A Single-shot Multi-path Interference Resolution for Mirror-based Full 3D Shape Measurement with a Correlation-based ToF Camera

Shohei Nobuhara Takashi Kashino Takashi Matsuyama
Graduate School of Informatics, Kyoto University, Japan
{nob,kashino,tm}@vision.kuee.kyoto-u.ac.jp

Kouta Takeuchi Kensaku Fujii
NTT Media Intelligence Laboratories, NTT Corporation, Japan
{takeuchi.kouta,fujii.kensaku}@lab.ntt.co.jp

Abstract

This paper is aimed at presenting a new algorithm for multi-path interference resolutions under mirror-based full 3D capture using a single correlation-based ToF camera. Our algorithm does not require additional captures or device modifications, and resolves the interference using a single ToF sensing that is also used for the 3D reconstruction as well. Evaluations with real images prove the concept of the proposed algorithm qualitatively and quantitatively.

1. Introduction

The goal of this paper is to realize a full 3D shape measurement using a single correlation-based ToF (C-ToF) camera with multiple mirrors behind the object. By utilizing the virtual viewpoints defined by the mirrors, the C-ToF camera can observe the object from different directions and can fuse depth-maps into a single full 3D shape.

A challenge in this mirror-based C-ToF sensing is the multi-path interference of C-ToF illuminations. Since C-ToF cameras utilize *four-bucket sampling* [2, 13, 14] in which the sensor outputs cross-correlations between an amplitude modulated infrared illumination cast to and returned from the scene in order to calculate the depth from the phase delay, additional illumination paths defined by the mirrors result in an unreliable depth calculation.

The key idea to solve this problem is to model the multi-path illumination explicitly by calibrating the mirror geometry in advance. We show that once a depth is hypothesized for a pixel, the multi-path model allows us to evaluate the likelihood of the assumption.

Our contribution is twofold. Firstly, we propose a new algorithm which can find a depth where the cross-correlations returned by the sensor is best described by taking the multi-path illuminations into account. By utilizing this our method can estimate the full 3D object shape by alternatingly optimizing the depth and the normal. Secondly we propose a new mirror calibration which exploits the interference. Our method can estimate not only the mirror geometry but also its reflection coefficient simultaneously.

Since our method requires a single regular C-ToF sensing only, it can be used to capture dynamic objects without capturing additional frames or device modifications.

2. Related works

Image-based full 3D capture with mirrors has been studied in the literature for years [3, 9, 15, 19]. In the context of active stereo with mirrors, Lanman *et al*. proposed an orthographic pattern projection [15] and Tahara *et al*. proposed an epipole-centered structured patterns [19] by making patterns projected from different paths do not collide spatially on the object surface.

On the other hand, as described in [2, 7, 13], multi-path illuminations from the light source to a correlation-based ToF pixel result in a wrong estimation of the depth, since it depends on computing correlations between the temporally-modulated source and the received light signals. This problem is known as *multi-path interference*, and frequency-division or time-division multiplexing can be a practical solution if it is caused by modulated illuminations from different C-ToF cameras.

Otherwise the interference occurs due to multiple reflections from a single light source, and a major difficultly is to estimate the target 3D structure and the multiple light paths simultaneously since the multiple paths depends on the 3D structure which is unknown at the capture timing.

To solve this problem there exist several approaches which utilize special devices [17, 20, 21], special illumination patterns [4, 6, 8, 11, 12], Lambertian scene assump-

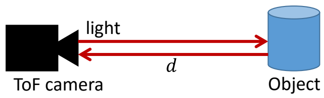


Figure 1: ToF camera. The camera and the light source do not coincide physically, but such displacement is assumed to be calibrated for each pixel.

tions [5, 10], *etc*. They utilize multiple captures of the scene or involve computationally expensive optimizations.

Compared with such studies, this paper proposes an algorithm that does not require special modifications to the ToF camera and the light source, similarly to [5, 10]. That is, our method can be operated only with the raw capture data of a single frame originally returned by off-the-shelf C-ToF cameras. The main difference with [5, 10] is the fact that our algorithm explicitly calibrates the mirrors that causes the interference, and the mirror calibration allows us a pixel-wise simple depth correction while [10] solves an inverse rendering to synthesize the observation. In addition, this paper proposes a mirror calibration utilizing the interference itself.

