A 607 MHz time-compressive computational pseudo-dToF CMOS image sensor

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, Naka Ward, Hamamatsu, Shizuoka, Japan 432-8011
[2] Research Institute of Electronics, Shizuoka University
3-5-1 Johoku, Naka Ward, Hamamatsu, Shizuoka, Japan 432-8011
[3] Institute for Datability Science, Osaka University
2-8, Yamadaoka, Suita, Osaka, Japan 565-0871
Email: kagawa@idl.rie.shizuoka.ac.jp

***Abstract* - This paper demonstrates a pseudo-direct time-of-flight (pseudo-dToF) CMOS image sensor that is robust to multipath interference (MPI) and has high distance accuracy and precision. This method uses an iToF-based image sensor, but can reconstruct the received light waveforms similar to those obtained by conventional dToF image sensors based on single-photon avalanche diode (SPAD). Therefore, this method has the advantages of both dToF and iToF depth image sensors such as high resolution, high accuracy, immunity to MPI, and motion-artifact-free. This paper presents a signal reconstruction scheme for our laboratory-designed time-compressive image sensor based on the charge domain compressive sensing. Two approaches to refine the depth resolution are explained: 1) Increasing the operating clock speed; 2) Oversampling in image reconstruction and quadratic fitting in depth calculation. Experimental results show the separation of two reflections 40 cm apart under an MPI condition, and a significant improvement in distance precision down to 1 cm order. These results suggest that this method could be a promising approach to virtually implement dToF imaging suitable for challenging environments with MPI.**

I. INTRODUCTION

Time-of-flight (ToF) depth imaging calculates the distance of an object by measuring the travel time of light emitted from the camera to the object and back to the camera. This method is increasingly utilized in the fields of robotics and automotive applications. In ToF cameras, each pixel provides a specific depth value. However, the target scene may involve multiple light paths that interact with the same pixel, resulting in depth images that can contain scene-dependent errors due to multipath interference (MPI).

There are two major methodologies for traditional ToF depth imaging: direct ToF (dToF) and indirect ToF (iToF). The dToF sensor, based on the single electron avalanche diode (SPAD) [1], directly measures the reflected light waveforms, making it immune to MPI. However, it requires a larger circuit area for time-to-digital converters and histogram builders. In contrast, the iToF sensor [2] has a smaller circuit area and can estimate the depth with a higher spatial resolution, but it is susceptible to MPI. This is because it calculates the depth from the number of charges that correlate the incident light waveform with a demodulation function applied to the modulator.

In this work, we propose a new measurement method called pseudo-dToF, which provides the advantages of both dToF and iToF. This method utilizes iToF-based high-speed charge modulators, allowing for high-resolution imaging similar to iToF image sensors. Moreover, the time-compressive sensing in the charge domain enables the sensor to reproduce the entire light waveform in a single shot. Therefore, pseudo-dToF realizes high accuracy in depth and motion artifact-free measurement. We also present two approaches to achieve higher temporal resolution and report the improvements in depth accuracy and precision by experiments.

II. PSEUDO–DTOF DEPTH IMAGING

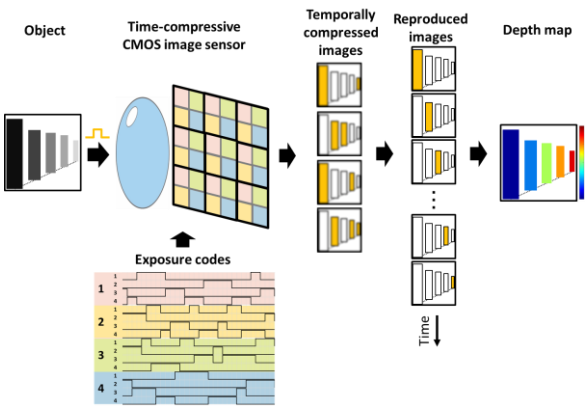


Fig. 1. Image acquisition and reconstruction flow

In our scheme, incident light signals are temporally compressed in the charge domain and reconstructed in three phases: pre-measurement, sensing, and signal reconstruction.

Firstly, we prepare exposure codes and measure the instrument response function (IRF) of the imaging optics and the image sensor in advance, which are used for signal reconstruction later.

In the sensing phase, the camera captures temporal signals of light emitted from a synchronized laser and reflected from objects. As shown in Fig. 1, the image sensor is composed of macro-pixels based on charge modulators. By applying exposure codes to the pixels

during the image shooting, multiple temporally compressed images (four in Fig. 1) are obtained at once.

