

Article

Resolving Multi-Path Interference in Compressive Time-of-Flight Depth Imaging with a Multi-Tap Macro-Pixel Computational CMOS Image Sensor

Masaya Horio ¹, Yu Feng ¹ , Tomoya Kokado ¹, Taishi Takasawa ², Keita Yasutomi ², Shoji Kawahito ², Takashi Komuro ³, Hajime Nagahara ⁴ and Keiichiro Kagawa ^{2,*}

- ¹ Graduate School of Integrated Science and Technology, Shizuoka University, Shizuoka 422-8011, Japan; mhori@idl.rie.shizuoka.ac.jp (M.H.); feng.yu.16@shizuoka.ac.jp (Y.F.); kokado.tomoya@mail.canon (T.K.)
- ² Research Institute of Electronics, Shizuoka University, Shizuoka 422-8011, Japan; ttakasawa@idl.rie.shizuoka.ac.jp (T.T.); kyasu@idl.rie.shizuoka.ac.jp (K.Y.); kawahito@idl.rie.shizuoka.ac.jp (S.K.)
- ³ Graduate School of Science and Engineering, Saitama University, Saitama 338-8570, Japan; komuro@mail.saitama-u.ac.jp
- ⁴ Institute of Datability Science, Osaka University, Osaka 565-0871, Japan; nagahara@ids.osaka-u.ac.jp
- * Correspondence: kagawa@idl.rie.shizuoka.ac.jp

Abstract: Multi-path interference causes depth errors in indirect time-of-flight (ToF) cameras. In this paper, resolving multi-path interference caused by surface reflections using a multi-tap macro-pixel computational CMOS image sensor is demonstrated. The imaging area is implemented by an array of macro-pixels composed of four subpixels embodied by a four-tap lateral electric field charge modulator (LEFM). This sensor can simultaneously acquire 16 images for different temporal shutters. This method can reproduce more than 16 images based on compressive sensing with multi-frequency shutters and sub-clock shifting. In simulations, an object was placed 16 m away from the sensor, and the depth of an interference object was varied from 1 to 32 m in 1 m steps. The two reflections were separated in two stages: coarse estimation based on a compressive sensing solver and refinement by a nonlinear search to investigate the potential of our sensor. Relative standard deviation (precision) and relative mean error (accuracy) were evaluated under the influence of photon shot noise. The proposed method was verified using a prototype multi-tap macro-pixel computational CMOS image sensor in single-path and dual-path situations. In the experiment, an acrylic plate was placed 1 m or 2 m and a mirror 9.3 m from the sensor.

Keywords: time-of-flight depth imaging; multi-path interference; multi-tap macro-pixel; charge modulators; compressive sensing; CMOS image sensor



Citation: Horio, M.; Feng, Y.; Kokado, T.; Takasawa, T.; Yasutomi, K.; Kawahito, S.; Komuro, T.; Nagahara, H.; Kagawa, K. Resolving Multi-Path Interference in Compressive Time-of-Flight Depth Imaging with a Multi-Tap Macro-Pixel Computational CMOS Image Sensor. *Sensors* **2022**, *22*, 2442. <https://doi.org/10.3390/s22072442>

Academic Editor: Lucio Pancheri

Received: 19 February 2022

Accepted: 19 March 2022

Published: 22 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Time-of-flight (ToF) [1] depth imaging is a technique for measuring the depth between a camera and objects based on the round-trip time of light emitted from the camera, assuming that the speed of light is constant. ToF has been applied to a variety of applications [2] such as autonomous driving, robot navigation, modeling of objects, and gesture recognition in entertainment due to the small device size, real-time measurement, and the ability to measure texture-less surfaces. However, ToF has a disadvantage in that the measurement accuracy is degraded by multi-path interference [3,4]. There are four major types of multi-path interference: (1) surface reflections from transparent objects, (2) multiple reflections between objects, (3) sub-surface scattering, and (4) volumetric scattering caused by bulky weak scattering media such as smoke or fog. In this paper, we focus on surface reflections.

There are two types of ToF image sensors: direct (dToF) [5] and indirect (iToF) [6]. In dToF, impulse light with a duration of nanoseconds is mostly used, and the temporal reflected light waveform is measured pixel by pixel by using a single-photon avalanche

diode (SPAD) [7]. Comparison of dToF and iToF is summarized in Table 1. In iToF, on the other hand, the inner products or correlations of the reflected light waveform with multiple time-window functions are measured. Then, the temporal or phase delay is calculated. To perform the inner-product operation, charge modulators are utilized. iToF is further classified into amplitude modulation continuous wave (AMCW) iToF [8] and pulse iToF [9]. AMCW iToF uses sinusoidally modulated light, and pulse iToF uses rectangular pulsed light with a duration of tens of nanoseconds. dToF provides the genuine waveform of the reflected light for many temporal sampling points (typically hundreds of points or more) with a high temporal resolution by using high-precision time-to-digital converters (TDCs). Although separation of multi-path components is easier by dToF than by iToF because the whole waveform of the reflected light is obtained, large digital circuits are required to build the histograms of photon arrival times. iToF is advantageous for achieving a small sensor size because signal detection is performed in the charge domain in pixels, and large digital or analog circuits are required for detection. However, the number of temporal or phase sampling points of the detected signal is only a few, typically three or four. Therefore, multi-path component separation is more difficult in iToF than in dToF.

Table 1. Comparison of dToF and iToF.

	dToF	iToF
Detector	SPAD	Charge modulator
Pixel size	Relatively large	Small
Pixel readout circuits	Time-to-digital converter and histogram builder	The same as ordinary CMOS image sensors (pixel source followers, column correlated double sampling circuits, and analog-to-digital converters)
Immunity to multi-path interference	Good	No

To decompose or compensate for the surface-reflection-type multi-path interference, methods that use multiple modulation frequencies or delays have been demonstrated [10–14]. However, these methods are vulnerable to motion artifacts, since multiple images with some scanning are necessary. Recently, deep-learning-based methods have been emerging [15,16]. Reference [16] proposed Deep ToF, which uses a single depth image taken by an ordinary AMCW iToF camera. Then, the multi-path interference is corrected by deep learning in real time. This method uses information about neighboring pixels or the scene to correct multi-path interference caused by multiple reflections. However, it could be difficult to resolve multi-path interference when neighboring pixels do not provide any cue to correct the multi-path interference. This problem could arise for the case of surface-reflection-type multi-path interference. To achieve motion-artifact-free single-shot imaging acquisition in ToF, specially designed image sensors are necessary.

In our previous study [17,18], we have proposed a method for multi-path component separation using multi-aperture-based temporally compressive pulse iToF. This method requires no large on-chip digital circuits, and the number of temporal sampling points is much more than that of conventional iToF, by taking advantage of the sparsity of the ToF signal. Figure 1 shows a method for separating multi-path components based on compressive sensing [19–22]. In this method, 15 inner products of the reflected pulse light with binary random shutters are acquired at the same time. Then, 32-bin reflected light histograms are reproduced. Thus, the drawback of iToF, i.e., the limited number of sampling points, is alleviated, and multiple reflected light paths are separated. However, this method requires a custom lens array, so it is not easy to reconfigure the optics. In addition, it is necessary to compensate for the parallax among the lenses. In the framework in [18], the depth resolution is the same as the width of the minimum time window, although sub-window resolution is achieved in ordinary iToF cameras.

We have proposed a multi-tap macro-pixel-based compressive ultra-high-speed CMOS image sensor [23]. This image sensor can implement multi-frequency shutters and sub-clock

shifting. Such features are useful for multi-path component separation in a single shot. By using a macro-pixel structure instead of a multi-aperture structure, it is possible to capture images using ordinary single-aperture lenses so that there is no disparity problem. In this paper, we evaluate the potential performance of separating dual-path components based on the multi-tap macro-pixel CMOS image sensor with multiple modulation frequencies and sub-clock shifting by simulation. Then, we demonstrate dual-path component separation in temporally compressive pulse iToF with a prototype image sensor. The depth estimation is performed in two steps: (1) separation of multiple peaks using a compressive sensing solver TVAL3 [24,25] and (2) refining the depths with sub-clock accuracy by fitting the result obtained at step 1 by a nonlinear optimization method [26]. Step 2 enables sub-clock-resolution depth estimation, which was not achieved in our previous study. Although this method is not suitable for real-time processing, the purpose of this study is to investigate the potential performance of the proposed scheme.

