

RESEARCH PAPER

Open Access



Phase disambiguation using spatio-temporally modulated illumination in depth sensing

Takahiro Kushida^{1*}, Kenichiro Tanaka¹, Takahito Aoto², Takuya Funatomi¹ and Yasuhiro Mukaigawa¹

Abstract

Phase ambiguity is a major problem in the depth measurement in either time-of-flight or phase shifting. Resolving the ambiguity using a low frequency pattern sacrifices the depth resolution, and using multiple frequencies requires a number of observations. In this paper, we propose a phase disambiguation method that combines temporal and spatial modulation so that the high depth resolution is preserved while the number of observation is kept. A key observation is that the phase ambiguities of temporal and spatial domains appear differently with respect to the depth. Using this difference, the phase can disambiguate for a wider range of interest. We develop a prototype to show the effectiveness of our method through real-world experiments.

Keywords: Time-of-flight camera, Phase shifting, Computational photography

1 Introduction

Depth measurement is widely used in applications such as augmented reality, factory automation, robotics, and autonomous driving. In the computer vision field, there are two well-known techniques for measuring scene depth using active illumination. One is the time-of-flight camera, which uses temporally modulated illumination to measure the travel time of light; the other is the phase shifting, which uses temporally modulated illumination to find the correspondence between the projector and the camera for triangulation.

A common problem is how to resolve the periodic ambiguity of the phase because either measurement gives the phase that is defined between 0 to 2π . Typical solution is to use multiple frequencies to resolve the phase ambiguity. However, the phase ambiguity still exists in the frequency of the greatest common divisor, which requires several measurements to obtain a wider range of interest. Another possible approach is to use a low frequency that sacrifices

the depth resolution. The aim of this study is to resolve the phase ambiguity in fewer observations, where both the wider range of interest and the better resolution of the depth are guaranteed.

A key observation of this paper is that the phase ambiguities of the time-of-flight (ToF) and the phase shifting appear differently on the depth domain. Since the temporal phase is proportional to the depth, the depth candidates from the phase appear at equal intervals along with the depth. On the other hand, the spatial phase is defined as the disparity domain; hence, the depth candidates appear at gradually increasing intervals. Based on this difference, the phase ambiguity can be resolved by combining temporal and spatial modulation. Because the candidate depth that satisfies both measured phases seldom appears, the number of phase can be reduced to one for each frequency. In this paper, we discuss ordinary ToF and phase shifting in the same framework. We show that precise depth can be measured in a wide range by combining temporal and spatial modulation. We also reveal the resolution and the range of interest theoretically, analyze the recoverability, and build a prototype to show the effectiveness of our method via real-world experiments.

*Correspondence: kushida.takahiro.kh3@is.naist.jp

¹Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma 630-0192, Japan

Full list of author information is available at the end of the article

This paper extends its preliminary version [1] with the following differences. Extensions have been made to (1) reveal the depth resolution and the range of interest of our proposed method, (2) develop an efficient implementation, and (3) confirm that the unrecoverable depth due to ambiguity seldom exists by simulation.

The rest of the paper is organized as follows. The related work is discussed in Section 2, a brief review of the ordinary time-of-flight and phase shifting algorithms are provided in Section 3, a spatio-temporal modulation technique is proposed in Section 4, the resolution and range of interest of our method is analyzed in Section 5, experiments with a prototype system is shown in Section 6, and we conclude with some discussions in Section 7.

2 Related work

Active depth measurements have been widely studied in the computer vision field. Earlier work used a projector-camera system to convert the projector's pixel index into multiple projection images based on the gray code [2]. The phase shifting approach [3] recovers subpixel correspondences by detecting the phase of the sinusoid. Gupta and Nayer [4] unwrapped the phase from slightly different frequencies so that it became robust to indirect light transport with a small budget of projection numbers. Mirdehghan et al. [5] proposed an optimal code for the structured light technique. The time-of-flight method is another way to measure depth. It emits amplitude modulated light, and a delayed signal is detected that corresponds to the scene depth [6]. Because the range of interest and the depth resolution are tradeoffs, a better resolution is obtained by limiting the range of interest [7]. We combine these techniques to realize both better resolution and wider range of interest.

Another problem regarding the ToF is multi-path interference due to indirect light transport. Recovering the correct depth of multi-path scenes has been broadly studied using a parametric model [8, 9], K -sparsity [10, 11], frequency analysis [12], and data-driven approaches [13–15]. Because the scene depth can be recovered by the first-returning photon, the depth can be obtained after recovering light-in-flight imaging [16–21]. Multi-path interference is mitigated by combining ToF and projector. Naik et al. [22] combined the ToF camera and a projector-camera system to mitigate a multi-path that uses direct-global separation [23]. Similar ideas are implemented with the ToF projectors that can modulate both spatially and temporally [24, 25]. In both cases, direct-global separation is utilized to mitigate multi-path interference. We also use a similar system for phase disambiguation not only for mitigating multi-path.

To obtain fine resolution, Gupta et al. [26] proposes the optimal code for ToF modulation. Gutierrez-Barragan et al. [27] proposes an optimization approach for designing

practical coding functions under hardware constraints. Kadambi et al. [28] uses the polarization cue to recover the smooth surface. Our method is more fundamental layer; hence, these techniques can be incorporated with our method to boost the resolution. An interferometer can also obtain micrometer resolution of a small size object. Interferometry gives micrometer resolution [29] in a carefully controlled environment. Li et al. [30] recover micro-resolution ToF using the superheterodyne technique. Maeda et al. [31] leverages the heterodyne technique to the polarization imaging to obtain the accurate depth.

Phase unwrapping is a subproblem in the depth measurement. The phase has to be unwrapped with either the phase shifting or the ToF; otherwise, the estimated depth have 2π ambiguity. The number of observations can be reduced by sacrificing the spatial resolution. The projector's coordinates can be obtained from a single image using a color code [32], a wave grid pattern [33], and a light-field ToF [34]. Our method falls into this class but does not sacrifice the spatial resolution nor require many patterns. Our method leverages the asymmetric relations of spatial and temporal wrapping to solve the ambiguity of the phase.

3 Depth measurement techniques using modulated illumination

Before explaining our method, we briefly review the ToF and phase shifting methods. We respectively explain them as the phase measurements using temporally or spatially modulated light.

3.1 Temporal modulation (time-of-flight)

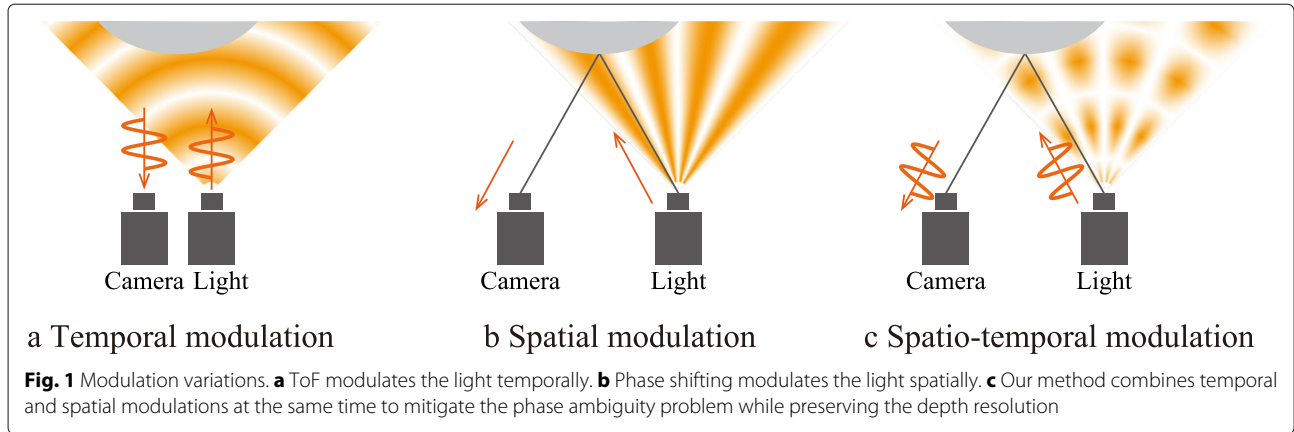
The ToF camera emits the temporally modulated light as shown in Fig. 1a. It measures the amplitude decay and phase delay of the modulated light, and the phase delay corresponds to the time it takes for the light to make a round trip.