3. Measurement model

3.1. Correlation-based time-of-flight camera

As illustrated in Figure 1, the time-of-flight principle measures the distance to the target by measuring the time in which the light emitted from the camera travels at the speed c hits the target at d distance and then travels back to the camera in time τ:

$$d = c\tau/2. \tag{1}$$

Since measuring τ is not a trivial task in particular for short-range measurement, correlation-based ToF (C-ToF) cameras modulate the amplitude of the emitted light in the time domain, and measure the phase delay between the emitted and the received light signals. Let $S(t)$ denote the emitted light signal modulated at f Hz:

$$S(t) = \cos 2\pi f t. \tag{2}$$

The received light can be modeled as

$$G(t) = A\cos\left(2\pi f t + \psi\right) + B, \tag{3}$$

where A is the attenuation caused by the reflection and the distance the light traveled. B denotes the background illumination. ψ is the phase delay caused by the round-trip distance from the camera to the target:

$$\psi = 2\pi f\tau. \tag{4}$$

In order to estimate the phase ψ, C-ToF sensors utilize the correlation between $S(t)$ and $G(t)$:

$$\begin{aligned} C(\tau) &= S(t) * G(t) \\ &= \lim_{T\to\infty}\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} G(t)S(t+\tau)dt \\ &= \tfrac{A}{2}\cos(\psi - 2\pi f\tau) + B. \end{aligned} \tag{5}$$

That is, C-ToF sensors output four correlation values of Eq (5) where $2\pi f\tau = 0, \pi/2, \pi, 3\pi/2$:

$$C_0 = C(0) = \tfrac{A}{2}\cos(\psi - 0) = \tfrac{A}{2}\cos\psi + B, \tag{6}$$

$$C_1 = C(\tfrac{\pi}{2}) = \tfrac{A}{2}\cos(\psi - \tfrac{\pi}{2}) = \tfrac{A}{2}\sin\psi + B, \tag{7}$$

$$C_2 = C(\pi) = \tfrac{A}{2}\cos(\psi - \pi) = -\tfrac{A}{2}\cos\psi + B, \tag{8}$$

$$C_3 = C(\tfrac{3\pi}{2}) = \tfrac{A}{2}\cos(\psi - \tfrac{3\pi}{2}) = -\tfrac{A}{2}\sin\psi + B, \tag{9}$$

and estimate the parameters A, B and ψ in $G(t)$ by

$$\psi = 2\pi f\tau = \arctan\left(\frac{C_3 - C_1}{C_2 - C_0}\right), \tag{10}$$

$$A = \frac{1}{2}\sqrt{(C_3 - C_1)^2 + (C_2 - C_0)^2}, \tag{11}$$

$$B = \frac{1}{4}(C_3 + C_2 + C_1 + C_0). \tag{12}$$

This algorithm is called *four-bucket sampling* [2, 13, 14], and we propose an extension of this for multi-path cases.

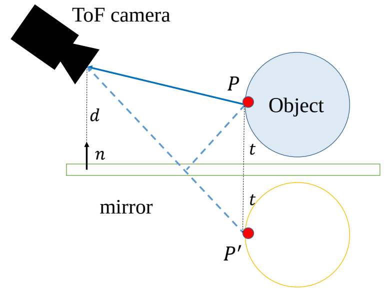


Figure 2: Mirror-based multi-view measurement

3.2. Mirror-based multi-view imaging

Figure 2 illustrates the measurement model. A C-ToF camera observes the target at $\boldsymbol{p}$ directly and also indirectly at $\boldsymbol{p}'$ via a mirror of normal $\boldsymbol{n}$ and the distance d from the camera center. Here the point $\boldsymbol{p}'$ is the mirror of $\boldsymbol{p}$ given by:

$$\boldsymbol{p} = \boldsymbol{p}' + 2t\boldsymbol{n}, \tag{13}$$

where t denotes the distance from $\boldsymbol{p}$ to the mirror plane. The projection of $\boldsymbol{p}'$ to $\boldsymbol{n}$ gives

$$t + d = -\boldsymbol{n}^\top \cdot \boldsymbol{p}'. \tag{14}$$

From these two equations, we obtain

$$\boldsymbol{p} = -2(\boldsymbol{n}^\top \cdot \boldsymbol{p}' + d)\boldsymbol{n} + \boldsymbol{p}'. \tag{15}$$