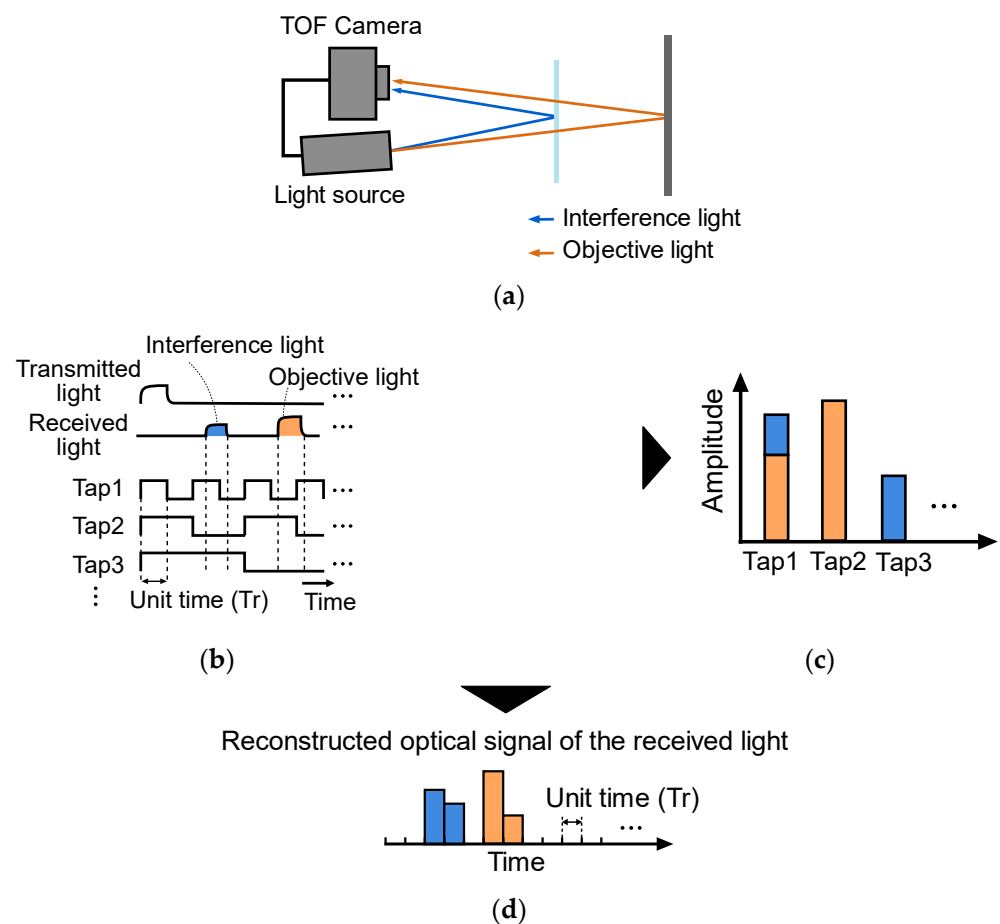


Figure 1. Separation of multi-path interference components: (a) optical setup that includes a transparent object causing multi-path interference, (b) compression of optical signals with temporal shutters, (c) compressed signals, and (d) reconstructed optical signal.

In Section 2, the multi-tap macro-pixel compressive image sensor architecture and depth imaging based on the temporally compressive pulse iToF are described. Section 3 shows system modeling and performance evaluation by simulation. In Section 4, dual-path component separation is demonstrated. In the experiment, a transparent acrylic plate was placed between the camera and an objective mirror to introduce multi-path interference. Section 5 discusses the issues of the proposed method. Section 6 concludes this paper.

2. Temporally Compressive Time-of-Flight Depth Imaging

2.1. Multi-Tap Macro-Pixel Computational CMOS Image Sensor

Figure 2 shows the architecture of a multi-tap macro-pixel computational CMOS image sensor. The sensor is mainly composed of a shutter controller, a clock tree, a lateral electric field charge modulator (LEFM) [27,28] driver, an array of macro-pixels, and readout circuits (RDOs). The specifications are summarized in Table 2.

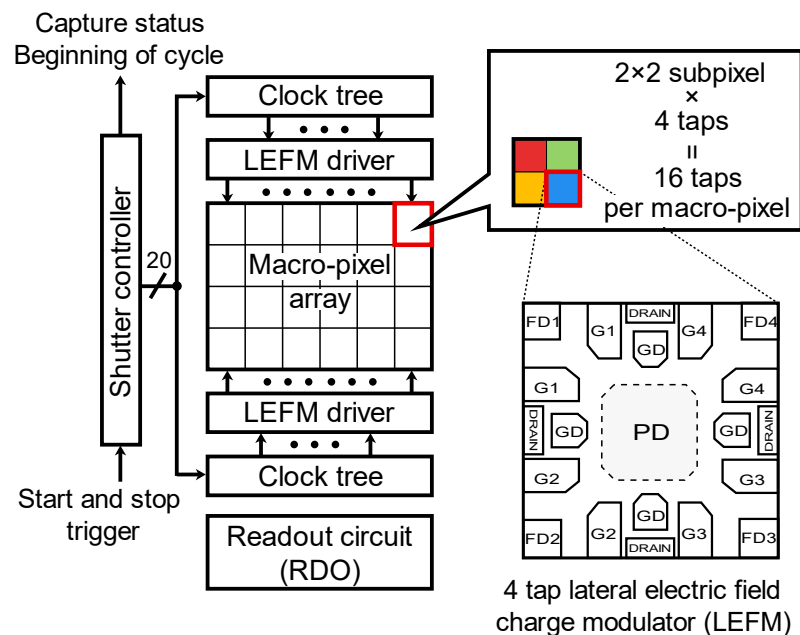


Figure 2. Structure of multi-tap macro-pixel computational CMOS image sensor. Each macro-pixel is composed of four subpixels. The subpixel is implemented by a four-tap LEFM with charge drain.

Table 2. Specifications of multi-tap macro-pixel computational CMOS image sensor.

Technology	0.11 μm CMOS image sensor process
Chip size	7.0 mm \times 9.3 mm
Valid subpixels	134 \times 110
Subpixel pitch	22.4 μm \times 22.4 μm
Subpixel count per macro-pixel	2 \times 2
Tap count per subpixel	4
Shutter length per tap	8 to 256 bits by 8 bits

The sensor was fabricated in a 0.11 μm CMOS image sensor process. A macro-pixel is composed of 2 \times 2 subpixels. Each subpixel is implemented by a four-tap LEFM, which has four storages (FD1-4). Charge transfer to the FDs and charge drain is controlled by five gate signals (G1-4 and GD). The shutter pattern of each gate is programmable. The shutter length per tap can be set to 8–256 bits in steps of 8 bits. Here, the shutter length is defined as the length of binary shutters and represented by bits. In Figure 1b, the shutter code in a unit time is given by one bit.

This sensor compresses images when the shutter length is longer than 16 bits (=the number of the total taps in the macro-pixel or the number of measured signals per macro-pixel). The mathematical representation of compressive sensing is briefly mentioned below. When we consider an N -dimensional column vector x , an $M \times N$ matrix A , and an M -dimensional column vector y , the relationship is denoted by

$$y = Ax. \quad (1)$$

Note that x is the original input signal, and y is the signal measured through a known measurement matrix A . When $N > M$, the original signal x is compressed. We can reproduce the original input signal x by solving the inverse problem from y and A based on a sparsity constraint.

Figure 3 shows the imaging process of a multi-tap macro-pixel computational CMOS image sensor. First, each tap detects the optical signal from the objects using different temporal shutters. Then, different temporally compressed images are obtained. After solving the inverse problem from the temporally compressed images and shutter patterns based on the sparsity, the time-sequential images are reproduced.

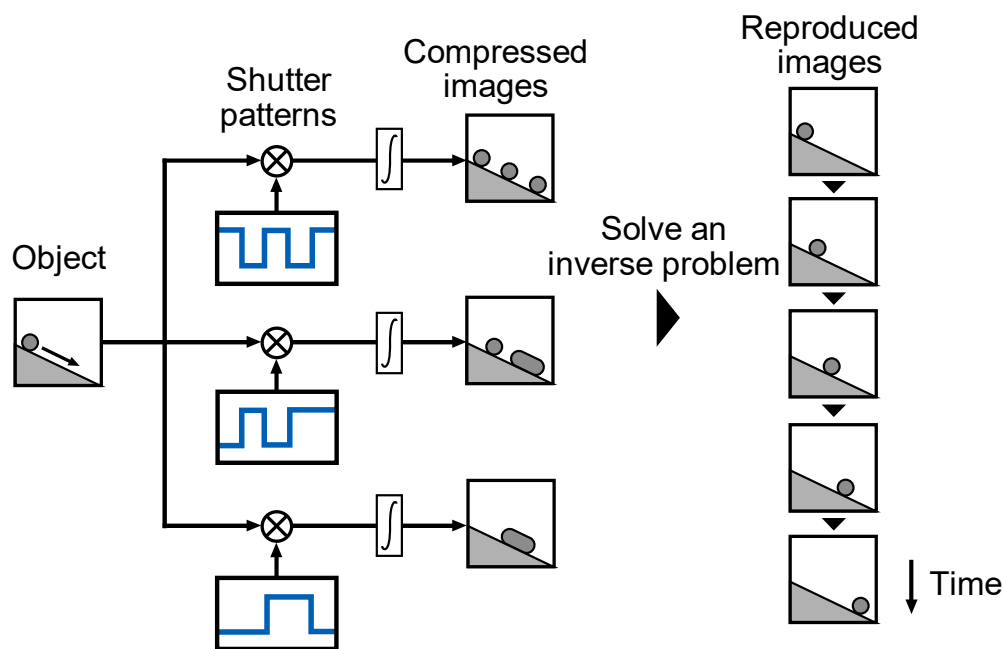


Figure 3. Image acquisition and reproduction flow of temporally compressive imaging. The object images are compressed in time with the temporal shutter patterns. Then, sequential images are reproduced by solving an inverse problem based on sparsity constraint.

For simplicity, the image acquisition and reproduction are explained based on a single pixel; i.e., x is a temporal waveform of the received light for a single pixel. The coded shutter is represented by the measurement matrix A in compressive sensing. It is assumed that A is composed of M sets of temporal shutters given by the $M \times N$ -dimensional matrix. Then, Equation (1) is rewritten as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{M1} & \cdots & a_{MN} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}, \tag{2}$$

where each row of A describes one temporal sequence of a shutter. The total number of taps is the same as the number of taps in a subpixel times the number of subpixels in a macro-pixel, namely $M = 4 \times 4 = 16$. If the original input signal x is K -sparse, which means only K elements have non-zero values and all the other elements are zero, x is estimated by the following optimization based on the l_0 -norm:

$$\hat{x}^{(0)} = \underset{x}{\operatorname{argmin}} \|x\|_0 \text{ subject to } y = Ax. \tag{3}$$

Here, $\|x\|_m$ shows l_m -norm of x . Note that l_0 -norm is defined as the number of the non-zero elements in x . However, solving this problem is hard because this is a huge combinatorial problem. Therefore, l_1 -norm minimization is used instead, as follows:

$$\hat{x}^{(1)} = \underset{x}{\operatorname{argmin}} \|x\|_1 \text{ subject to } y = Ax. \tag{4}$$

Total variation (TV) minimization is widely used to solve l_1 -norm minimization, where the original input signal x is estimated by

$$\hat{x}^{(TV)} = \underset{x}{\operatorname{argmin}} \sum_i \|D_i x\|_1 \text{ subject to } y = Ax. \tag{5}$$

Here, D_i is the differential operator that subtracts an adjacent element from the i -th element.