In the reconstruction phase, the input optical waveforms for all subpixels are reproduced by solving the inverse problem based on the sparsity regularization. Thus, temporally sequential images or transient images of light are obtained. Subsequently, the temporal peak positions of the light waveform are detected with a quadratic curve fitting. Finally, they are converted to the object's depth using the speed of light.

a) Multi-tap macro-pixel CMOS image sensor

This sensor utilizes iToF-based charge modulators, which contribute to small pixel size or high-spatial-resolution imaging.

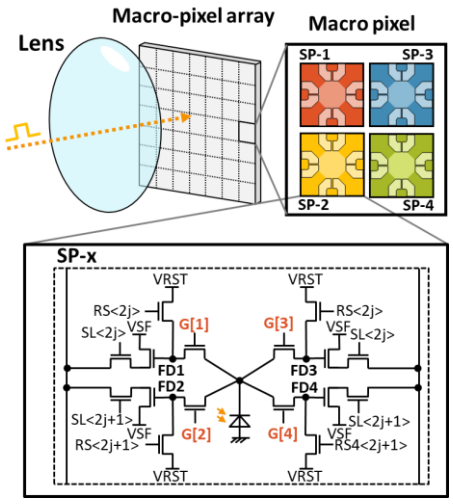


Fig. 2. Image sensor configuration.
One macro pixel is composed of 2×2 four-tap subpixels.

Figure 2 shows the pixel structure, which comprises an array of four sub-pixels (SPs) for each macro-pixel. Each sub-pixel is implemented by a four-tap LEFM charge modulator [3]. The LEFM charge modulator is composed of a photodiode and four sets of a charge transfer gate and a storage diode (tap). It outputs the integral value of a time-variant optical signal $\boldsymbol{g(t)}$ within a designated time window function $\boldsymbol{\omega_i(t)}$ as a pixel value $\boldsymbol{Q_i}$ (Eq. 1). Here, the index $\boldsymbol{i}$ identifies a tap.

$$Q_i = \int_0^{T_{exp}} g(t) \cdot \omega_i(t) dt \qquad \text{Eq. 1}$$

The high-speed operation of LEFM pixels allows for high temporal resolution. In our previous research, we confirmed that the sensor can be driven at a clock frequency of up to 303 MHz [4].

b) Compressive sensing

Compressive sensing [5] is an efficient sampling method that reconstructs more data points from fewer samples when the original signal is sparse. We employ this principle to reconstruct the entire incident light waveform from the compressed output images, resulting in dToF-like signals with high-depth accuracy and robustness to MPI, like the conventional dToF method. Note that compressed images are obtained in a single shot, which makes our sensor motion-artifact-free.

Table 1. Sensor architecture

Technology	0.11 μm FSI CIS
Chip size	7.0 mm^H×9.3 mm^V
Macro-pixel size	22.4 μm^H×22.4 μm^V
Effective sub-pixel count	212^H×188^V
Sub-pixel count per macro-pixel	2^H×2^V
Maximum exposure code length	256 bits
Maximum modulation frequency	303 MHz
Maximum clock frequency	607 MHz
Image readout frame rate	21 fps
Power consumption	2.8 W

Our sensor utilizes a coded exposure pattern that is a temporal series of random binary values to create time-compressed signals. When the exposure code is 0, no charge is transferred to a specific tap, and when it is 1, the charge is stored in the tap. Therefore, the number of charges accumulated in each tap corresponds to the correlation value between the input light signal and the coded exposure pattern, resulting in temporally compressed images. The exposure codes are repeatedly applied to increase the pixel value.

Time-compressive sensing and signal reconstruction are as follows. Consider the input signal is $\boldsymbol{x}$ and the corresponding measurement signal is $\boldsymbol{y}$, their relation can be expressed by the following linear equation:

$$\boldsymbol{y = Ax = (WH)Px} \qquad \text{Eq. 2}$$

Here, matrix $\boldsymbol{A}$ is a spatio-temporal observation matrix that includes the spatial IRF of the imaging optics $\boldsymbol{P}$, the exposure code $\boldsymbol{W}$ and the sensor's temporal IRF $\boldsymbol{H}$. The dimensions of $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{A}$ are N, M, and $M \times N$, respectively. In cases where $N > M$, i.e., the measured signal is lower in dimensionality than the original input signal, the signal is compressed. Retrieving the original input signal $\boldsymbol{x}$ from $\boldsymbol{y}$ is an ill-posed problem. However, if $\boldsymbol{x}$ is K-sparse, meaning that only K elements have non-zero values, we can determine the estimated solution for $\boldsymbol{x}$ by optimizing the L1 norm.