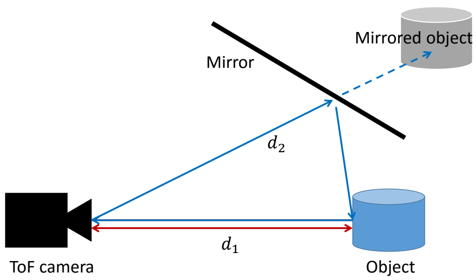


Figure 3: Interference by direct and indirect illuminations. The direct illumination (red) travels $2d_1$, and the indirect one (blue) travels $d_1 + d_2$.

This can be rewritten as

$$\tilde{\boldsymbol{p}} = S\tilde{\boldsymbol{p}}' = \begin{bmatrix} H & -2d\boldsymbol{n} \\ \boldsymbol{0}_{1\times 3} & 1 \end{bmatrix} \tilde{\boldsymbol{p}}', \tag{16}$$

where $H = I_{3\times 3} - 2\boldsymbol{n}\boldsymbol{n}^\top$ is Householder matrix and $\tilde{x}$ denotes the homogeneous coordinate of x.

3.3. C-ToF interference by direct and indirect illuminations

As illustrated in Figure 3, suppose a C-ToF camera pixel observes a point on the object surface such that it is illuminated directly by the light source (red) as well as indirectly via a mirror (blue). In this case the reflected light can be modeled as

$$\begin{aligned} G(t) &= G_1(t) + G_2(t) \\ &= A_1 \cos(2\pi f t + \psi_1) + A_2 \cos(2\pi f t + \psi_2) + B, \end{aligned} \tag{17}$$

and the correlation can be given by

$$\begin{aligned} C(\tau) &= S(t) * (G_1(t) + G_2(t)) \\ &= S(t) * G_1(t) + S(t) * G_2(t) \\ &= \tfrac{A_1}{2}\cos(\psi_1 - 2\pi f \tau) + \tfrac{A_2}{2}\cos(\psi_2 - 2\pi f \tau) + B. \end{aligned} \tag{18}$$

Hence the four correlation values at $2\pi f \tau = 0, \pi/2, \pi, 3\pi/2$ satisfy

$$\begin{bmatrix} \cos\psi_1 & \cos\psi_2 & 1 \\ \sin\psi_1 & \sin\psi_2 & 1 \\ -\cos\psi_1 & -\cos\psi_2 & 1 \\ -\sin\psi_1 & -\sin\psi_2 & 1 \end{bmatrix} \begin{bmatrix} \frac{A_1}{2} \\ \frac{A_2}{2} \\ B \end{bmatrix} = \begin{bmatrix} C_0 \\ C_1 \\ C_2 \\ C_3 \end{bmatrix}. \tag{19}$$

From this equation, we obtain $B = \frac{1}{4}(C_0 + C_1 + C_2 + C_3)$ as done in Eq (12) and

$$\begin{bmatrix} \cos\psi_1 & \cos\psi_2 \\ \sin\psi_1 & \sin\psi_2 \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} C_0 - C_2 \\ C_1 - C_3 \end{bmatrix}, \tag{20}$$

by subtracting the even rows and odd rows in Eq (19).

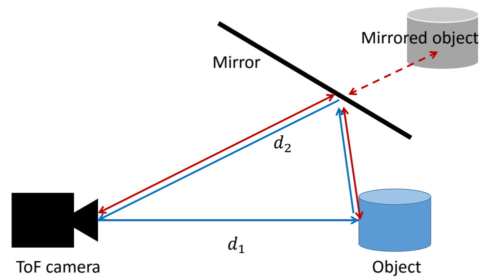


Figure 4: Four-bucket sampling on mirrored image pixels. The direct illumination to the mirrored object (red) travels $2d_2$, and the indirect one to the same target point (blue) travels $d_1 + d_2$.

3.3.1 Four-bucket sampling on real image pixels

Suppose the mirror parameter S in Eq (16) and also the intrinsic parameter K of the C-ToF camera are calibrated beforehand. If we assume the distance to the object projected to a pixel $\boldsymbol{q}$ is d_1, then we can obtain the distance to its mirror as follows.