2.2. Modeling of Multi-Path Interference in ToF

The detected light waveform $g(t)$ at the ToF camera, shown in Figure 4c, is modeled as follows:

$$g(t) = L(t) * f(t), \tag{6}$$

where $f(t)$ (Figure 4a) is the scene response function, $L(t)$ (Figure 4b) is the waveform of the total system response, and $*$ is the convolution operator. In the multi-path scenario, $f(t)$ is modeled as

$$f(t) = \sum_{i=1}^K a_i \delta\left(t - \frac{2d_i}{c}\right), \tag{7}$$

where a_i is the amplitude and d_i is the depth of the i -th reflection, c is the speed of light, K is the total number of reflections, and $\delta(x)$ is the δ -function. $L(t)$ is represented by the convolution of the light source waveform $L_0(t)$ (Figure 4d) and the sensor response $h_s(t)$ (Figure 4e):

$$L(t) = L_0(t) * h_s(t). \tag{8}$$

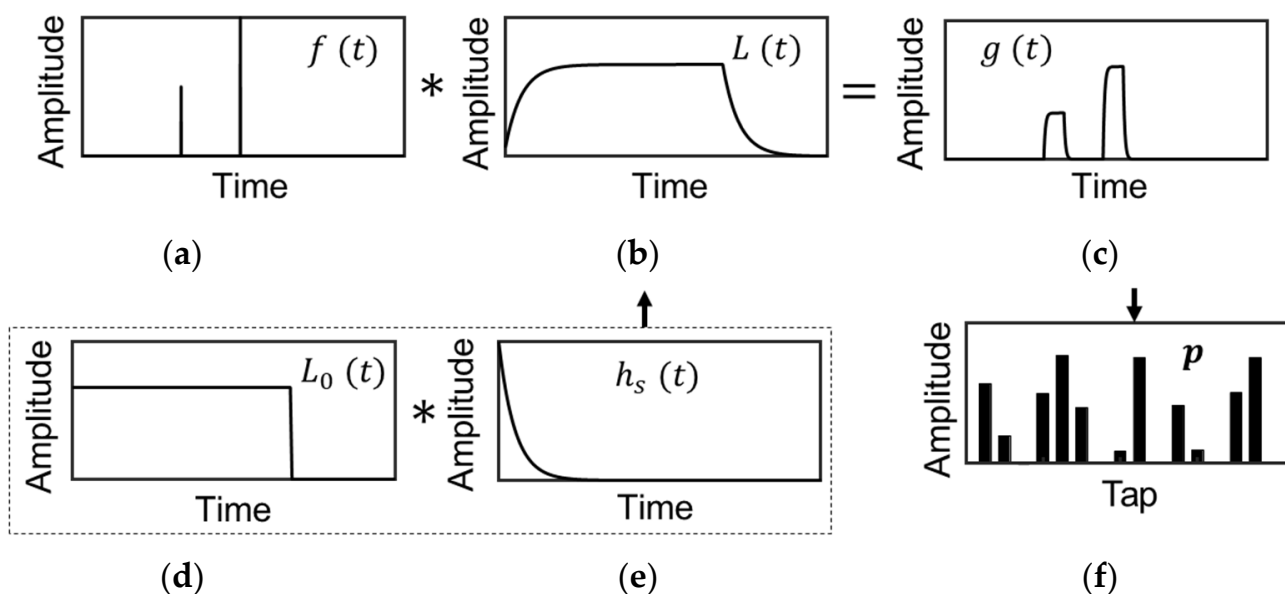


Figure 4. Modeling of multi-path interference in temporally compressive pulse iToF: (a) Scene response $f(t)$. (b) Total system response $L(t)$. (c) Detected light waveform $g(t)$. (d) Light source waveform $L_0(t)$. (e) Sensor response $h_s(t)$. (f) Sensor output p .

The sensor response $h_s(t)$ is further represented by the convolution of the response of the charge modulator $h_{s1}(t)$ and the photodiode response $h_{s2}(t)$. When either of these responses is dominant, $h_s(t)$ can be approximated by the first-order delay:

$$h_s(t) = h_{s1}(t) * h_{s2}(t) \sim \exp\left(-\frac{t}{\tau}\right), \quad (9)$$

where τ is the time constant of the modulator or photodiode. Thus, the detected light waveform $g(t)$ at the ToF camera is shown in Figure 4c and is modeled as

$$g(t) = L(t) * f(t) \sim L_0(t) * \exp\left(-\frac{t}{\tau}\right) * \sum_{i=1}^K a_i \delta\left(t - \frac{2d_i}{c}\right). \quad (10)$$

The image sensor puts out M compressed signals $\{p_m\}$ ($m = 1, \dots, M$) (Figure 4f), which is also denoted by a vector representation, $\mathbf{p} = (p_1, \dots, p_M)^T$. Note that \mathbf{x}^T means the transpose of a vector, \mathbf{x} . When the shutter pattern for the m -th tap is written as $w_m(t)$, the detected signal p_m is given by the inner product or correlation between $w_m(t)$ and $g(t)$:

$$p_m = \int w_m(t)g(t)dt. \quad (11)$$

2.3. Selection of Exposure Patterns

As mentioned in Section 2.1, 16 different shutters are applied to the multi-tap macro-pixel computational CMOS image sensor. In this paper, 32-bit shutter patterns shown in Figure 5 are used. The shutter patterns of subpixel 1 are shifted by half a clock from those of subpixel 2. Time windows for slower frequencies can increase the depth range, while high-frequency time windows improve the depth resolution. By using half-clock shifting, the depth resolution can be improved by a factor of two without shortening the width of the minimum time window or increasing the operating frequency of the sensor.

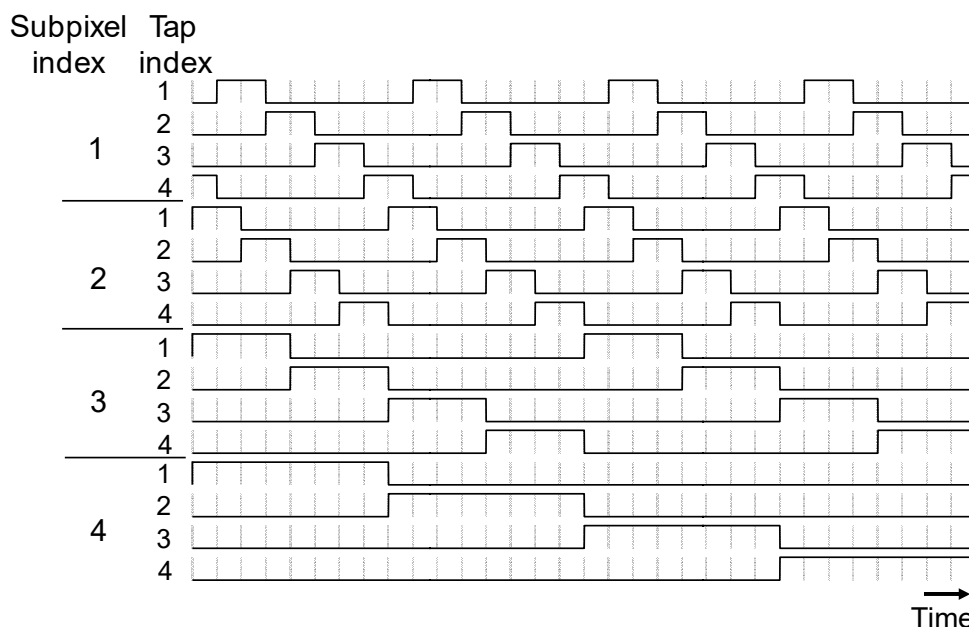


Figure 5. Temporally coded shutters (shutter length of 32 bits) with multiple frequencies and sub-clock shifting. The shutters of subpixel 1 are a half-clock-shifted version of those of subpixel 2.

When ordinary four-tap pulse iToF images sensors [29] are used for the dual-path scene, the shutter patterns only for one frequency can be applied at once, e.g., the shutter patterns for subpixel 4 in Figure 5. They cannot resolve the multi-path interference if

two reflections are detected in the same time windows. It is possible for them to emulate the shutter patterns in Figure 5. However, image acquisition should be performed four times for each set of the shutter patterns for each subpixel. Thus, this implementation will suffer from motion artifacts. The benefit of our multi-tap macro-pixel computational CMOS image sensor is single-shot image acquisition. A drawback is that the macro-pixel is larger in size than ordinary four-tap pixels.

2.4. Solving the Inverse Problem and Depth Refinement

The depth estimation process is pixelwise and composed of two stages. We assume that there are two reflections: objective light and interference light. This algorithm can be easily extended to more reflections.

First, we reconstruct the scene response function denoted by \hat{x} from the sensor output \mathbf{p} , which is a measured version of \mathbf{y} , and the measurement matrix \mathbf{A} . x is a discrete version of $f(t)$, and \mathbf{A} describes the shutter patterns $\{w_m(t)\}$ and the total system response $L(t)$. Next, we find peaks in \hat{x} and obtain K peak depths $\{d_i\}$ and their amplitudes $\{a_i\}$ ($i = 1, \dots, K$). When two adjacent peaks are too close, only one peak is found. To overcome this problem, we add two peaks at both ends of the highest peak that is most likely to include the objective light. Their amplitudes are half of the peak. Then, for all combinations of two arbitrary peaks among the $K + 2$ peaks, their depths and amplitudes are refined by a nonlinear search algorithm. Refinement is performed by minimizing the following evaluation function for two peaks i and j ($i, j = 1, \dots, K + 2, i \neq j$):

$$E\left(a = \frac{a_j}{a_i}, d_i, d_j\right) = \left| F(a, d_i, d_j) - \frac{\mathbf{p}}{\sum_{m=1}^M p_m} \right|^2. \quad (12)$$

Here, the amplitudes of the sensor output are normalized. F is the forward problem function and provides a simulated sensor output for a , d_i , and d_j , where a is the ratio of the two amplitudes. The sensor output is normalized by the sum of the elements, i.e., $\frac{(y_1, \dots, y_M)^T}{\sum_{m=1}^M y_m}$. The set of the two peaks that gives the minimal evaluation value is selected as a solution.