The ToF camera measures the correlation between the signals emitted and those received. For each frequency, the phase delay is calculated from the correlations with N_T reference signals, which are temporally shifted. For the k -th signal, the correlation $i_k(x)$ at the camera pixel x is represented as

$$i_k(x) = g\left(t + \frac{2\pi k}{N_T}\right) * s(x, t) \quad (1)$$

$$= \frac{A(x)}{2} \cos\left(\phi_T(x) + \frac{2\pi k}{N_T}\right) + O(x), \quad (2)$$

where $g\left(t + \frac{2\pi k}{N_T}\right)$ is the reference signal with the shifted phase $2\pi k/N_T$, s is the returned signal, the $*$ operator represents the correlation, A is the amplitude decay, ϕ_T is the phase delay, and O is the ambient light. In the case of



$N_T = 4$, the phase ϕ_T and the amplitude A of the returned signal can be recovered by a direct conversion method from multiple observations while changing the phase $\frac{2\pi k}{N_T}$ as

$$\phi_T(x) = \arctan\left(\frac{i_3(x) - i_1(x)}{i_0(x) - i_2(x)}\right), \quad (3)$$

$$A(x) = \sqrt{(i_3(x) - i_1(x))^2 + (i_0(x) - i_2(x))^2}. \quad (4)$$

The depth d is obtained as

$$d(x) = \frac{c}{2\omega_T} \phi_T(x), \quad (5)$$

where ω_T is the modulation frequency and c is the speed of light.

3.2 Spatial modulation (phase shifting)

The phase shifting spatially modulates the projection pattern. Finding the correspondences between the projector and camera pixels is the main part of the spatial phase shifting. The idea is to project the sinusoidal pattern as shown in Fig. 1b and measure the phase of the sinusoid for each pixel, which corresponds to the projector's pixel coordinates.

The observed intensity of the camera $I_l(x)$ for l -th shift is represented as

$$I_l(x) = A(x) \cos\left(\phi_S(x) - \frac{2\pi l}{N_S}\right) + O(x), \quad (6)$$

where ϕ_S is the spatial phase of the projection pattern due to disparity. There are three unknown parameters, which are the offset O , the amplitude $A(x)$, and the phase $\phi_S(x)$; therefore, they can be recovered from $N_S \geq 3$ observations while changing the phase of the pattern. In the case of $N_S = 4$, the spatial phase ϕ_S and the amplitude A can be recovered in the same way as the ToF as

$$\phi_S(x) = \arctan\left(\frac{I_3(x) - I_1(x)}{I_0(x) - I_2(x)}\right), \quad (7)$$

$$A(x) = \sqrt{(I_3(x) - I_1(x))^2 + (I_0(x) - I_2(x))^2}. \quad (8)$$

From the estimated disparity, the scene depth can be recovered using the triangulation theory. For example, when the parallel stereo is assumed, the depth is inversely proportional to the disparity as

$$d(x) = \frac{bf}{x - \frac{\phi_S(x)}{\omega_S}} \quad (9)$$

where $x - \frac{\phi_S(x)}{\omega_S}$ is the disparity, ω_S is the spatial angular frequency of the projection pattern, f is the focal length, and b is the baseline of the pro-cam system. Here, x represents the horizontal pixel position.

3.3 Phase ambiguity and depth resolution

A common problem in both temporal and spatial methods is 2π ambiguity, where the phase is wrapped when the depth exceeds the maximum depth of interest. A naive approach is using a low frequency to avoid the phase ambiguity. However, a tradeoff exists between the range of interest and the depth resolution. While the phase ambiguity does not appear at a lower frequency, the depth resolution becomes low as shown in Fig. 2a. With a higher frequency, the depth resolution improves while the phase ambiguity becomes significant, and the depth cannot be uniquely recovered for a wide range of interest as shown in Fig. 2b.

The phase ambiguity is usually relaxed by using multiple frequencies in either a temporal or a spatial domain. However, multiple captures are required, and it sacrifices real-time possibility as shown in Fig. 2c. We propose a hybrid approach of disambiguation that can take advantage of a different nature in temporal and spatial modulation.

4 Proposed method

We propose a hybrid method of temporal and spatial modulation as shown in Fig. 1c. The phase ambiguity can be resolved by using both temporal and spatial phases instead of using multiple frequencies in either domain.

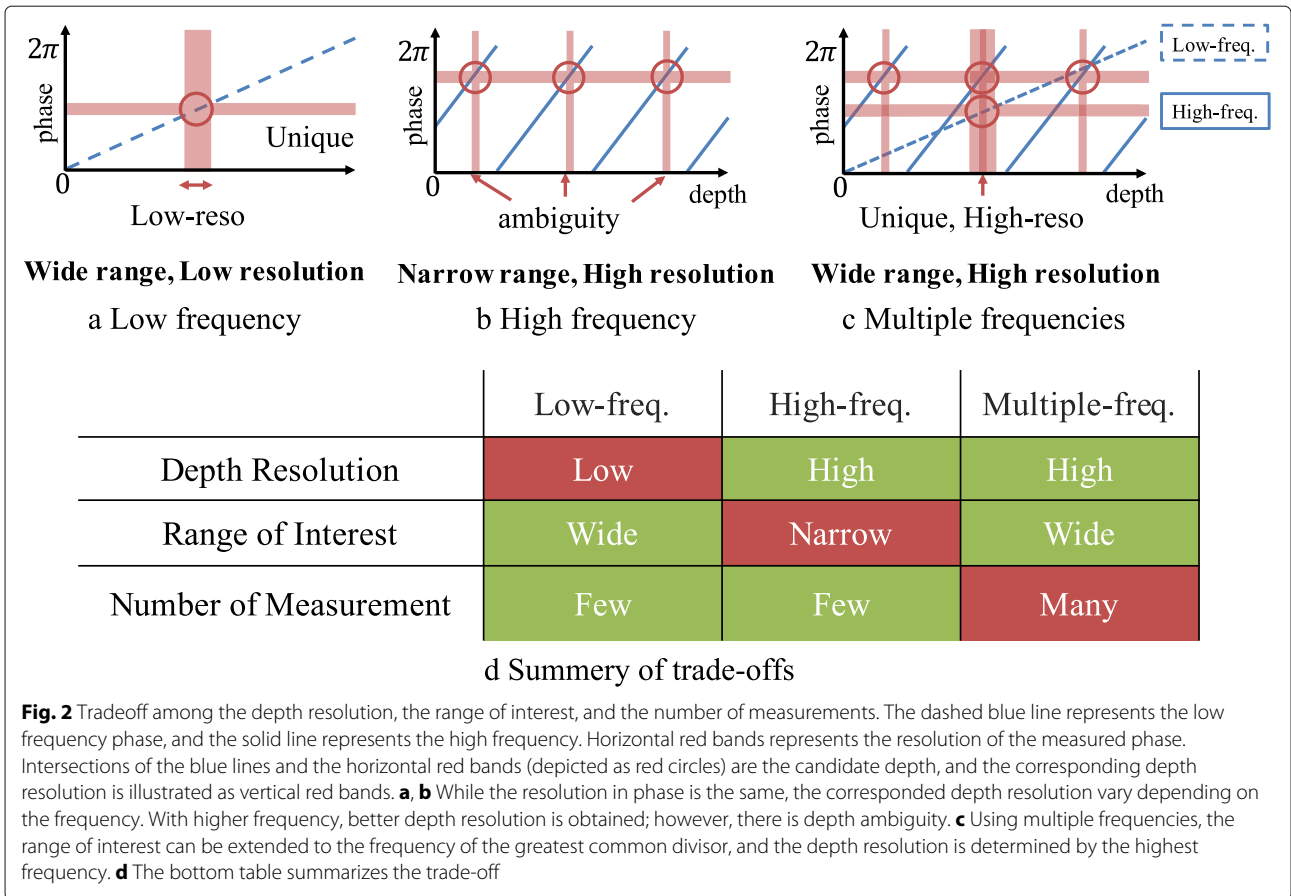


Fig. 2 Tradeoff among the depth resolution, the range of interest, and the number of measurements. The dashed blue line represents the low frequency phase, and the solid line represents the high frequency. Horizontal red bands represents the resolution of the measured phase. Intersections of the blue lines and the horizontal red bands (depicted as red circles) are the candidate depth, and the corresponding depth resolution is illustrated as vertical red bands. **a, b** While the resolution in phase is the same, the corresponded depth resolution vary depending on the frequency. With higher frequency, better depth resolution is obtained; however, there is depth ambiguity. **c** Using multiple frequencies, the range of interest can be extended to the frequency of the greatest common divisor, and the depth resolution is determined by the highest frequency. **d** The bottom table summarizes the trade-off