The position of a 3D point projected to a pixel $\boldsymbol{q}$ of the C-ToF camera is given by

$$\boldsymbol{p} = \frac{d_1 K^{-1}\tilde{\boldsymbol{q}}}{\|K^{-1}\tilde{\boldsymbol{q}}\|}, \tag{21}$$

where d_1 is the distance from the camera to the 3D point, and its mirror is given by

$$\tilde{\boldsymbol{p}}' = S\tilde{\boldsymbol{p}}. \tag{22}$$

Hence by computing the length of $\boldsymbol{p}'$ as d_2, the round trip distances of the direct and the indirect illuminations are given as $2d_1$ and $d_1 + d_2$.

Once obtained such distances, then Eqs (1) and (4) return the corresponding delays τ_i and the phases $\psi_i (i = 1, 2)$ immediately. Therefore, the left-side 2×2 matrix of Eq (20) can be expressed as a function of d_1:

$$\Psi(d_1) = \begin{bmatrix} \cos\psi_1 & \cos\psi_2 \\ \sin\psi_1 & \sin\psi_2 \end{bmatrix}, \tag{23}$$

and we can solve Eq (20) for A_1 and A_2 as

$$\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \Psi^{-1}(d_1) \begin{bmatrix} C_0 - C_2 \\ C_1 - C_3 \end{bmatrix}, \tag{24}$$

for the hypothesized d_1.

3.3.2 Four-bucket sampling on mirrored image pixels

Similarly to Section 3.3.1, we can consider a four-bucket sampling for pixels corresponding to mirrored images as

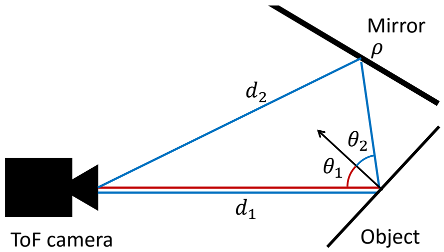


Figure 5: Multi-path illuminations. The direct light (red) illuminates the object with the angle θ_1, and the indirect one (blue) illuminates with the angle θ_2 via the mirror of the reflectance coefficient ρ.

illustrated in Figure 4. In this case the target is illuminated as same as Figure 3, but the reflected light is captured by the camera via the mirror. Therefore, the two light path lengths are given as $d_1 + d_2$ and $2d_2$.

4. Analysis by synthesis approach

Eq (24) returns A_1 and A_2 once a depth d_1 to the target is hypothesized. In other words, this equation itself cannot determine if the hypothesized d_1 was correct or not. This section proposes an algorithm that seeks the best depth d_1 for each pixel by examining a *goodness* of the returned A_1 and A_2 for possible d_1 candidates by introducing a surface reflection model.

As illustrated in Figure 5, both the direct and the indirect illuminations are delivered to the camera along the same path once they are reflected at the object surface. Hence the difference on A_1 and A_2 can be modeled by considering how they illuminate the object from different directions.

By considering the illumination attenuation by the inverse square of the distance and the reflectance coefficients of the mirror and the surface, the difference on A_1 and A_2 can be modeled as

$$A_2 = \frac{\mathcal{R}_2}{\mathcal{R}_1}\frac{d_1^2}{d_2^2}\rho A_1, \tag{25}$$

where ρ is the reflectance coefficient of the mirror and $\mathcal{R}_1$ and $\mathcal{R}_2$ are the reflectance coefficients of the object surface for each illumination. d_1 and d_2 are the direct and the indirect distances from the light to the surface.

While there exist many analytical [18] and learning-based [16] reflection models, we here employ Lambertian model in which the reflection depends only on the incident light angle:

$$A_2 = \frac{\cos\theta_2}{\cos\theta_1}\frac{d_1^2}{d_2^2}\rho A_1, \tag{26}$$

where θ_1 and θ_2 denote the incident angles to the object surface (Figure 5).

Notice that Eq (26) requires the object surface normal to compute the incident angles while the 3D shape itself is unknown. This is a chicken-and-egg problem and we solve this by an iterative approach as described later. In other words, our algorithm assumes the normals to be given in estimating the depth, and considers that $\frac{\cos\theta_2}{\cos\theta_1}$ is also a function of d_1 for a particular pixel $\boldsymbol{q}$.