3. Simulation

3.1. Simulation Method

The simulation flow is shown in Figure 6. For simplicity, only photon shot noise is considered, whereas sensor random noise is neglected. The shot noise is given as follows: First, we determine the total number of photons for the M taps, which is given by N_{op} , to define the intensity of the reflected light or signal level. The sensor output \mathbf{p} is scaled to make $\sum_{m=1}^M p_m$ equal to N_{op} . Then, shot noise is given to each scaled p_m based on the Poisson distribution to generate \mathbf{p}' electrons. The depth and amplitude are estimated from \mathbf{p}' by the method explained above. This process is repeated R times to determine the precision and accuracy of the estimated depths, which are evaluated by using the relative standard deviation (RSD) and relative mean error (or simply error). Note that RSD and error refer to precision (or uncertainty) and accuracy (or nonlinearity), respectively. We compare these indexes for different N_{op} 's and a 's.

As for the multi-path interference, single-path and dual-path scenarios are simulated. The single-path situation is essential in the depth estimation. It is necessary to test if the proposed algorithm correctly works for this simplest case. The resolution of multi-path interference in the dual-path situation is the main issue in this study. It is expected that the estimated depth of the objective light becomes erroneous as the number of photons decreases or the amplitude of the interference light increases. With the depth of the objective light fixed, the dependency of the estimated depths for both objective and interference light on the depth of the interference light is investigated.

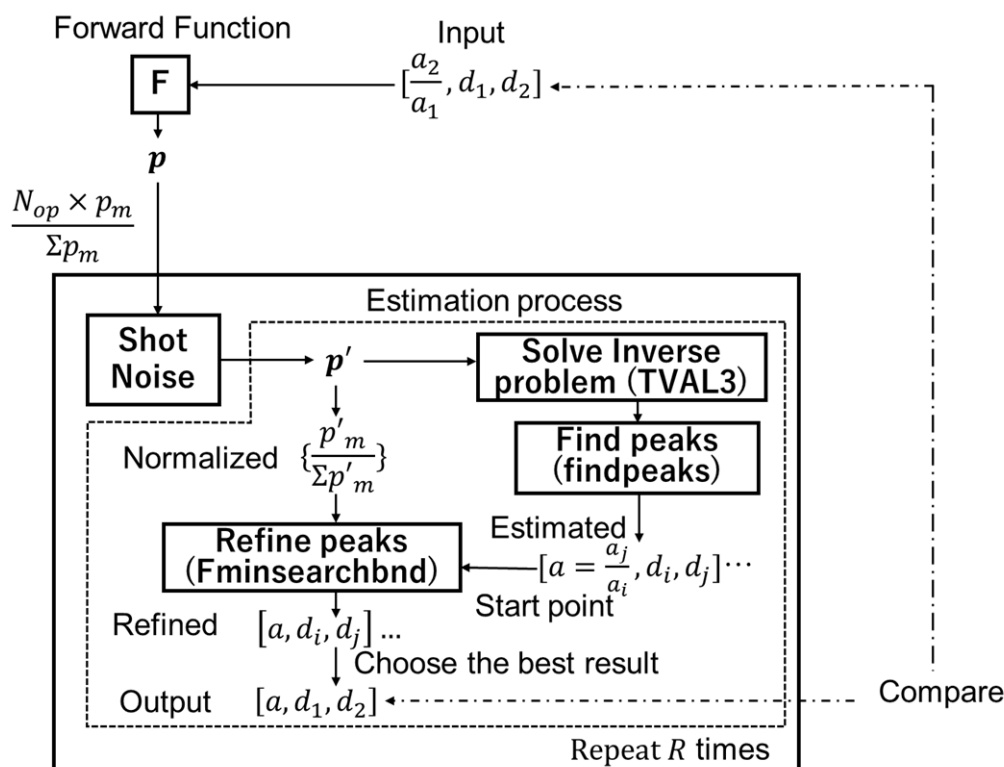


Figure 6. Simulation flow. The forward function generates a pixel value vector p from the depths of the objective and interference light and the amplitude ratio. These parameters are estimated in two stages from a noisy pixel value vector p' with shot noise.

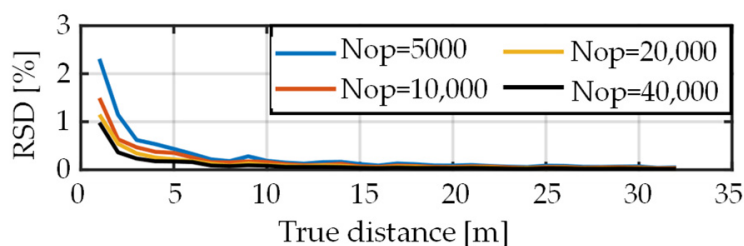
For the number of photons, two different cases are discussed. In reality, the number of the received photons at the ToF image sensor decays in inverse proportion to the square of the depth of the object. In the simplest case, this decay is not considered to investigate the natural behavior of the proposed algorithm. Then, the decay of the received photon number is considered.

The simulation conditions were as follows: The clock frequency for the shutter generator was 73 MHz. Therefore, the unit time was 13.7 ns which corresponded to approximately 2 m in depth. The measurable depth range was 32.64 m for the shutter length of 32 bits. The time step in the simulations was 0.05 ns (≈ 0.75 cm in depth). The time constant τ of the system response in Equation (9) was set to 1 ns. The duration of the laser pulse was as long as the shortest time window, namely 13.7 ns. R was set to 100.

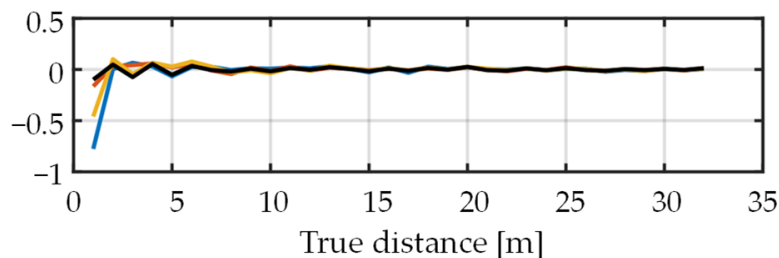
To generate the shot noise based on the Poisson distribution, the MATLAB function `imnoise` was used. TVAL3 was used as a compressive sensing solver [24,25]. The parameters of TVAL3 were as follows: `nonneg` was true, μ was 2^4 , β was 2^{-2} , μ_0 was 2^1 , β_0 was 2^{-3} , and other parameters were the default values. To find peaks, `findpeaks` in MATLAB Signal Processing Toolbox was used. To refine the peak positions, `fminsearchbnd` [26] was used to apply non-negative constraints.

3.2. Single Path

First, we simulated single paths (with no interference light) for the following conditions: N_{op} was 5000, 10,000, 20,000, or 40,000. The amplitude ratio a was 1. Depth-dependent decay of the number of photons was not considered. The depth d was varied from 1 to 32 m in steps of 1 m. The simulation results in Figure 7 show that depth errors were very small over the range. It was confirmed that the proposed method worked normally even for the single-path scenario.



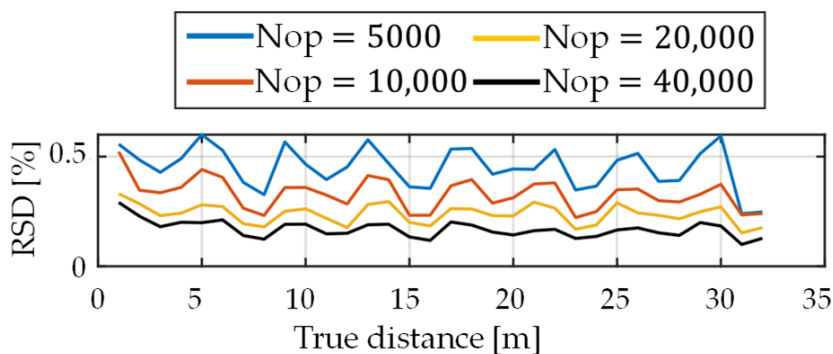
(a)



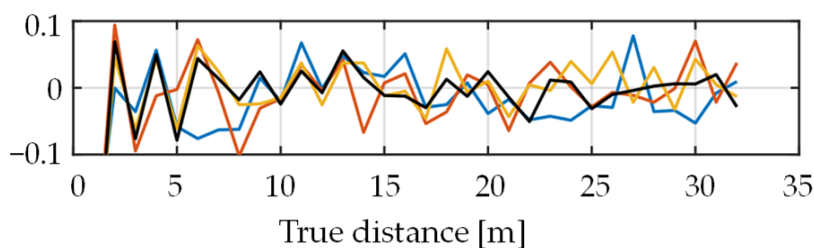
(b)

Figure 7. Results of single path simulation when depth-dependent decay of the number of photons is not considered. (a) Relative standard deviation and (b) relative mean error of the estimated depth.

Then, simulation was conducted for a realistic case considering the depth-dependent decay of light intensity. Figure 8 shows the simulation results. As N_{op} decreased, RSD increased because the shot noise relatively increased. The waveforms of RSD were oscillating a little, but mostly constant. No nonlinearity was observed, as shown in Figure 8b.



(a)



(b)

Figure 8. Results of single path simulation when depth-dependent decay of the number of photons is considered. (a) Relative standard deviation and (b) relative mean error of the estimated depth.

3.3. Dual-Path

In the dual-path scenario, two situations were considered. In Type A, the total photon number was changed while the amplitude was fixed. In Type B, the amplitude was changed with the total photon number fixed. All the conditions are shown in Table 3. d_1 and a_1 are for the objective light. d_2 and a_2 are for the interference light. The reflection's depth d_1 and amplitude a_1 were always 16 m and 1, respectively. The interference depth d_2 was moved from 1 m to 32 m in 1 m steps. Because this is a numerical investigation, the interference light can come after the objective light.