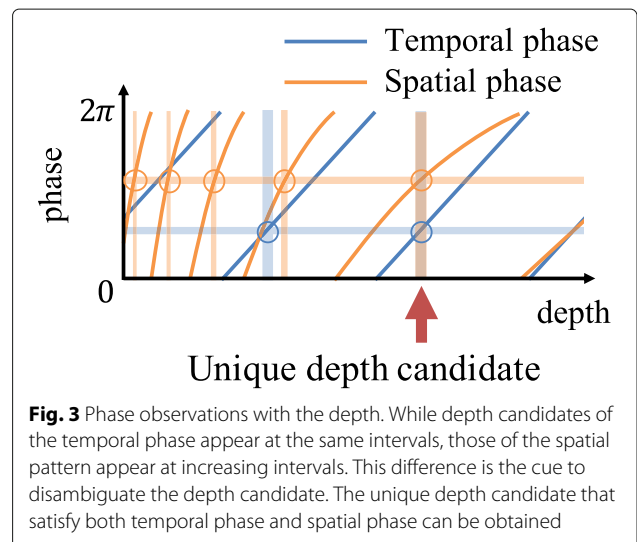
4.1 Spatio-temporal phase disambiguation

Our key idea is that the depth candidates from the ambiguity of the temporal and spatial phases are different. In the case of the temporal phase, the intervals of the depth candidates are constant along the depth because the depth is proportional to the phase, as shown in Eq. (5). On the other hand, the spatial phase is defined in the disparity domain. Because the depth is inversely proportional to the disparity (as shown in Eq. (9)), the intervals of depth candidates increase along with the depth. Figure 3 shows the phase observations along with the scene depth. Multiple depth candidates correspond to a single phase. The depth candidates appear at the same interval for the temporal phase, while the intervals of the spatial phase increase. This difference is a key feature of our method to resolve the phase ambiguity.

Depths that satisfy both temporal and spatial phases seldom appear. The unwrapped phase is not restricted by the greatest common divisor, and the set of temporal and spatial phases is unique for the wider range of interest. The candidate depths can be respectively obtained from the following equations as

$$d_T = \frac{c}{2\omega_T} (2\pi n_T + \phi_T) \tag{10}$$

$$d_S = \frac{bf}{x - \frac{2\pi n_S + \phi_S}{\omega_S}} \tag{11}$$



The integer pair (n_T, n_S) that satisfies $d_T = d_S$ seldom exists. Therefore, the phase ambiguity problem can be resolved using phases of different domains.

4.2 Phase recovery and depth estimation

Defining I_0 as the irradiance, the emitted signal from the projector with the k -th temporal shift and the l -th spatial shift $I(p, t, k, l)$ can be expressed as

$$I(p, t, k, l) = I_0 \left(\frac{1}{2} \cos \left(\omega_T t + \frac{2\pi k}{N_T} \right) + \frac{1}{2} \right) \left(\frac{1}{2} \cos \left(\omega_S p - \frac{2\pi l}{N_S} \right) + \frac{1}{2} \right), \quad (12)$$

where t is time and p is the projector's pixel. The returned signal $r(x, t, k, l)$ at the camera pixel x is represented as

$$r(x, t, k, l) = I_0 \kappa(x) \left(\frac{1}{2} \cos \left(\omega_T t - \phi_T(x) - \frac{2\pi k}{N_T} \right) + \frac{1}{2} \right) \left(\frac{1}{2} \cos \left(\phi_S(x) - \frac{2\pi l}{N_S} \right) + \frac{1}{2} \right) + o(x), \quad (13)$$

where κ is the reflectance of target object, $o(x)$ is the ambient light, $\phi_T(x)$ is the phase delay corresponding to the round trip time, and $\phi_S(x)$ is the phase corresponding to the disparity $(x - p)$. The intensity is the correlation with the reference signal $g_{\omega_T}(t)$ [35] as

$$i(x, k, l) = \int_0^T r(x, t, k, l) g_{\omega_T}(t) dt \approx A(x) \left(\frac{1}{2} \cos \left(\phi_T(x) + \frac{2\pi k}{N_T} \right) + \frac{1}{2} \right) \left(\frac{1}{2} \cos \left(\phi_S(x) - \frac{2\pi l}{N_S} \right) + \frac{1}{2} \right) + O(x), \quad (14)$$

where T is the exposure time. The temporal phase ϕ_T and spatial phase ϕ_S are obtained from 8 observations with $N_T = 4$ and $N_S = 4$ as

$$\begin{cases} \phi_T(x) = \arctan \frac{i(x, 3, 0) - i(x, 1, 0)}{i(x, 0, 0) - i(x, 2, 0)} \\ \phi_S(x) = \arctan \frac{i(x, 0, 3) - i(x, 0, 1)}{i(x, 0, 0) - i(x, 0, 2)}. \end{cases} \quad (15)$$

Now, we have two phases: the temporal phase ϕ_T and the spatial phase ϕ_S . Depth estimation from the two phases is similar to the unwrapping problem in both the multi-frequency phase shifting and the ToF, and it can be solved by searching a lookup table [4]. The observed phases should respectively equal to the phases computed from the same depth, the computed phase $\tilde{\phi}_T(d)$, $\tilde{\phi}_S(d)$ is

obtained as

$$\tilde{\phi}_T(d) = \frac{2\omega_T d}{c} \bmod 2\pi \quad (16)$$

$$\tilde{\phi}_S(d, x) = \omega_S \left(x - \frac{bf}{d} \right) \bmod 2\pi. \quad (17)$$

A lookup table is built for each horizontal pixel position x of the camera because the spatial phase depends on the pixel position. The table \mathcal{T}_x at the horizontal position x consists of the vector $\Phi_{D_i, x} = [\tilde{\phi}_T(D_i), \tilde{\phi}_S(D_i, x)]$ of the candidate depth D_i as

$$\mathcal{T}_x(D_i) = \Phi_{D_i, x} = [\tilde{\phi}_T(D_i), \tilde{\phi}_S(D_i, x)]. \quad (18)$$

For each pixel, the depth can be estimated by searching the lookup table as

$$\hat{d}(x) = \arg \min_d \|\mathcal{T}_x(d) - [\phi_T(x), \phi_S(x)]\|_2^2. \quad (19)$$

Efficient implementation In practice, building the lookup table for each horizontal pixel position is not necessary. Although the spatial phase and corresponding depth depends on the position of camera pixel, the disparity does not depend on the position of the camera pixel. The depth of all camera pixels can be obtained by only one look up table by building from the pair of temporal phase and the disparity after converting the measured phase to the disparity. The disparity is obtained from the measured spatial phase ϕ_S and pixel position x as

$$\delta(x, \phi_S(x)) = x - \frac{\phi_S(x)}{\omega_S} \quad (20)$$

$$= \frac{bf}{\tilde{d}}, \quad (21)$$

where δ represents the disparity and \tilde{d} is the wrapped depth. The table \mathcal{T}' consists of the vector $\Phi'_{D_i} = [\tilde{\phi}_T(D_i), \tilde{\delta}(D_i)]$ of the candidate depth D_i as

$$\tilde{\delta}(D_i) = \frac{bf}{D_i} \bmod \frac{2\pi}{\omega_S} \quad (22)$$

$$\mathcal{T}'(D_i) = \Phi'_{D_i} = [\tilde{\phi}_T(D_i), \tilde{\delta}(D_i)], \quad (23)$$

where $\tilde{\delta}$ is the computed disparity from candidate depths. For each pixel, the depth can be estimated by searching the lookup table as

$$\hat{d}(x) = \arg \min_d \|\mathcal{T}'(d) - [\phi_T(x), \delta(x, \phi_S)]\|_2^2. \quad (24)$$

5 Analysis of the proposed method

Depth resolution The resolution is better than ToF in a near range and better than phase shifting in a far range.