As a result, Eq (26) can be expressed as

$$A_2 = \alpha(d_1)A_1, \tag{27}$$

and hence Eq (20) can be rewritten as

$$\begin{bmatrix}\cos\psi_1 & \cos\psi_2\\ \sin\psi_1 & \sin\psi_2\end{bmatrix}\begin{bmatrix}A_1\\ \alpha(d_1)A_1\end{bmatrix} = \Psi(d_1)\begin{bmatrix}A_1\\ \alpha(d_1)A_1\end{bmatrix} = \begin{bmatrix}C_0 - C_2\\ C_1 - C_3\end{bmatrix}. \tag{28}$$

This is an overconstrained system for A_1 and the MLE is given by

$$A_1^* = \frac{1}{2}\left(\frac{C_0 - C_2}{\cos\psi_1 + \alpha(d_1)\cos\psi_2} + \frac{C_1 - C_3}{\sin\psi_1 + \alpha(d_1)\sin\psi_2}\right). \tag{29}$$

Inversely we can evaluate how the hypothesized d_1 as well as the surface normal are correct by measuring

$$E(d_1) = \left\| \Psi(d_1)\begin{bmatrix}A_1^*\\ \alpha(d_1)A_1^*\end{bmatrix} - \begin{bmatrix}C_0 - C_2\\ C_1 - C_3\end{bmatrix}\right\|. \tag{30}$$

The same discussion can be applied for mirrored image pixels in Section 3.3.2, by substituting Eq (26) with

$$A_2 = \frac{\cos\theta_1}{\cos\theta_2}\frac{d_2^2}{d_1^2}\frac{1}{\rho}A_1. \tag{31}$$

5. Mirror-based full 3D shape measurement

Given the objective function $E(d_1)$ in Eq (30), we can introduce an algorithm that alternatingly estimates the 3D shape as well as its normal as follows.

Step 1. Capture the object with mirrors, and obtain four correlation values for each pixel.

Step 2. Initialize the normal corresponding to each pixel by using the depth-map without considering interference.

Step 3. Find the best depth for each pixel by seeking the best d_1 that minimizes $E(d_1)$ (Eq (30)).

Step 4. Update the normal using the estimated depth-map and repeat Step 3 until the estimated depth converges.

In computing the normal for each pixel, we simply assumed that it is approximated by the average of the surface normals of four 3D triangles defined by each pixel and its four neighboring pixels. Let $\boldsymbol{p}(\boldsymbol{q})$ denote the 3D point corresponding to a C-ToF pixel $\boldsymbol{q}$. By denoting the four neighboring pixels of a pixel $\boldsymbol{q}$ in question as $\boldsymbol{q}_i (i = 0, \ldots, 3)$ in CCW order in the image plane, the surface normal at $\boldsymbol{p}(\boldsymbol{q})$ is approximated by

$$\frac{1}{4} \sum_{i=0}^{3} \frac{\boldsymbol{e}_i \times \boldsymbol{e}_{i+1 \bmod 4}}{\|\boldsymbol{e}_i \times \boldsymbol{e}_{i+1 \bmod 4}\|}, \tag{32}$$

where $\boldsymbol{e}_i = \boldsymbol{p}(\boldsymbol{q}_i) - \boldsymbol{p}(\boldsymbol{q})$.

6. Mirror calibration

Obviously the 3D shape estimation quality by the proposed algorithm depends on the quality of the mirror parameters, *i.e.* its geometry S in Eq (16) as well as the reflectance coefficient ρ in Eq (26). This section introduces a new calibration method of these parameters by using a reference object that follows the reflection model assumed in Section 4.

Suppose a reference object with well-localizable feature points, such as a chess board, is located where the capture target is supposed to be positioned, and it is captured by the fixed C-ToF camera with and without mirrors. The capture without the mirrors returns the ground truth 3D geometry of the feature points since it does not involve any interferences. The capture with mirrors returns a 3D geometry of the same feature points distorted due to the interference.