Based on this principle, the reflected light waveform $\boldsymbol{x}$ of all sub-pixels is reconstructed from the four compressed images $\boldsymbol{y}$ and the pre-measured observation matrix $\boldsymbol{A}$. In this process, total variation is minimized instead of the L1 norm as shown in Eq. 3. Here, $\boldsymbol{D}_i$ shows a spatio-temporal differential operator. This process is performed by TVAL3 [6], a compressed sensing solver, using an iterative method.

$$\hat{\boldsymbol{x}}^{(TV)} = \boldsymbol{\arg\min_x} \sum_i \|\boldsymbol{D}_i \boldsymbol{x}\|_1, \ \boldsymbol{s.t.}\ \boldsymbol{y = Ax} \qquad \text{Eq. 3}$$

Thus, transient images of light are reproduced, providing depth information. The depth accuracy and precision are dependent on the frame rate of the transient images. The temporal resolution of the reconstructed signal is determined by the reciprocal of

this frame rate, which is also equal to the minimum duration of the exposure code. For example, if our sensor operates at 303 MHz and uses a 32-bit exposure code, the temporal resolution is 3.3 ns (0.5 m in depth), and the measurable distance is 16 m.

To generate a depth map from the reconstructed images, the reconstructed waveform for each pixel is analyzed to determine the temporal peak position, which corresponds to the round-trip ToF. Finally, it is converted into a depth value using the speed of light.

III. IMPROVEMENT OF DEPTH RESOLUTION

So far, we have succeeded in ToF imaging at an operating clock frequency of 303 MHz [4]. This allowed the minimum time window with a duration of 3.3 ns, which corresponded to a 0.5 m resolution. Generally, higher depth resolution can be achieved simply by increasing the operating frequency of the charge modulator. However, driving the charge modulation at a higher frequency is highly dependent on the fabrication technology and modulator structure. It was challenging to achieve a minimum time window duration of less than 3.3 ns for our modulators. To overcome such restrictions, the following two approaches were applied to obtain higher depth resolution.

a) Sub-time-window shifting

Here, sub-time-window shifting of the exposure code is introduced to enhance the temporal resolution without changing the minimum time window duration, which is determined by the modulator design. To implement the sub-time-window shifting, we double the PLL clock frequency from 303 MHz to 607 MHz, while maintaining the minimum time window duration at 3.3 ns, equivalent to two clocks. Then, as shown in Figure 4, the falling and rising edges of the time window are shifted by one clock, which is half of the minimal time window. As a result, the temporal resolution is halved down to 1.75 ns.

Figure 5 shows the temporal sensor IRF measured

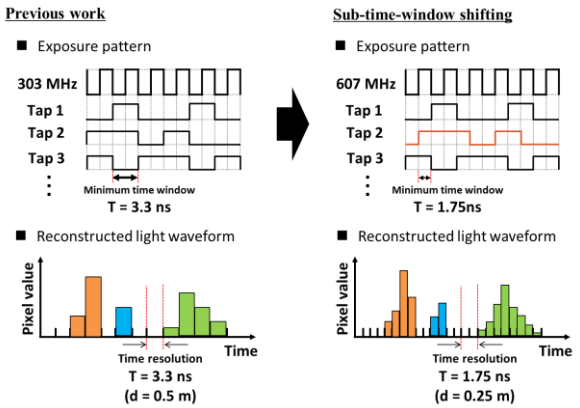


Fig. 4. Improvement of temporal resolution by sub-time-window shifting

when this sensor was operated at 607 MHz. The exposure code was a 32-bit random binary code, and a 445 nm laser with an FWHM of 60 ps was used. The results were averaged for each tap. It can be observed that charge modulation was successfully performed.

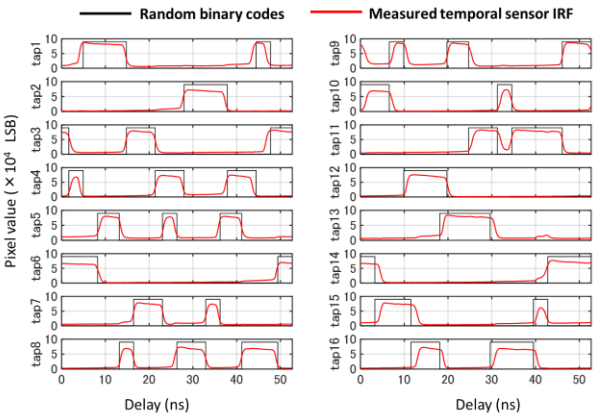


Fig. 5. Measured temporal sensor IRF for the random codes at 607 MHz

b) Oversampling

Another approach to improve the temporal resolution is oversampling the sensor's IRFs, i.e., increasing the number of data points of the pre-measured observation matrix used for reconstruction. Previously, the sensor response was sampled once per bit of the exposure code. However, in this study, we have sub-sampled 10 points per bit (10× oversampling), as depicted in Fig. 6. The temporal (or depth) resolution is then improved, resulting in a denser reconstructed light waveform. Since the signal waveforms can be obtained as in dToF imaging, we can perform fitting to refine the depth. After the peak position in the waveform is detected from the reproduced signal waveform, the five data points around the peak are considered in fitting with a quadratic curve.

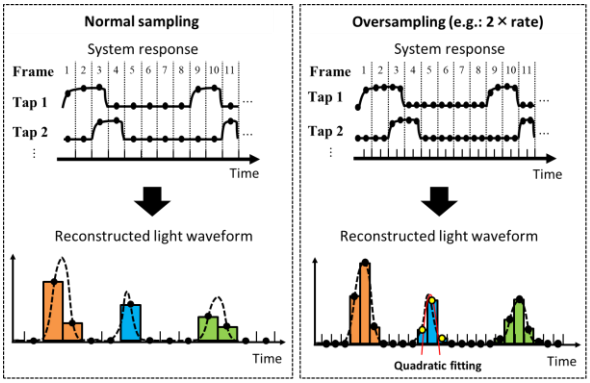


Fig. 6. Oversampling pre-measured sensor's IRF and fitting

The proposed method works better with oversampling, as denser light waveforms are utilized in fitting, resulting in more accurate depth estimation. Moreover, even if there are multiple peaks due to MPI, each peak can be distinguished and separately detected. Note that the depth resolution can be improved without changing the image sensor hardware or the total measurement time, although the reconstruction process takes longer.

IV. EXPERIMENTS

To evaluate the improvement of depth resolution after increasing the clock frequency from 303 MHz to 607 MHz, we compared the peaks separation performance under MPI conditions. In this experiment,

interference light was introduced by a weak diffusive plastic sheet placed in front of an objective panel with the letters "SU" as shown in Fig. 7a. The letters look blurry due to the diffuser. The distance between the two objects was set to 0.65 m and 0.40 m, and experiments were conducted with the sensor operating at 303 MHz and 607 MHz. A short-pulse semiconductor laser (PicoQuant LDH-IB-450-M-P, 443 nm) with a pulse width of 228 ps was used.

Figure 7b shows the reconstruction of optical waveforms at two clock frequencies. At 303 MHz, the interference light due to the diffuser was merged with the objective reflection for both clearances and could not be separated. However, when the clock frequency was increased to 607 MHz, the two reflections were completely separated for the 0.65 m clearance. For the 0.40 m clearance, although there was an overlap between the two signals, the object's reflection peak could be distinguished. These results show that the MPI is resolved in the 607 MHz operations.

Next, a ToF imaging experiment was conducted to verify the effectiveness of oversampling for a ratio of 10. Note that the time/depth resolutions at 303 MHz, 607 MHz, and 607 MHz with 10× sampling are 3.3 ns/0.5 m, 1.75 ns/ 0.25 m, and 0.175 ns/ 0.025 m respectively. A pulsed semiconductor laser with a wavelength of 660 nm and an FWHM of 2.5 ns was used. Figure 8 shows the configuration of the targets and the depth maps for each condition. 100-image averaging was applied to improve the SNR. Fig. 9 compares the reconstructed optical waveforms and their fitting curves with a quadratic approximation. The mean and standard deviation of the depth is quantitatively compared in Table 1, which demonstrates the effectiveness of the proposed method.

IV. CONCLUSION

In this paper, we demonstrated the concept, benefits, and implementation of pseudo-dToF imaging. Several techniques such as increasing clock frequency from 303 MHz to 607 MHz and oversampling have been applied to improve the depth resolution and estimated depth precision. While there are still some issues such as long processing time, vulnerability to ambient light and so on, the proposed method remains a promising technique for applications like autonomous vehicles, and robotics in challenging environments with MPI.

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] A. R. Ximenes, *et al.*, "A 256 × 256 45/65nm 3D-Stacked SPAD-Based Direct TOF Image Sensor for LiDAR Applications with Optical Polar Modulation for up to 18.6 dB Interference Suppression", ISSCC 2018, pp. 96–98.