The resolution of ordinary ToF and phase shifting is respectively represented as [6, 25]

$$\Delta d_T = \frac{c\pi}{\omega_T} \frac{\sqrt{B}}{2\sqrt{8A}}, \quad (25)$$

$$\Delta d_S = \frac{2\pi d^2}{bf\omega_S} \frac{\sqrt{B}}{2\sqrt{8A}}, \quad (26)$$

where A and B are the number of photo-electrons that the sensor can accumulate and represents the amplitude and the DC component, respectively. We suppose that A and B are the parameters of the hardware and are independent from the scene. However, the returned light is influenced by the light falloff in real; hence, a future work is expected to include this effect to analyze more accurately.

Figure 4 shows the depth resolution of ToF and phase shifting along with the depth according to Eqs. (25) and (26). The resolution of ToF is constant at any depth while the resolution of phase shifting is proportional to the square of the depth. The proposed method achieves the resolution that is close to the better resolution of either phase shifting or time-of-flight as shown in Fig. 4.

The depth d_{cross} is defined by the depth where the resolution of ToF is equal to the resolution of phase shifting. In the range near than d_{cross} , the resolution of our method is better than ToF and close to phase shifting. In the range far than d_{cross} , the resolution of our method is better than phase shifting and close to ToF. The depth d_{cross} is given as

$$\Delta d_S = \Delta d_T \quad (27)$$

$$d_{\text{cross}} = \sqrt{\frac{cbf\omega_S}{2\omega_T}}. \quad (28)$$

When we want to improve the resolution of pure ToF, the maximum range of this system should be designated shorter than d_{cross} .

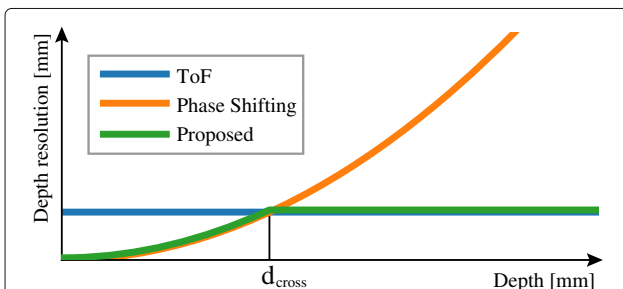


Fig. 4 Depth resolution along with the depth. According to Eqs. (25) and (26), the resolution of ToF is constant (blue) and the resolution of phase shifting is proportional to the square of the depth (orange). The depth d_{cross} is the depth where the lines of the resolution of ToF and the resolution of the phase shifting is crossed. The proposed method can achieve the resolution that is close to the phase shifting in near range before d_{cross} and the resolution that is close to the ToF in far range after d_{cross} (green)

Range of interest The range of interest (ROI) of the proposed method is determined by the relative relation between the temporal and the spatial frequencies.

Nearest range When the spatial frequency is too high compared with the temporal frequency, the phase ambiguity problem cannot be resolved because multiple candidate depths exist within the resolution of the ToF, as shown in Fig. 5a. The spatial frequency varies depending on the depth because the projection is perspective. As the distance is shorter, the spatial frequency is higher. This property gives the nearest ROI of the proposed method. The nearest ROI d_{min} is where the wrapping distance of spatial phase is equal to the resolution of the ToF at the given temporal and spatial frequencies as

$$d_S|_{n_S=n'_S} - d_S|_{n_S=n'_S-1} = \frac{\Delta d_T}{2}, \quad (29)$$

where $d_S|_{n_S=n'_S}$ is the unwrapped depth and $d_S|_{n_S=n'_S-1}$ is the neighbor depth candidate from Eq. (11). Substituting Eq. (17) and transforming the expression, the minimum depth of the range of interest d_{min} can be obtained as¹

$$d_{\text{min}} = \frac{\Delta d_T}{4} + \frac{1}{2} \sqrt{\frac{\Delta d_T^2}{4} + \frac{\omega_S b f \Delta d_T}{\pi}}. \quad (30)$$

Farthest range When the spatial frequency is too low compared with the temporal frequency, the phase ambiguity problem cannot be resolved because multiple candidate depths exist within the resolution of the spatial phase shifting, as shown in Fig. 5b. Because the resolution of the spatial phase shifting is inversely proportional to the depth, the farthest ROI d_{max} is determined. The farthest ROI d_{max} is where the wrapping distance of temporal phase is equal to the resolution of the phase shifting as

$$d_T|_{n_T=n'_T} - d_T|_{n_T=n'_T-1} = \frac{\Delta d_S}{2}, \quad (31)$$

where $d_T|_{n_T=n'_T}$ is the unwrapped depth and $d_T|_{n_T=n'_T-1}$ is the neighbor depth candidate from Eq. (11). Substituting Eq. (16), Eq. (26), and transforming the expression, the farthest ROI d_{max} can be obtained as²

$$d_{\text{max}} = \sqrt{\frac{\omega_S b f c^2 \pi}{\omega_T^2 \Delta d_T}}. \quad (32)$$

Unrecoverable point There are few unrecoverable depths in the proposed method. Figure 6 shows that the pair of temporal and spatial phases corresponding to the depth. The vertical axis is the temporal phase, and the horizontal axis is the spatial phase. The color of the curves represents the depth. The intersections of

¹Please see Appendix for the derivation.

²Please see Appendix for the derivation.

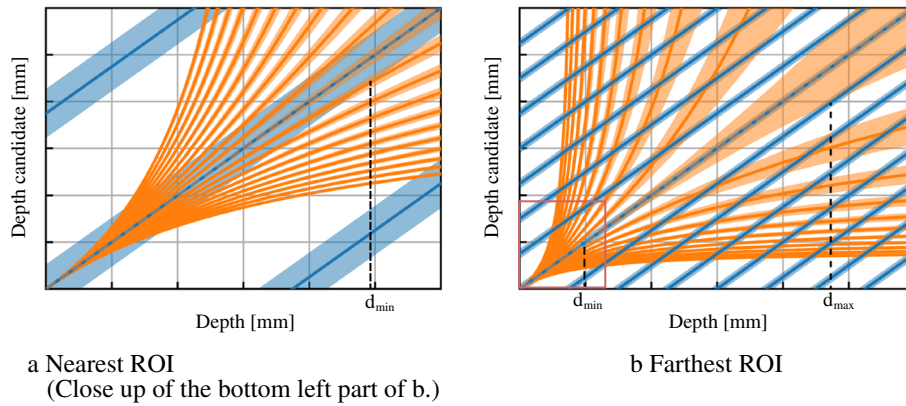


Fig. 5 Upper and lower bound of the ROI. Orange lines represent the candidate depths of spatial modulation; blue lines represent the candidate depths of temporal modulation. The width of the line shows the resolution. **a** If the depth is near than d_{min} , several candidate depths from spatial modulation (orange lines) exist within the resolution of temporal modulation (blue band). **b** On the other hand, if the depth is longer than d_{max} , several candidate depths from temporal modulation (blue lines) exist within the spatial resolution (orange band)

the curves are unrecoverable depth because different depths have the same phase pair. This is a limitation of this method; however, these points generally appeared sparsely in the image hence can be estimated by looking at neighbor pixels of the image.

We confirm that the unrecoverable points seldom exists via simulation. We evaluate the percentage of unrecoverable pixels in an image using an indoor dataset [36]. Temporal phases and spatial phases were respectively rendered, and the depth image is estimated by our method from these phase images. The temporal frequency is set

to 50 MHz, and spatial frequency is $1/0.6 \text{ mm}^{-1}$. One hundred scenes were selected from the dataset randomly.

The results are shown in Fig. 7. Depths of some pixels cannot be recovered due to multiple candidates. The average ratio of the uncovered pixel in each image is less than 5%. These points exist sparsely in the image; hence, it is possible to select the candidate by looking around their pixels.

Brightness of the pattern One may think that the temporal phase cannot be obtained if the spatial pattern is

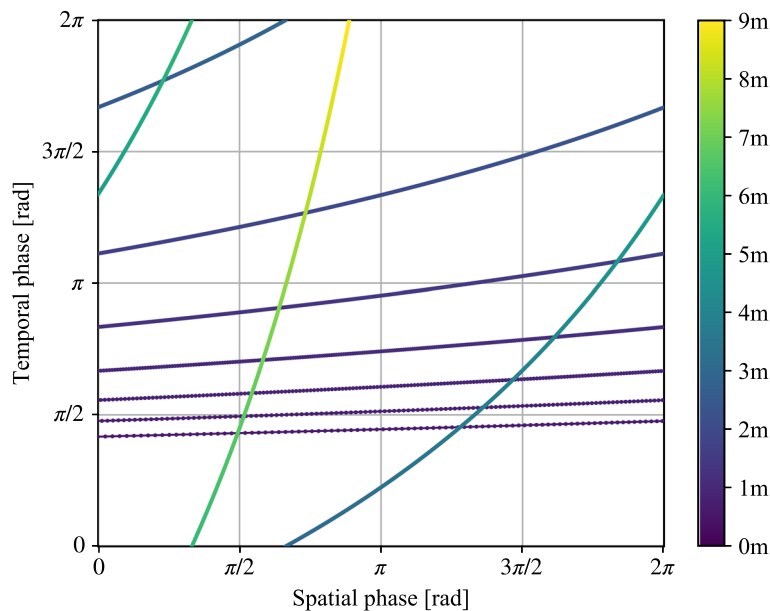
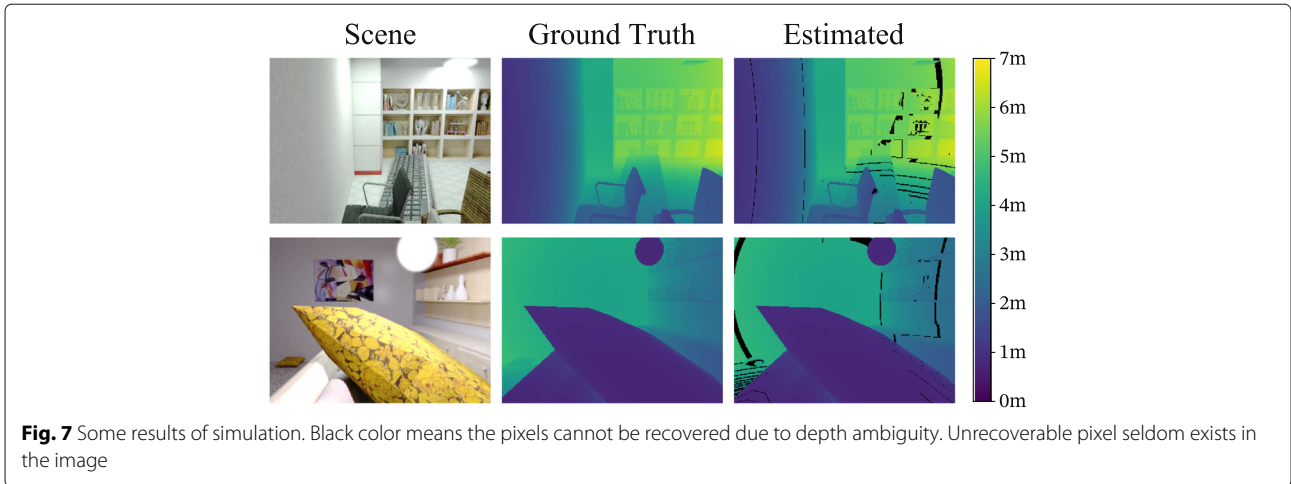


Fig. 6 The transition of temporal and spatial phases with respect to the depth. The vertical axis represents the temporal phase and the horizontal axis represents the spatial phase. The color represents the depth. The intersections of the curves have the same phase pair at the different depths. These depths cannot be recovered uniquely



completely black. Because the spatial sinusoidal pattern is projected, all the pixels have a chance to obtain the photons unless the spatial pattern is extremely low. A possible solution is to add the constant value to the spatial pattern so that there are no pixels that are always black. In this case, the observation Eq. (14) is rewritten as

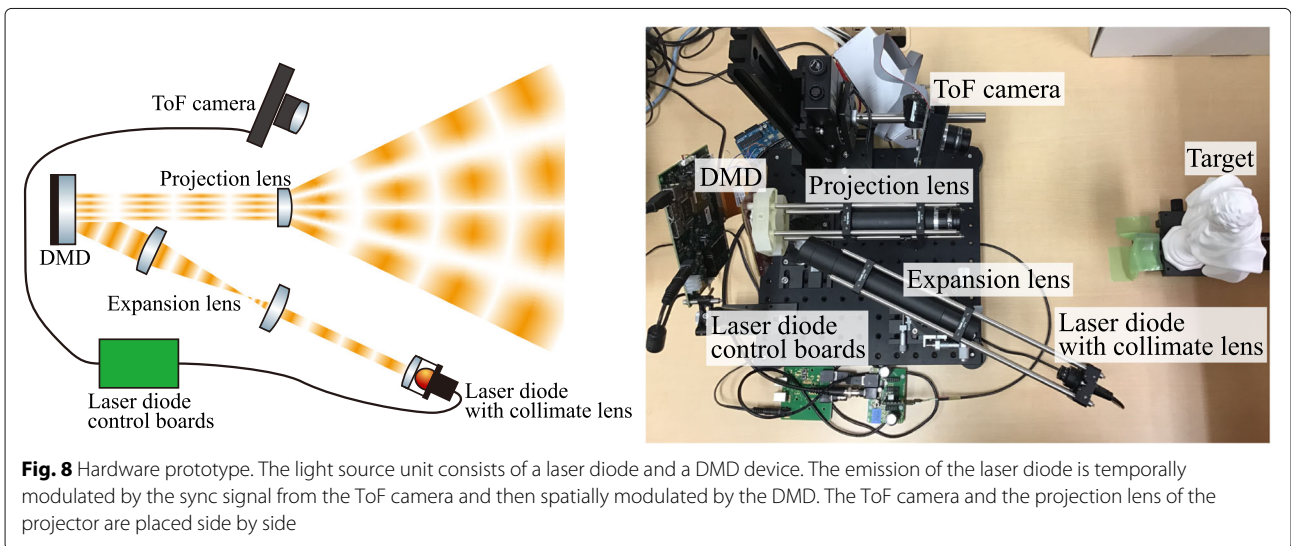
$$i(x, k, l) = A(x) \left(\frac{1}{2} \cos \left(\phi_T(x) + \frac{2\pi k}{N_T} \right) + \frac{1}{2} \right) \left(A_S \cos \left(\phi_S(x) - \frac{2\pi l}{N_S} \right) + O_S \right) + O(x), \tag{33}$$

where A_S and O_S ($0 < O_S - A_S$ and $O_S + A_S \leq 1$) are the amplitude and offset of the spatial modulation, respectively. Analogous to Eq. (14), both phases can be obtained by the same equations as Eq. (15) in the $N_T, N_S = 4$ case. So, it is not necessary to increase the number of observations.

6 Experiment

We demonstrated the effectiveness of our proposed method with real-world experiments.

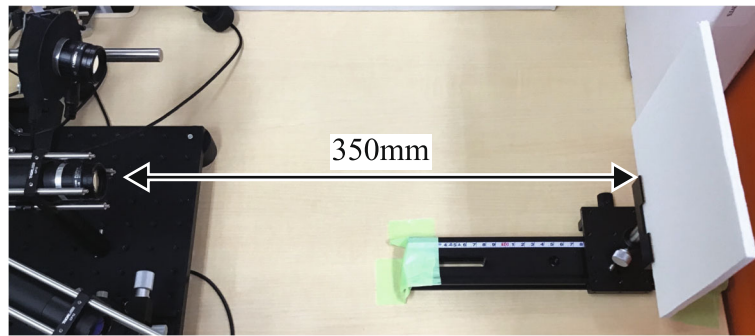
Hardware prototype We developed a hardware prototype that can illuminate a scene with a spatio-temporal modulated pattern. Our prototype was built onto a ToF camera (Texas Instruments OPT8241-CDK-EVM). The light source was replaced with a laser diode and a DMD system that can project the spatial pattern. The light source was an 830-nm laser diode (Hamamatsu Photonics L9277-42), and its emission was synchronized with the ToF sensor. The light emitted by the diode was collimated and expanded through lenses, and then reflected onto a DMD device (Texas Instruments DLP6500) that had 1920×1080 pixels. Finally, the spatio-temporal pattern was projected onto the scene through a projection lens, as shown in Fig. 8.



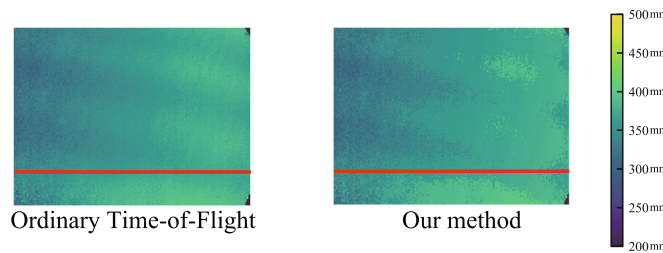
First, the measurement system was calibrated in a standard way for the pro-cam systems using a reference board [37]. The phase of the ToF on each pixel was then calibrated to share the same coordinates as the pro-cam system. A white plane board was captured while its position was moved for the phase calibration. For each measurement of the board, the pair of the raw phase and the ground-truth depth was obtained because the depth of the board was measured by the ordinary phase shifting.

The parameter to recover the depth from the phase was calibrated by line fitting.

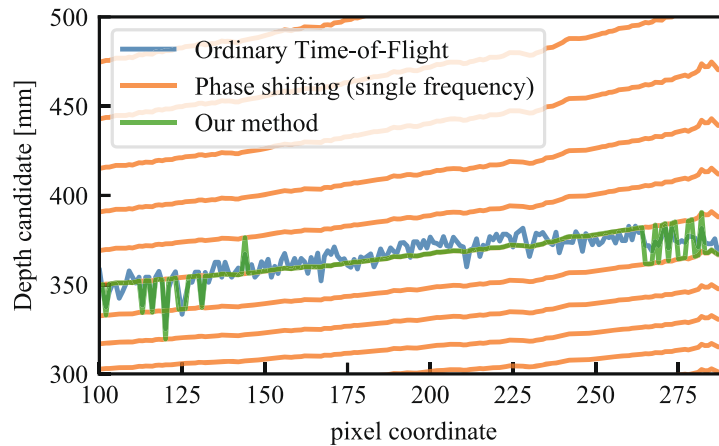
Result First, we measured a white planar board and placed it at approximately 350 mm from the camera and slightly slanted it, as shown in Fig. 9a. The temporal frequency was 60 MHz, and the period of the spatial pattern was 60 pixels on the projection image. The baseline between the camera and the projector was approximately



a Setup



b Depth images



c Cross-section of the depth image

Fig. 9 Results with a white planar board. Ordinary ToF, phase shifting (single high frequency), and our method are compared. **a** The object was placed at a slight slant. **b** The estimated depth images. Because the depth cannot be identified in the phase shifting, the depth image cannot be visualized. **c** The cross-section of the red line is shown. While the ordinary ToF is noisy and phase shifting has many candidates, our method recovers a smooth and unique depth candidate

70 mm, and the focal length of the projection lens was 35 mm.

The depths were obtained by an ordinary ToF with a single low frequency, phase shifting with single high frequency, and our method for the comparison. Figure 9b shows the estimated depth images. Both the ToF and our method recover the global depth. The depth image with phase shifting cannot be visualized because it has multiple depth candidates. The cross-section of the red line is shown in Fig. 9c. While the depth measured by the ordinary ToF is noisy and there are many depth candidates due to phase ambiguity in the phase shifting, our method recovers a smooth

surface while resolving the phase ambiguity. The region near the edge is not correctly disambiguated because the resolution of the temporal measurement exceeds the interval of the phase shifting. The ToF resolution near the edge is lower than what we expected because the illumination is very low near the edge. However, decreasing the spatial frequency might have mitigated it.

Finally, we measured a plaster bust and placed it approximately 400 mm from the camera, as shown in Fig. 10a. The estimated depth images are shown in Fig. 10b. The cross-section of the depth is shown in Fig. 10c. Our method recovers a unique and smooth depth.