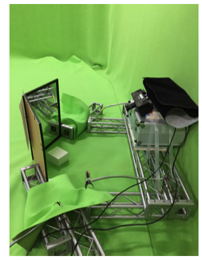
The goal of the calibration is to estimate the optimal mirror parameter such that it minimizes Eq (30) for the depth given by the ground truth 3D geometry, and it can be formulated as a non-linear optimization problem as follows.

Let the number of feature points found in the reference object be N, the ground truth depth for i-th feature point be $\hat{d}_1^i$. The calibration parameters consist of the mirror normal $\boldsymbol{n}$, the distance d, and the reflectance coefficient ρ. Given $\boldsymbol{n}$ and d, we can compute the mirrored position of each ground truth 3D points by Eq (16), and the corresponding distance d_2. Hence $\Psi(d_1)$ in Eq (30) can be expressed as $\Psi(\boldsymbol{n}, d)$.

Similarly since $\alpha(d_1)$ is a function of ρ, d_1 and d_2, it can be expressed as $\alpha(\boldsymbol{n}, d, \rho)$. Here the ground truth 3D geometry of the reference object can be used to provide the object surface normal to compute θ_1 and θ_2 for $\alpha(\boldsymbol{n}, d, \rho)$ (Eq (26)).

As a result, Eq (30) can be used as an error function of $\boldsymbol{n}$, d, and ρ for each of N feature points, and their optimal values can be given by solving:

$$\min_{\boldsymbol{n}, d, \rho} \sum_{i=1}^{N} \left(E(\boldsymbol{n}, d, \rho)|_{d_1 = \hat{d}_1^i} + E'(\boldsymbol{n}, d, \rho)|_{d_1 = \hat{d}_1^i} \right), \tag{33}$$


(a) Side-view


(b) Top-view

Figure 6: Capture environment

Figure 7: Captured ToF image

where $E(\boldsymbol{n}, d, \rho)|_{d_1 = \hat{d}_1^i}$ and $E'(\boldsymbol{n}, d, \rho)|_{d_1 = \hat{d}_1^i}$ denote the errors for i-th feature point on the pixels corresponding to the real and mirrored images (Eq (27)).

7. Evaluations

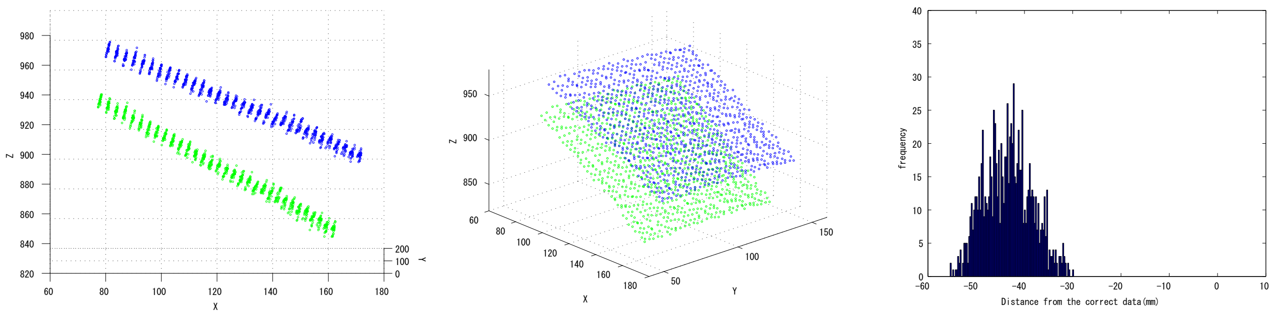
7.1. Environment

As shown in Figure 6 we used a C-ToF camera (PMD CamCube 3.0, 200×200 resolution, 20 MHz modulation frequency) to capture the object. The two mirrors are located behind the object with approximately 120°, and they are calibrated by the algorithm in Section 6.