Table 3. Conditions of dual-path simulation.

		Type A (Variable Total Photons)	Type B (Variable Interference Reflection Amplitude)
Shutter length			32 bits
Minimal time window duration			13.7 ns
Light source pulse width			13.7 ns
Number of total taps per macro-pixel			16
Number of total electrons (N_{op})		5000, 10,000, 20,000, 40,000	20,000
Amplitude	Objective	$a_1 = 1$	$a_1 = 1$
	Interference	$a_2 = 0.1$	$a_2 = 0.1, 0.5, 1.0$
Depth	Objective	$d_1 = 16$ m	$d_1 = 16$ m
	Interference	$d_2 = 1 - 32$ m	$d_2 = 1 - 32$ m

Firstly, simulations without the depth-dependent decay of the number of photons were performed. Figure 9 shows the results of the simulation. Both the relative error and RSD of d_1 (objective reflection) became large as N_{op} decreased or a_2 increased, especially when the objective light and interference light were merged. As shown in the results of Type A (Figure 9a–d), the absolute value of the relative mean error for different light intensities was <1.2% in both d_1 and d_2 for all given N_{op} 's except when d_2 was around 1 m or 16 m. The results of Type B are shown in Figure 9e–h. The absolute value of the relative mean error for different amplitudes of the interference light was <0.8% in both d_1 and d_2 except when d_2 was around 1 m, 16 m, or 26 m. The reason why the depth error at around 1 m is relatively large is that the relative errors become large as the distance becomes small. Other issues related to the large errors are discussed below.

Figure 10 shows the simulation results for Type B when the depth-dependent decay of photons was considered. N_{op} was adjusted to be 5000 when the depth was 4 m and the amplitude was 1, and it was in inverse proportion to the square of depth. RSD and error of d_1 became large as the contribution of the interference light increased. The behavior of RSD and error of d_2 is interesting. As the amplitude of the interference light becomes small, both RSD and error become large. At far distances, the nonlinearity of d_2 for $a_2 = 0.1$ is significantly large. This is possibly because d_2 becomes closer to $d_1 = 16$ m.

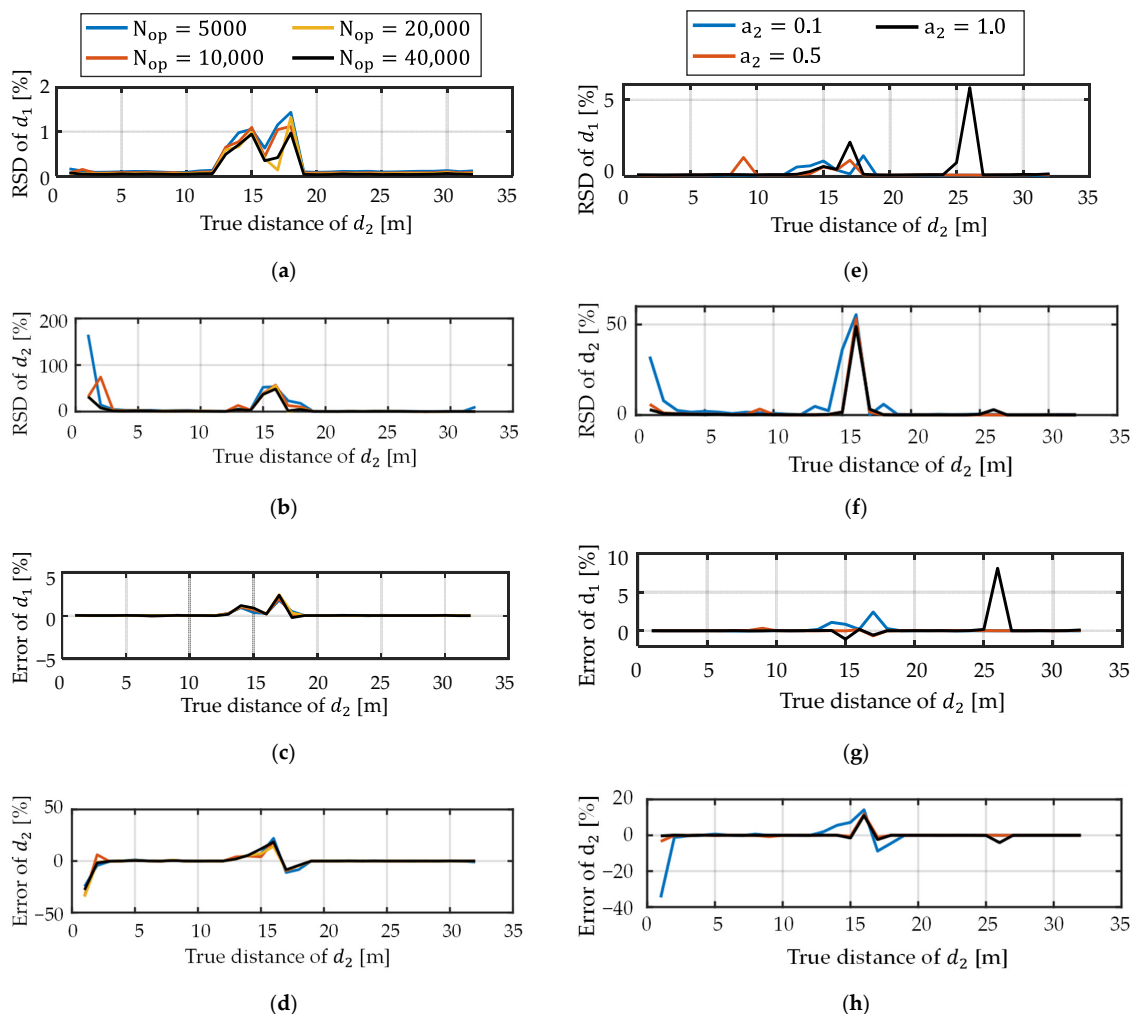


Figure 9. Results of dual-path simulations: (a–d) Type A. N_{op} was changed. (e–h) Type B. a_2 was changed. (a,e) Relative standard deviation of d_1 . (b,f) Relative standard deviation of d_2 . (c,g) Relative mean error of d_1 . (d,h) Relative mean error of d_2 .

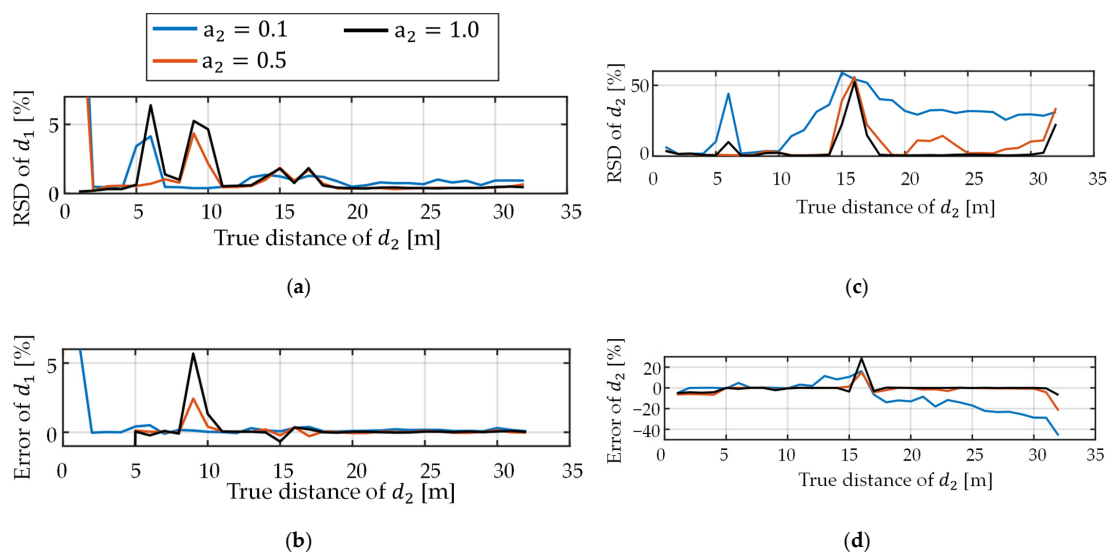


Figure 10. Results of dual-path simulations when depth-dependent decay of the number of photons was considered in Type B. N_{op} was set to 5000 at 4 m for the amplitude of 1. (a) Relative standard error and (b) relative mean error of d_1 . (c) Relative standard error and (d) relative mean error of d_2 .

4. Experiment Using a Multi-Tap Macro-Pixel Computational CMOS Image Sensor

4.1. Experimental System

In this section, the aforementioned method is demonstrated with a prototype multi-tap macro-pixel computational CMOS image sensor shown in Table 2 and Figure 2. The experimental system setup is shown in Figure 11. A semiconductor pulsed laser (Tama Electric, Hamamatsu, Japan, Model LDS-320, $\lambda = 850$ nm) emitted light with a duration of 21 ns toward the objective mirror and the acrylic plate through a mirror having a hole in the middle (holed mirror) to realize a semi-coaxial configuration. The distance between the acrylic plate and the holed mirror was 1 or 2 m. The objective mirror was placed 9.3 m away from the holed mirror. The 16 compressed images for the 16 32-bit shutters shown in Figure 5 were acquired in a single shot. The shutter controller frequency was set to 73 MHz. In this experiment, no imaging lens was used. The objective and interference light directly illuminated the image sensor. In solving the inverse problem, it was assumed that the subpixels in the same macro-pixel were uniformly illuminated. The saturation level of the sensor was 67558 LSB. The read noise of the sensor was 376 LSB.

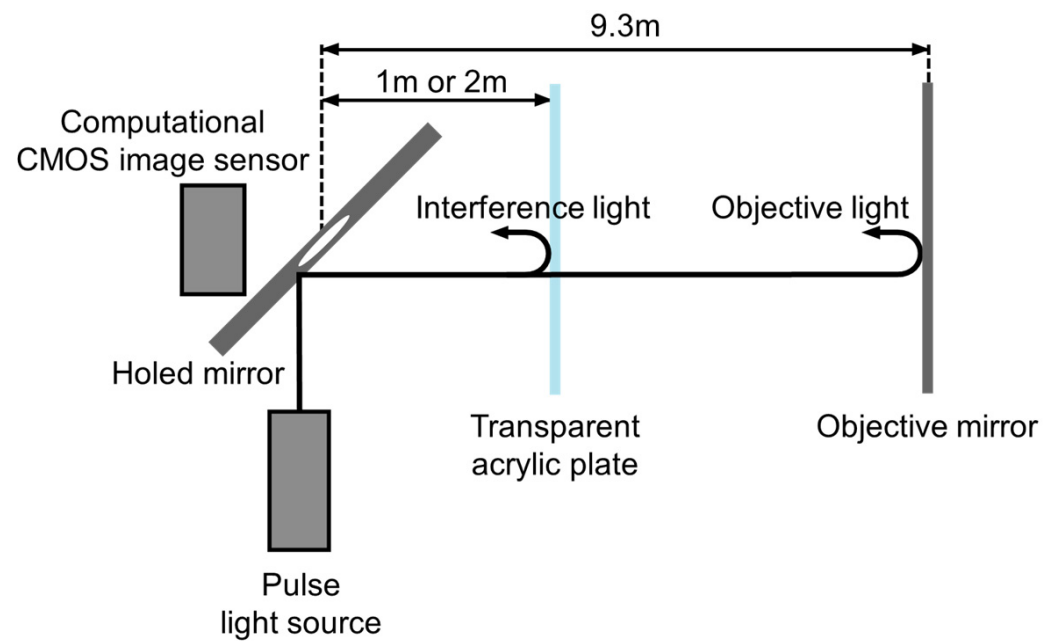


Figure 11. Experimental setup for dual-path ToF depth imaging. The objective mirror and transparent acrylic plate constitute dual-path interference.

The correlation of the total system response of the ToF camera system and the coded shutters, which was equivalent to the sensor output p , was measured before the ToF imaging. The measurement was performed while the emission timing delay of the pulsed laser t was scanned over the measurable range. The measured functions are denoted as follows:

$$p'_m(t) = \int w_m(\xi)L(\xi - t)d\xi. \quad (13)$$

The measured responses shown in Figure 10 were used in every step of the inverse problem solving. The measurement matrix A was generated by down-sampling $\{p'_m(t)\}$, and the sensor output in the refinement stage was given by the superposition of the sensor output in Figure 12 for different times of flight (=light pulse delays) such as

$$p_m = \int p'_m(t)f(t)dt. \quad (14)$$

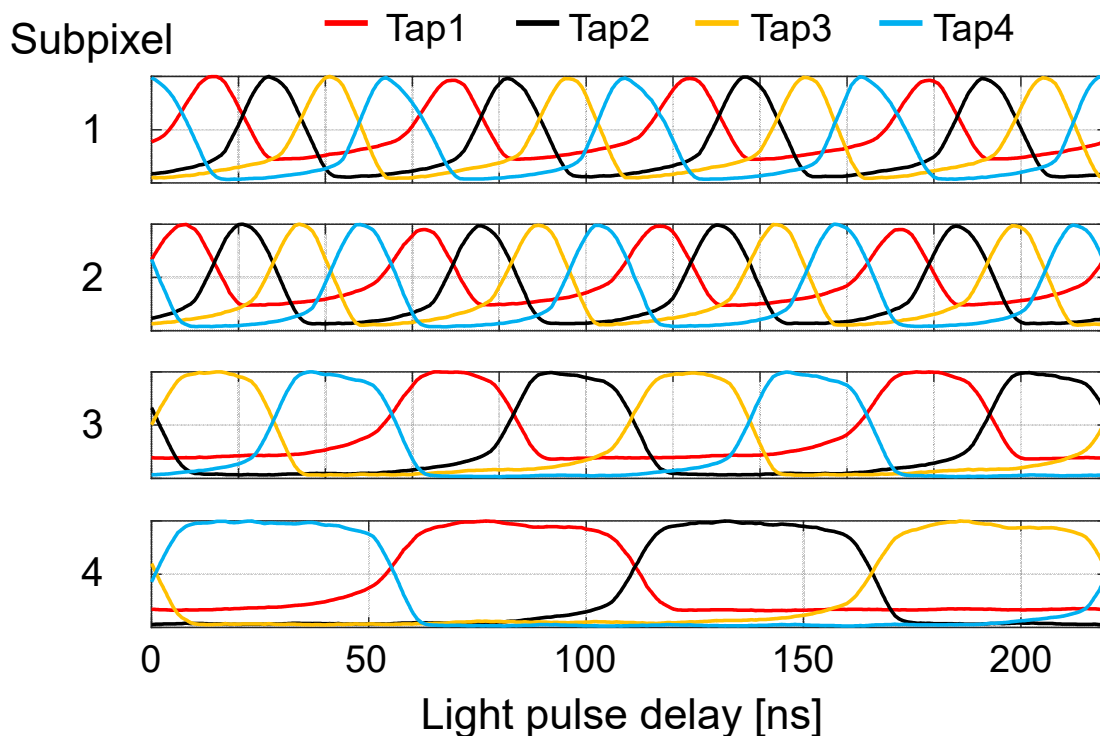


Figure 12. Measured sensor output versus light pulse delay. The waveforms are normalized by the maximum for each tap.

4.2. Measured and Processed Results

Figure 13 shows 16 temporally compressed images when the objective mirror and the acrylic plate were placed at 9.3 m and 1 m, respectively. For comparison, single-path scenes were also measured. Figure 13a,b show the captured images for the single-path scenes. The former is for only the acrylic plate, and the latter is for only the mirror. Figure 13c is for the dual-path situation. Two reflections from both acrylic plate and mirror are overlapped. Note that the interference light from the acrylic plate illuminated the bottom half of the image sensor. Figure 14a,b show the reproduced depth images (33×52 pixels) for the acrylic plate placed at 1 m and the mirror placed at 9.3 m, respectively. The whole image sensor was illuminated. Figure 14c,d shows the depths reproduced from the images in Figure 13 for the dual-path scenario.

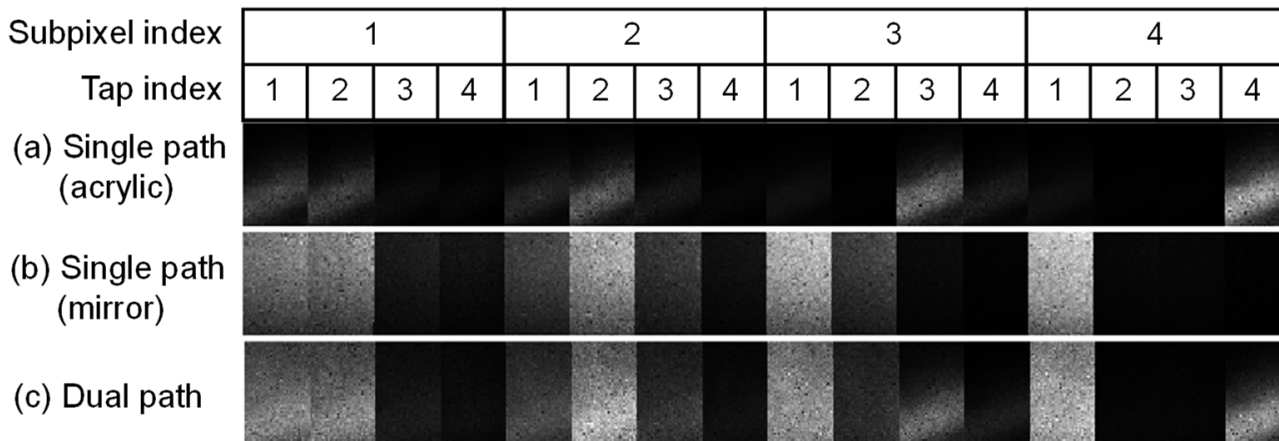


Figure 13. Captured images in the dual-path situation for the objective mirror and the acrylic plate placed at 9.3 m and 1 m, respectively. One hundred images are averaged.

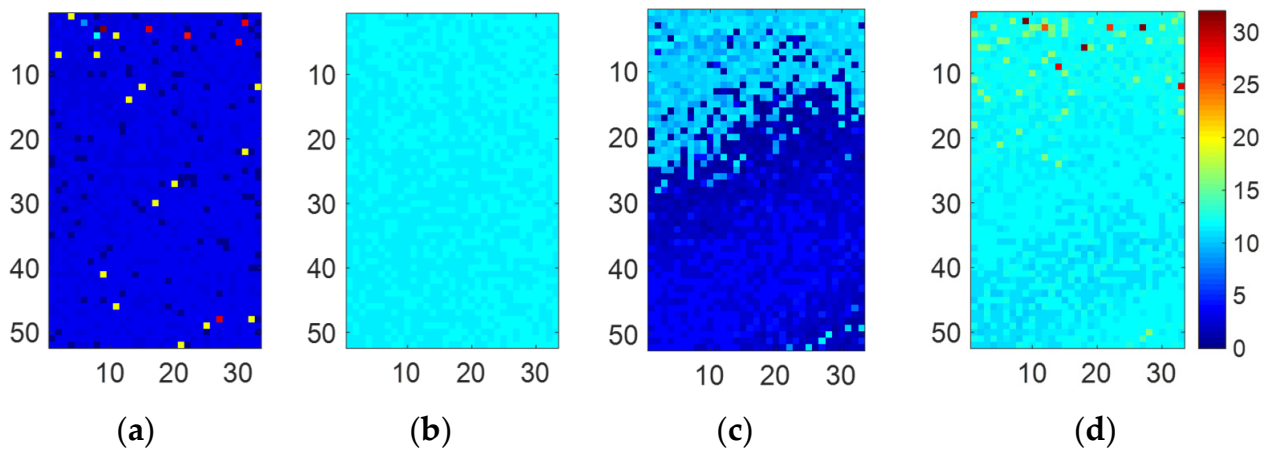


Figure 14. Measured depth images (33×52 pixels): (a) Acrylic plate placed at 1 m (single path). (b) Mirror placed at 9.3 m (single path). (c) Acrylic plate depth (dual-path). (d) Mirror depth (dual-path). In the dual-path situation (c,d), the depths are similar to those in the single-path situations (a,b).

The total processing time for the dual-path scene was 74 min. The percentages of the processing time for the first stage (coarse depth estimation by TVAL3) and the second stage (depth refinement by `fminsearchbnd`) were 0.7% and 99.3%, respectively. Specifications of the PC are as follows: CPU: Intel Core i7-9700 (3 GHz, eight cores), RAM: 16 GB. Signal processing was conducted on MATLAB Version 9.3.0.713570.

Figure 15 shows the depth histograms for the single-path and dual-path situations. Table 4 summarizes the mean peak positions of the histograms, where the pixels with pixel values more than half of the maximum pixel value are considered. As shown in the table, multiple depth components were separated with an absolute error less than 0.2 m compared with the depths acquired in the single-path scenario.

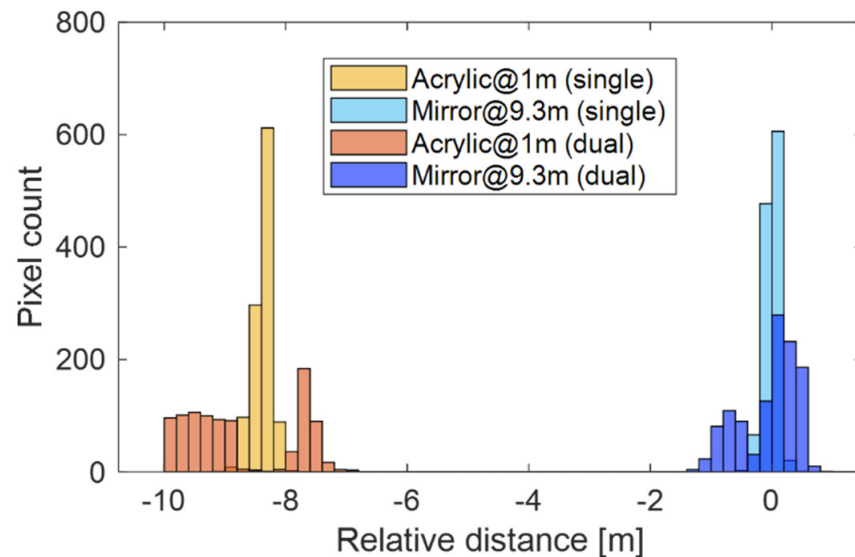


Figure 15. Depth histograms for the single-path and dual-path situations. The distance between the acrylic plate and the objective mirror was 8.3 m.

Table 4. Measured relative mean depths based on the single-path situation (m).

	Acrylic Plate at 1 m	Acrylic Plate at 2 m	Mirror at 9.3 m
Single path	−8.5172	−7.2929	0
Dual path	−8.3404	−7.3693	−0.1826

5. Limitations

In the simulation, one of the drawbacks of the proposed method is a relatively large depth error when the two light reflections are very close. In Figure 8, the depth error of the interference light is significantly large between 14 m and 18 m. This can happen when the two pulses are merged. The most effective way to improve the separation is to speed up the operating frequency of the image sensor to shorten the minimal time window duration. Because the sensor is still a prototype, it has problems such as low photosensitivity, large pixel size, and slow operating frequency.

In Figure 9e–h, a huge error is observed at 26 m. At this distance, d_1 and d_2 are estimated to be 17.96 m and 24.39 m when no shot noise is considered, although they should be 16 m and 26 m, respectively. Figure 16 compares the simulated sensor output p' for $d_1 = 16$ m, $d_2 = 26$ m, and $d_1 = 17.96$ m, $d_2 = 24.39$ m. The signals are similar in terms of the peak positions. Therefore, the solution could converge to a local minimum. A depth refinement method or better shutter patterns that can find the global minimum should be explored.

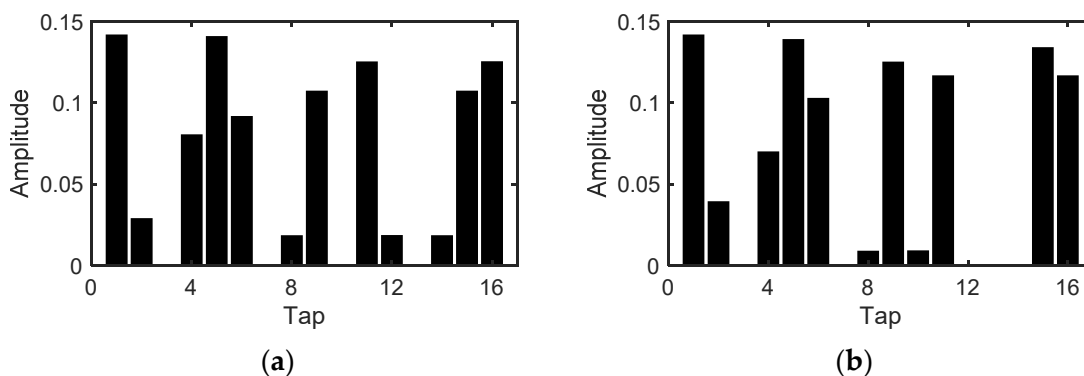


Figure 16. Simulated sensor output, p' without shot noise: (a) $d_1 = 16$ m, $d_2 = 26$ m. (b) $d_1 = 17.96$ m, $d_2 = 24.39$ m. They look very similar.

The number of resolvable paths is of great concern. Basically, multi-path resolution using the shutter patterns shown in Figure 5 is a combinatorial problem. We assume a situation that the reflected light from an object placed at depth d_i is detected by our computational CMOS image sensor using n -tap charge modulators. For the shutter patterns for a shutter period of T_j , k_j -th time window detects the reflection. k_j is denoted by

$$k_j = \left\lfloor \frac{2d_i}{c} \bmod \frac{T_j}{n} \right\rfloor + 1. \tag{15}$$

Here, $\lfloor x \rfloor$ means the integer that is not more than x . In our proposed method, we prepare multiple T_j 's. Note that sub-clock shift is not considered here. For different T_j 's, k_j 's are detected from the captured image. However, when there are multiple reflections, k_j becomes ambiguous. Namely, it is not easy to tell which signal for each T_j comes from the same reflection. When signal peaks are separated for a certain T_j and k_j is specified for a reflection, we can resolve the multi-path interference more easily. There are two cases: (1) k_{j1} (for higher frequency) is resolved, but k_{j2} is not determined. (2) k_{j2} (for lower frequency) is resolved, but k_{j1} is not determined. For case 1, k_{j2} is written as follows when $T_{j2} = 2T_{j1}$:

$$k_{j2} = \left\lfloor qT_{j1} + \frac{(k_{j1} - 1)T_{j1}}{n} \bmod \frac{T_{j2}}{n} \right\rfloor + 1. \tag{16}$$

Here, q is zero or a positive integer and mod means the modulation operator. The problem is to find q .

In case 2, another ambiguity arises. Because the time window duration for T_{j2} is twice longer than that for T_{j1} , there are two choices as follows:

$$k_{j1} = \left\lfloor \frac{(k_{j2} - 1)T_{j2}}{n} \bmod \frac{T_{j1}}{n} \right\rfloor + 1 \text{ or } \left\lfloor \frac{(k_{j2} - 1)T_{j2}}{n} \bmod \frac{T_{j1}}{n} \right\rfloor + 2. \quad (17)$$

It is also necessary for the precision and accuracy to be small enough to determine the combinations of the detected signals for the multiple frequencies. To improve the capability of multi-path resolution, it is effective to increase the number of taps of the charge modulator. However, trade-off among the pixel size, response time, and separability should be considered.

Although only surface reflection was incorporated in the mathematical model in this paper, we believe that the proposed method could be extended to compensate multiple reflections and volumetric scattering because they significantly increase the path length or time of flight of light. The most difficult point is that the reflected light waveform is not the same as that of the emitted light. Probably, spatio-temporal point spread functions for these multi-path components should be also estimated. On the other hand, separation of sub-surface scattering could be more challenging and requires different techniques other than the application of temporal shutter patterns because the light intensity caused by the sub-surface scattering decays very quickly, e.g., in less than hundreds of picoseconds.

Improvement of the algorithm is also necessary. Although it is assumed that the incident light signal is uniform over a macro-pixel, we should consider the point spread function to realize subpixel resolution. Development of a non-iterative real-time signal processing algorithm is also key. In this paper, our aim was to investigate the potential of the multi-tap macro-pixel computational CMOS image sensor in multi-path separation. As shown in Section 4.2, the processing time for the depth refinement stage occupies 99.3% of the total processing time. Most of the processing time was dedicated to the nonlinear optimization. As the next step, algorithms suitable for real-time processing should be explored [30].

6. Conclusions

In this paper, the separation of multi-path components in time-of-flight depth imaging using a multi-tap macro-pixel computational CMOS image sensor was demonstrated. The macro-pixel has 2×2 subpixels, and each subpixel is implemented by a four-tap lateral electric field charge modulator (LEFM). Thus, 16 images for different temporal shutters are acquired in a single shot. Multiple frequency-time windows and sub-clock shifting were adopted to separate the multi-path interference components and to achieve both long distance range and high distance resolution. The following two-step depth estimation was used to investigate the potential of the proposed scheme for multi-path component separation. In the simulation without depth-dependent decay of the number of photons, the absolute value of the relative mean error was $<1.2\%$ for both d_1 (objective depth) and d_2 (interference depth), except when d_2 was very small (at 1 m) or two reflections were merged (at 16 m). When the amplitudes of the two reflections were the same, the error at $d_2 = 26$ m became large, probably because the solution could converge to a local minimum. For the simulation considering the depth-dependent decay of light intensity, the nonlinearity of the interference light increased significantly at long distances, probably under the effect of the strong objective light. In the experiment, we separated two reflections from an acrylic plate placed at 1 m or 2 m and a mirror placed at 9.3 m. The absolute value of the mean error compared with the single-path situation was less than 0.2 m.