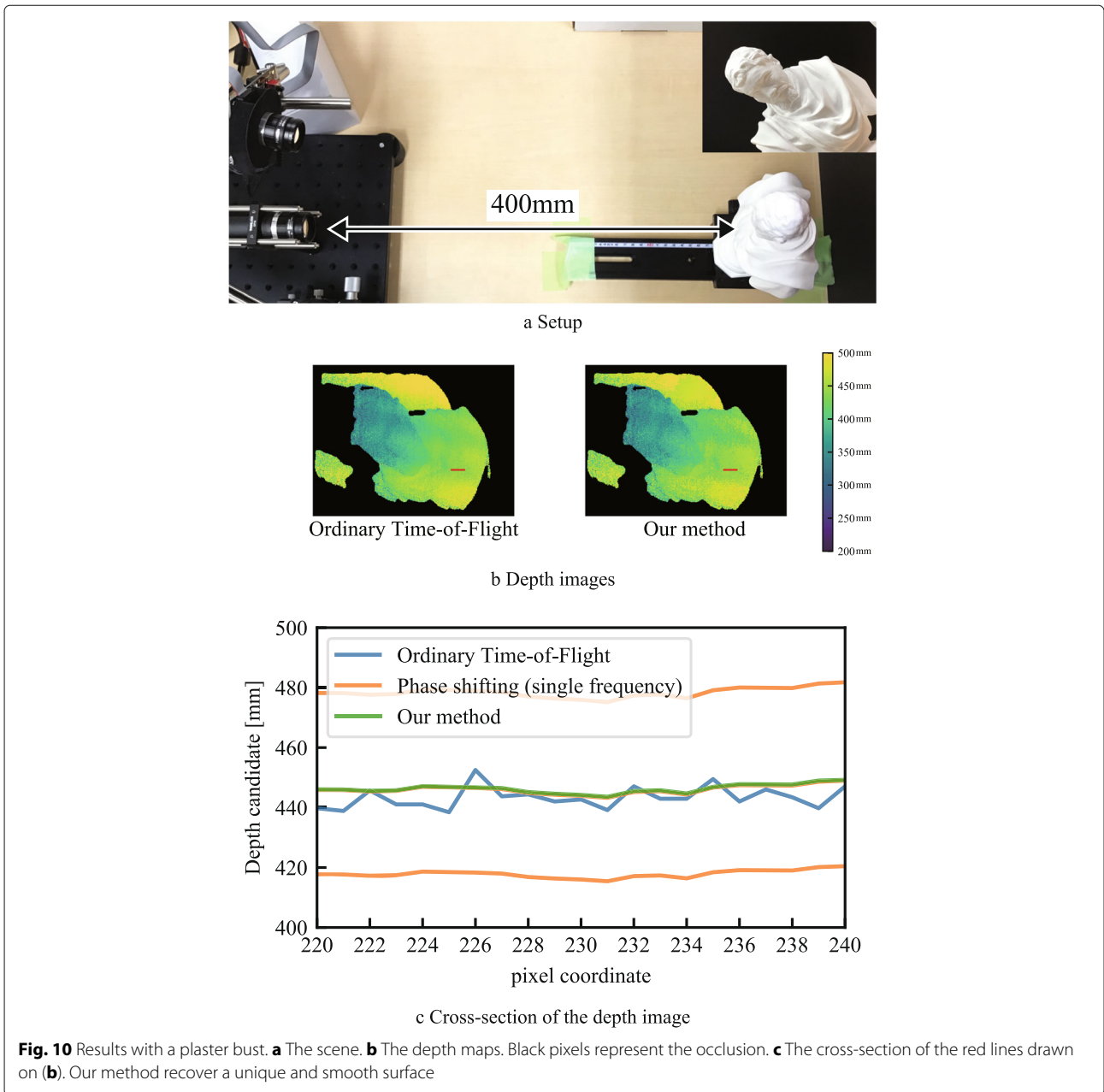


Fig. 10 Results with a plaster bust. **a** The scene. **b** The depth maps. Black pixels represent the occlusion. **c** The cross-section of the red lines drawn on **(b)**. Our method recover a unique and smooth surface

7 Conclusion

We developed a depth sensing method that uses spatio-temporally modulated illumination. We showed that the phase ambiguities of the temporal and spatial modulations are different, so it is possible to effectively resolve the ambiguities while reducing the observations and preserving the depth resolution.

Our proposed method inherits not only the strength of time-of-flight camera and active stereo using projector-camera system but also the weakness of them. While the proposed method can archive better resolution and wider range of interest, it may suffer from occlusion, which scarifies the ToF camera's potential. However, in practice, the current ToF camera is not a co-axial setup and it does not much suffer from occlusion. If the spatial-temporal projector is configured in the micro-baseline setup similarly to a ToF camera, the system does not much suffer from occlusion.

In this paper, depths of the ToF measurement are defined as the distance between a camera and a target; on the other hand, depths of the projector-camera system of phase shifting is defined as the distance between a center of baseline and a target. In practice, the difference should be correct for implementation although this is not affected to our key idea. Indeed, this model mismatch is absorbed by calibration step to build a look up table.

Our hardware prototype has some limitations. Because the DMD produces the sinusoidal pattern by controlling the mirrors on and off, it can make artifacts to the ToF. We ignored this effect, but it should be considered to control the DMD or to use a solid spatial light modulator appropriately. The quality of the spatio-temporally modulated illumination of our prototype is not very high. The temporal phase contains a systematic distortion, and the spatial resolution of the projector is currently limited to 64 pixels on the DMD, corresponding to 4 pixels on the camera, because the pattern is blurred. This might be due to the collimation and the alignment accuracy of the optics or the diffraction on the DMD. The light source cannot emit a spatial pattern that is equal to or less than the camera pixel's size, resulting in diminished phase shifting. In future implementations, we will develop a better light source unit to improve the temporal phase measurements and generate higher spatial resolutions.

Appendix

Derivation of Eq. (30)

We reshow Eq. (29) for the derivation of Eq. (30) as

$$d_S|_{n_S=n'_S} - d_S|_{n_S=n'_S-1} = \frac{\Delta d_T}{2}. \quad (29)$$

$d_S|_{n_S=n'_S-1}$ is the neighbor depth candidate as

$$d_S|_{n_S=n'_S-1} = \frac{bf}{x - \frac{2\pi(n'_S-1)+\phi_S}{\omega_S}}. \quad (A.1)$$

The unwrapped depth $d_S|_{n_S=n'_S}$ that satisfy Eq. (29) is the minimum depth of the range of interest d_{\min} as

$$d_S|_{n_S=n'_S} = \frac{bf}{x - \frac{2\pi n'_S + \phi_S}{\omega_S}} = d_{\min}. \quad (A.2)$$

Substituting Eqs. (A.1) and (A.2) to Eq. (29),

$$d_{\min} - \frac{bf}{x - \frac{2\pi(n'_S-1)+\phi_S}{\omega_S}} = \frac{\Delta d_T}{2},$$

$$d_{\min} - \frac{bf}{x - \frac{2\pi n'_S + \phi_S}{\omega_S} - \frac{2\pi}{\omega_S}} = \frac{\Delta d_T}{2}. \quad (A.3)$$

Substituting Eq. (A.2) to the denominator part,

$$d_{\min} - \frac{bf}{\frac{bf}{d_{\min}} - \frac{2\pi}{\omega_S}} = \frac{\Delta d_T}{2}. \quad (A.4)$$

Multiplying both sides of the equation by $\frac{bf}{d_{\min}} - \frac{2\pi}{\omega_S}$ and rearranging the equation,

$$d_{\min} \left(\frac{bf}{d_{\min}} - \frac{2\pi}{\omega_S} \right) - bf = \frac{\Delta d_T}{2} \left(\frac{bf}{d_{\min}} - \frac{2\pi}{\omega_S} \right) \quad (A.5)$$

$$d_{\min}^2 - \frac{\Delta d_T}{2} d_{\min} + \frac{\omega_S}{2\pi} \frac{\Delta d_T}{2} bf = 0. \quad (A.6)$$

Solving the quadratic equation for d_{\min} , we obtain

$$d_{\min} = \frac{\Delta d_T}{4} + \frac{1}{2} \sqrt{\frac{\Delta d_T^2}{4} + \frac{\omega_S bf \Delta d_T}{\pi}}, \quad (30)$$

where the other solution is always negative and the out of range of $d_{\min} > 0$.

Derivation of Eq. 32

Substituting Eqs. (10) and (26) to Eq. (31),

$$\frac{c}{2\omega_T} (2\pi n_T + \phi_T) - \frac{c}{2\omega_T} (2\pi(n_T - 1) + \phi_T) \quad (A.7)$$

$$= \frac{1}{2} \frac{2\pi d_{\max}^2 \sqrt{B}}{bf\omega_S 2\sqrt{8A}},$$

$$\frac{c\pi}{\omega_T} = \frac{\pi d_{\max}^2 \sqrt{B}}{bf\omega_S 2\sqrt{8A}}. \quad (A.8)$$

Rearranging the equation,

$$d_{\max}^2 = \frac{\omega_S bfc 2\sqrt{8A}}{\omega_T \sqrt{B}}. \quad (A.9)$$

Substituting Eq. (25) to Eq. (A.9) to cancel A and B ,

$$d_{\max}^2 = \frac{\omega_S bfc^2 \pi}{\omega_T^2 \Delta d_T}. \quad (A.10)$$

Therefore,

$$d_{\max} = \sqrt{\frac{\omega_s b f c^2 \pi}{\omega_T^2 \Delta d T}}, \quad (32)$$

because $d_{\max} > 0$.

Abbreviations

ToF: Time-of-flight; DMD: Digital mirror device

Acknowledgements

We thank all the people who gave us various insightful and constructive comments.

Authors' contributions

TK contributed to the concept, conducted experiments, and wrote the manuscript; KT and TA contributed to the concept and optical design and edited the manuscript; and TF and YM supervised the project and improved the representation. The authors reviewed and approved the final manuscript.

Funding

This work is partly supported by JST CREST JPMJCR1764 and JSPS Kaken grant JP18H03265 and JP18K11922.

Availability of data and materials

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma 630-0192, Japan. ²University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8577, Japan.

Received: 26 August 2019 Accepted: 10 March 2020

Published online: 07 April 2020

References

- Kushida T, Tanaka K, Takahito A, Funatomi T, Mukaigawa Y (2019) Spatio-temporal phase disambiguation in depth sensing. In: Proc. ICCP. <https://doi.org/10.1109/iccp.2019.8747338>
- Inokuchi S, Sato K, Matsuda F (1984) Range imaging system for 3-D object recognition. In: Proc. International Conference on Pattern Recognition. IEEE Computer Society Press, pp 806–808
- Salvi J, Fernandez S, Pribanic T, Llado X (2010) A state of the art in structured light patterns for surface profilometry. *Pattern Recog* 43. <https://doi.org/10.1016/j.patcog.2010.03.004>
- Gupta M, Nayer S (2012) Micro phase shifting. In: Proc. CVPR. IEEE, pp 813–820. <https://doi.org/10.1109/CVPR.2012.6247753>
- Mirdehghan P, Chen W, Kutulakos KN (2018) Optimal structured light à la carte. In: Proc. CVPR. <https://doi.org/10.1109/cvpr.2018.00654>
- Lange R, Seitz P (2001) Solid-state time-of-flight range camera. *IEEE J Quantum Electron* 37(3):390–397
- Yasutomi K, Usui T, Han S.-m., Takasawa T, Keiichiro K, Kawahito S (2016) A submillimeter range resolution time-of-flight. *IEEE Trans Electron Devices* 63(1):182–188
- Heide F, Xiao L, Kolb A, Hullin MB, Heidrich W (2014) Imaging in scattering media using correlation image sensors and sparse convolutional coding. *Opt Express* 22(21):26338–50
- Kirmani A, Benedetti A, Chou PA (2013) Spumic: simultaneous phase unwrapping and multipath interference cancellation in time-of-flight cameras using spectral methods. In: IEEE International Conference on Multimedia and Expo (ICME), pp 1–6. <https://doi.org/10.1109/icme.2013.6607553>
- Freedman D, Krupka E, Smolin Y, Leichter I, Schmidt M (2014) SRA: Fast Removal of General Multipath for ToF Sensors. In: Proc. ECCV, pp 1–15. https://doi.org/10.1007/978-3-319-10590-1_16
- Qiao H, Lin J, Liu Y, Hullin MB, Dai Q (2015) Resolving transient time profile in ToF imaging via log-sum sparse regularization. *Opt Lett* 40(6):918–21
- Kadambi A, Schiel J, Raskar R (2016) Macroscopic interferometry: rethinking depth estimation with frequency-domain time-of-flight. In: Proc. CVPR, pp 893–902. <https://doi.org/10.1109/cvpr.2016.103>
- Marco J, Hernandez Q, Muñoz A, Dong Y, Jarabo A, Kim MH, Tong X, Gutierrez D (2017) DeepToF: off-the-shelf real-time correction of multipath interference in time-of-flight imaging. *ACM Trans Graph* 36(6):219–121912. <https://doi.org/10.1145/3130800.3130884>
- Tanaka K, Mukaigawa Y, Funatomi T, Kubo H, Matsushita Y, Yagi Y (2018) Material classification from time-of-flight distortions. *IEEE TPAMI*. <https://doi.org/10.1109/tpami.2018.2869885>
- Su S, Heide F, Wetzstein G, Heidrich W (2018) Deep end-to-end time-of-flight imaging. In: Proc. CVPR. <https://doi.org/10.1109/cvpr.2018.00668>
- Velten A, Willwacher T, Gupta O, Veeraraghavan A, Bawendi MG, Raskar R (2012) Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nat Commun* 3(745). <https://doi.org/10.1038/ncomms1747>
- Heide F, Hullin MB, Gregson J, Heidrich W (2013) Low-budget transient imaging using photonic mixer devices. *ACM ToG* 32(4):1
- Kitano K, Okamoto T, Tanaka K, Aoto T, Kubo H, Funatomi T, Mukaigawa Y (2017) Recovering temporal PSF using ToF camera with delayed light emission. *IP SJ Trans Comput Vis Appl* 9(15). <https://doi.org/10.1186/s41074-017-0026-3>
- Kadambi A, Whyte R, Bhandari A, Streeter L, Barsi C, Dorrington A, Raskar R (2013) Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM ToG* 32(6):1–10
- O'Toole M, Heide F, Xiao L, Hullin MB, Heidrich W, Kutulakos KN (2014) Temporal frequency probing for 5D transient analysis of global light transport. *ACM ToG* 33(4):1–11
- O'Toole M, Heide F, Lindell D, Zang K, Diamond S, Wetzstein G (2017) Reconstructing transient images from single-photon sensors. In: Proc. CVPR. <https://doi.org/10.1109/cvpr.2017.246>
- Naik N, Kadambi A, Rhemann C, Izadi S, Raskar R, Bing Kang S (2015) A light transport model for mitigating multipath interference in time-of-flight sensors. In: Proc. CVPR, pp 73–81. <https://doi.org/10.1109/cvpr.2015.7298602>
- Nayar SK, Krishnan G, Grossberg MD, Raskar R (2006) Fast separation of direct and global components of a scene using high frequency illumination. *ACM ToG* 25(3):935–944
- Whyte R, Streeter L, Cree MJ, Dorrington AA (2015) Resolving multiple propagation paths in time of flight range cameras using direct and global separation methods. *Opt Eng* 54:54–549. <https://doi.org/10.1117/1.OE.54.11.113109>
- Agresti G, Zanuttigh P (2018) Combination of spatially-modulated ToF and structured light for MPI-free depth estimation. In: ECCV Workshop on 3D Reconstruction in the Wild. IEEE. https://doi.org/10.1007/978-3-030-11009-3_21
- Gupta M, Velten A, Nayar SK, Breitbach E (2018) What are optimal coding functions for time-of-flight imaging?. *ACM ToG* 37(2):13–11318. <https://doi.org/10.1145/3152155>
- Gutierrez-Barragan F, Reza S, Velten A, Gupta M (2019) Practical coding function design for time-of-flight imaging. In: Proc. CVPR. <https://doi.org/10.1109/cvpr.2019.00166>
- Kadambi A, Taamazyan V, Shi B, Raskar R (2015) Polarized 3D: high-quality depth sensing with polarization cues. In: Proc. ICCV, pp 3370–3378. <https://doi.org/10.1109/iccv.2015.385>
- Gkioulekas I, Levin A, Durand F, Zickler T (2015) Micron-scale light transport decomposition using interferometry. *ACM ToG* 34(4):37–13714
- Li F, Willomitzer F, Rangarajan P, Gupta M, Velten A, Cossairt O (2018) Sh-tof: micro resolution time-of-flight imaging with superheterodyne interferometry. In: Proc. ICCP. <https://doi.org/10.1109/iccp.2018.8368473>
- Maeda T, Kadambi A, Schechner YY, Raskar R (2018) Dynamic heterodyne interferometry. In: Proc. ICCP. IEEE. <https://doi.org/10.1109/ICCPHOT.2018.8368471>
- Sagawa R, Kawasaki H, Furukawa R, Kiyota S (2011) Dense one-shot 3D reconstruction by detecting continuous regions with parallel line projection. In: Proc. ICCV. <https://doi.org/10.1109/iccv.2011.6126460>

33. Sagawa R, Sakashita K, Kasuya N, Kawasaki H, Furukawa R, Yagi Y (2012) Grid-based active stereo with single-colored wave pattern for dense one-shot 3D scan. In: 3DIMPVT. pp 363–370. <https://doi.org/10.1109/3DIMPVT.2012.41>
34. Jayasuriya S, Pediredla A, Sivaramakrishnan S, Molnar A, Veeraraghavan A (2015) Depth fields: extending light field techniques to time-of-flight imaging. In: 2015 International Conference on 3D Vision. pp 1–9. <https://doi.org/10.1109/3DV.2015.8>
35. Heide F, Heidrich W, Hullin M, Wetzstein G (2015) Doppler time-of-flight imaging. *ACM ToG* 34(4):36–13611
36. McCormac J, Handa A, Leutenegger S, J.Davison A (2017) SceneNet RGB-D: can 5m synthetic images beat generic ImageNet pre-training on indoor segmentation? <https://doi.org/10.1109/iccv.2017.292>
37. Zhang Z (2000) A flexible new technique for camera calibration. *TPAMI* 22:1330–1334. <https://doi.org/10.1109/34.888718>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