7.2. Quantitative evaluations

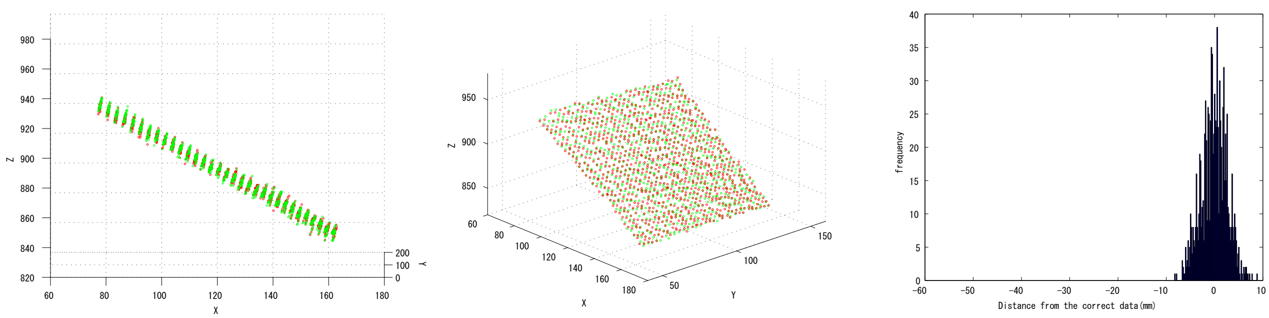
To evaluate the proposed algorithm quantitatively, this section shows results for a flat chessboard captured with a single mirror (the left-side mirror in Figure 6). Figure 7 shows the image captured by the ToF camera. This is the amplitude image given as A by Eq (11). The right and the left patterns in this image correspond to the real and its mirrored images of the board.

Figures 8 and 10 show results returned by the original four-bucket sampling algorithm without interference resolutions. The former corresponds to the real image (the right-side pattern of Figure 7), and the latter corresponds to the mirrored image (the left-side of Figure 7). In these figures, (a) and (b) clearly shows that the 3D points captured with the mirror (blue points) are largely distorted from the ground truth (green and red points). Here the ground truth 3D points in the real image are captured by removing the



(a) Side-view (b) Top-view (c) Histogram

Figure 8: 3D geometry returned by the ToF camera under interferences (real images). Green: ground truth, blue: under interference.



(a) Side-view (b) Top-view (c) Histogram

Figure 9: 3D geometry corrected by the proposed method (real images). Green: ground truth, red: proposed.

mirror from the scene to eliminate multi-path illuminations. For the mirror image side, the estimated 3D points are mirrored to the real side to be compared with the ground truth. Figures (c) show the histograms of displacements w.r.t. the ground truth.

The red points in Figure 9 and the green points in Figure 11 show results by the proposed method. The green and the red points are the ground truth. Compared with Figures 8 and 10, (a) and (b) prove the displacements are well corrected qualitatively. This point is also verified quantitatively by the histograms in (c) in which the residual displacements are approximately centered at zero.

7.3. Full 3D reconstruction

Figure 13 shows two objects *Prism* and *Box* used to evaluate our method in terms of full 3D shape capture with mirrors.

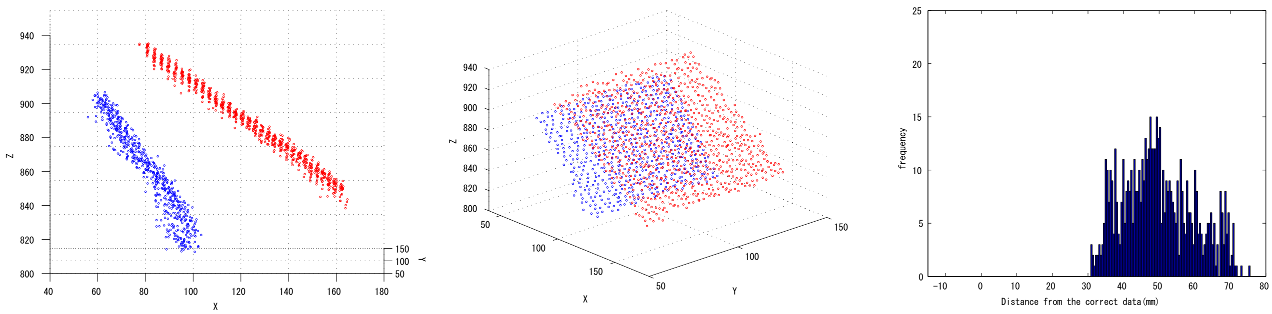
Figure 13(a) shows the 3D points captured without the mirrors. While it includes areas visible from the original camera viewpoint only, we use this as a reference. