (b) 10× oversampling.

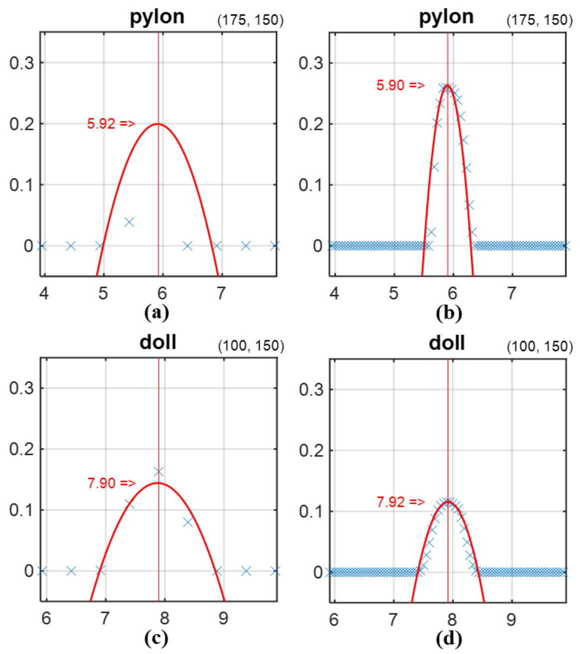

Fig. 5. Reconstructed waveforms and signal peaks: (a, c) normal sampling, (b, d) 10× oversampling. (Objects: pylon and doll, picked-up pixel's locations are marked with a red × in Fig. 4).