Author Contributions: Conceptualization, T.K. (Takashi Komuro), H.N. and K.K.; experiments and data processing, M.H., Y.F. and T.K. (Tomoya Kokado); algorithm and programming, Y.F.; hardware implementation, T.T.; image sensor design, T.K. (Tomoya Kokado), K.Y., S.K. and K.K.; supervision and project administration, K.K.; funding acquisition, H.N. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by JSPS Grant-in-Aid for Scientific Research (S) Numbers 17H06102 and 18H05240.

Acknowledgments: This work was supported by the VLSI Design and Education Center (VDEC), The University of Tokyo, in collaboration with Cadence Corporation, Synopsys Corporation, and Mentor Graphics Corporation. The authors would like to thank DB Hitek for the CIS chip fabrication.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Remondino, F.; Stoppa, D. *TOF Range-Imaging Cameras*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 68121.
2. He, Y.; Chen, S. Recent Advances in 3D Data Acquisition and Processing by Time-of-Flight Camera. *IEEE Access* **2019**, *7*, 12495–12510. [CrossRef]
3. Feigin, M.; Bhandari, A.; Izadi, S.; Rhemann, C.; Schmidt, M.; Raskar, R. Resolving Multi-path Interference in Kinect: An Inverse Problem Approach. *IEEE Sens. J.* **2016**, *16*, 3419–3427. [CrossRef]
4. Whyte, R. Multi-Path Interference in Indirect Time-of-Flight Depth Sensors. Available online: <https://medium.com/chronoptics-time-of-flight/multi-path-interference-in-indirect-time-of-flight-depth-sensors-7f59c5fcd122> (accessed on 28 April 2021).
5. Ximenes, A.R.; Padmanabhan, P.; Lee, M.-J.; Yamashita, Y.; Yaung, D.N.; Charbon, E. 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. In Proceedings of the 2018 IEEE International Solid—State Circuits Conference—(ISSCC), San Francisco, CA, USA, 11–15 February 2018; pp. 96–98.
6. Acerbi, F.; Moreno-Garcia, M.; Koklü, G.; Gancarz, R.M.; Büttgen, B.; Biber, A.; Furrer, D.; Stoppa, D. Optimization of Pinned Photodiode Pixels for High-Speed Time of Flight Applications. *IEEE J. Electron Devices Soc.* **2018**, *6*, 365–375. [CrossRef]
7. Cova, S.; Longoni, A.; Andreoni, A. Towards picosecond resolution with single-photon avalanche diodes. *Rev. Sci. Instrum.* **1981**, *52*, 408–412. [CrossRef]
8. Bamji, C.S.; Mehta, S.; Thompson, B.; Elkhatib, T.; Wurster, S.; Akkaya, O.; Payne, A.; Godbaz, J.; Fenton, M.; Rajasekaran, V.; et al. 1Mpixel 65 nm BSI 320MHz Demodulated TOF Image Sensor with 3 μm Global Shutter Pixels and Analog Binning. In Proceedings of the 2018 IEEE International Solid—State Circuits Conference—(ISSCC), San Francisco, CA, USA, 11–15 February 2018; pp. 94–96.
9. Yasutomi, K.; Okura, Y.; Kagawa, K.; Kawahito, S. A Sub-100 μm -Range-Resolution Time-of-Flight Range Image Sensor with Three-Tap Lock-In Pixels, Non-Overlapping Gate Clock, and Reference Plane Sampling. *IEEE J. Solid-State Circuits* **2019**, *54*, 2291–2303. [CrossRef]
10. Kadambi, A.; Whyte, R.; Bhandari, A.; Streeter, L.; Barsi, C.; Dorrington, A.; Raskar, R. Coded Time of Flight Cameras: Sparse Deconvolution to Address Multi-path Interference and Recover Time Profiles. *ACM Trans. Graph. (ToG)* **2013**, *32*, 167. [CrossRef]
11. Kitano, K.; Okamoto, T.; Tanaka, K.; Aoto, T.; Kubo, H.; Funatomi, T.; Mukaigawa, Y. Recovering Temporal PSF Using ToF Camera with Delayed Light Emission. *IPSJ Trans. Comput. Vis. Appl.* **2017**, *9*, 15. [CrossRef]
12. Bhandari, A.; Kadambi, A.; Whyte, R.; Barsi, C.; Feigin, M.; Dorrington, A.; Raskar, R. Resolving Multi-path Interference in Time-of-Flight Imaging via Modulation Frequency Diversity and Sparse Regularization. *Opt. Lett.* **2014**, *39*, 1705–1708. [CrossRef] [PubMed]
13. Kadambi, A.; Schiel, J.; Raskar, R. Macroscopic Interferometry: Rethinking Depth Estimation with Frequency-Domain Time-of-Flight. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 893–902.
14. Lickfold, C.A.; Streeter, L.; Cree, M.J.; Scott, J.B. Simultaneous Phase and Frequency Stepping in Time-of-Flight Range Imaging. In Proceedings of the 2019 IEEE SENSORS, Montreal, QC, Canada, 27–30 October 2019; pp. 1–4.
15. Agresti, G.; Schaefer, H.; Sartor, P.; Zanuttigh, P. Unsupervised Domain Adaptation for ToF Data Denoising with Adversarial Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5584–5593.
16. Marco, J.; Hernandez, Q.; Munoz, A.; Dong, Y.; Jarabo, A.; Kim, M.H.; Tong, X.; Gutierrez, D. DeepToF: Off-the-Shelf Real-Time Correction of Multi-path Interference in Time-of-Flight Imaging. *ACM Trans. Graph. (ToG)* **2017**, *36*, 219. [CrossRef]
17. Mochizuki, F.; Kagawa, K.; Miyagi, R.; Seo, M.-W.; Zhang, B.; Takasawa, T.; Yasutomi, K.; Kawahito, S. Separation of Multi-Path Components in Sweep-Less Time-of-Flight Depth Imaging with a Temporally-Compressive Multi-Aperture Image Sensor. *ITE Trans. Media Technol. Appl.* **2018**, *6*, 202–211. [CrossRef]
18. Mochizuki, F.; Kagawa, K.; Okihara, S.; Seo, M.-W.; Zhang, B.; Takasawa, T.; Yasutomi, K.; Kawahito, S. Single-Shot 200 Mfps 5×3 -Aperture Compressive CMOS Imager. In Proceedings of the 2015 IEEE International Solid-State Circuits Conference-(ISSCC) Digest of Technical Papers, San Francisco, CA, USA, 22–26 February 2015; pp. 1–3.
19. Baraniuk, R.G. Compressive Sensing [Lecture Notes]. *IEEE Signal Process. Mag.* **2007**, *24*, 118–121. [CrossRef]
20. Romberg, J. Imaging via Compressive Sampling. *IEEE Signal Process. Mag.* **2008**, *25*, 14–20. [CrossRef]
21. Dadkhah, M.; Deen, M.J.; Shirani, S. Compressive Sensing Image Sensors-Hardware Implementation. *Sensors* **2013**, *13*, 4961–4978. [CrossRef] [PubMed]

22. Duarte, M.F.; Davenport, M.A.; Takhar, D.; Laska, J.N.; Sun, T.; Kelly, K.F.; Baraniuk, R.G. Single-Pixel Imaging via Compressive Sampling. *IEEE Signal Process. Mag.* **2008**, *25*, 83–91. [[CrossRef](#)]
23. Kagawa, K.; Kokado, T.; Sato, Y.; Mochizuki, F.; Nagahara, H.; Takasawa, T.; Yasutomi, K.; Kawahito, S. Multi-Tap Macro-Pixel Based Compressive Ultra-High-Speed CMOS Image Sensor. In Proceedings of the 2019 International Image Sensor Workshop, Snowbird, UT, USA, 23–27 June 2019; pp. 270–273.
24. Li, C.; Yin, W.; Zhang, Y. TVAL3: TV Minimization by Augmented Lagrangian and Alternating Direction Algorithms. Available online: <https://www.caam.rice.edu/~optimization/L1/TVAL3/> (accessed on 28 April 2021).
25. Li, C. An Efficient Algorithm for Total Variation Regularization with Applications to the Single Pixel Camera and Compressive Sensing. Ph.D. Thesis, Rice University, Houston, TX, USA, 2010.
26. D’Errico, J. `fminsearchbnd`, `fminsearchcon`. Available online: <https://jp.mathworks.com/matlabcentral/fileexchange/8277-fminsearchbnd-fminsearchcon> (accessed on 28 April 2021).
27. Kawahito, S.; Baek, G.; Li, Z.; Han, S.-M.; Seo, M.-W.; Yasutomi, K.; Kagawa, K. CMOS Lock-in Pixel Image Sensors with Lateral Electric Field Control for Time-Resolved Imaging. In Proceedings of the International Image Sensor Workshop (IISW), Snowbird, UT, USA, 12–16 June 2013; Volume 361.
28. Seo, M.-W.; Kagawa, K.; Yasutomi, K.; Kawata, Y.; Teranishi, N.; Li, Z.; Halin, I.A.; Kawahito, S. A 10 Ps Time-Resolution CMOS Image Sensor with Two-Tap True-CDS Lock-in Pixels for Fluorescence Lifetime Imaging. *IEEE J. Solid-State Circuits* **2015**, *51*, 141–154.
29. Lee, S.; Yasutomi, K.; Morita, M.; Kawanishi, H.; Kawahito, S. A Time-of-Flight Range Sensor Using Four-Tap Lock-In Pixels with High near Infrared Sensitivity for LiDAR Applications. *Sensors* **2020**, *20*, 116. [[CrossRef](#)] [[PubMed](#)]
30. Conde, M.H.; Kagawa, K.; Kokado, T.; Kawahito, S.; Loffeld, O. Single-Shot Real-Time Multiple-Path Time-of-Flight Depth Imaging for Multi-Aperture and Macro-Pixel Sensors. In Proceedings of the ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 1469–1473.