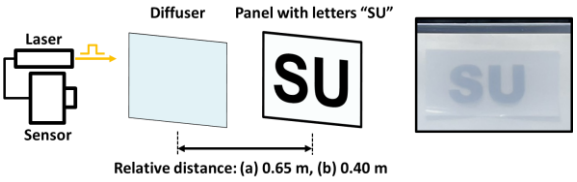
[2] Z. Zhao, *et al.*, "A Novel Imaging Method for Two-Tap Pulsed-Based Indirect Time-of-Flight Sensor," IEEE Sensors Journal, Vol. 23, No. 7, pp. 7017-7030, 2023.

[3] S. Kawahito, *et al.*, "CMOS Lock-In Pixel Image Sensors with Lateral Electric Field Control for Time Resolved Imaging," Proc. 2013 IISW, Vol. 361, 10-6.

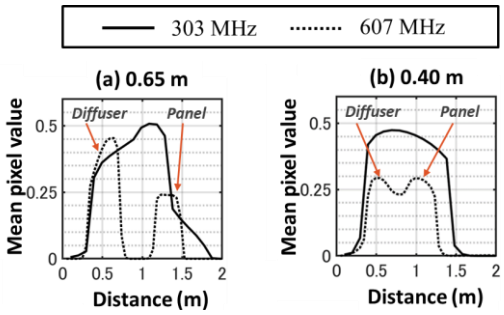
[4] K. Kagawa, *et al.*, "A dual mode 303 Megaframes per second charge domain time compressive computational CMOS image sensor," Sensors 2022, Vol. 22.5, 1953.

[5] R. Baranuik, "Compressive sensing [lecture notes]," IEEE Signal Processing Magasize, Vol. 24, No. 4, pp. 118-121, 2007.

[6] C. Li, *et al.*, "An efficient augmented Lagrangian method with applications to total variation minimization," Computational Optimization and Applications, Vol. 56, 2013.



a) Experiment setup for multi-path ToF imaging



b) Reconstructed light waveform at relative distances of 0.65 m and 0.40 m (Average of ROI 5×5 pixel values)

Fig. 7. MPI interference separation at 303 MHz and 607 MHz

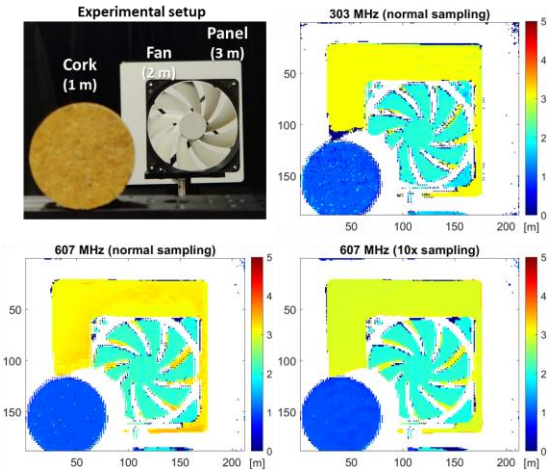


Fig. 8. ToF imaging experiment setup and depth maps

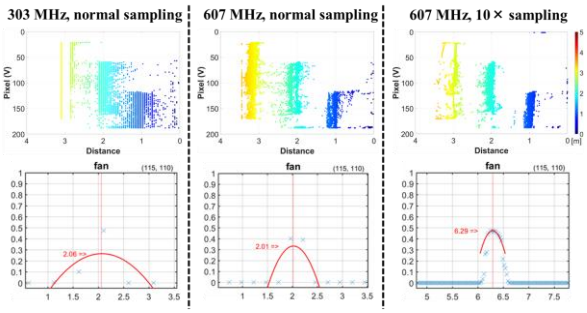


Fig. 9. Side view of depth maps and an example of peaks of reconstruction signal

Table 2. Mean and standard deviation of objects' depth (Average of 3×3 ROI pixels of 20 measurements)

	Real depth	303 MHz Normal sampling		607 MHz Normal sampling		607 MHz 10 × sampling	
		Mean [m]	Std [cm]	Mean [m]	Std [cm]	Mean [m]	Std [cm]
Cork	**1.00**	1.04	2.53	1.01	1.17	**1.00**	**0.94**
Fan	**2.00**	1.66	5.59	1.98	1.14	**1.99**	**0.72**
Panel	**3.00**	3.03	3.67	3.08	5.60	**3.00**	**0.57**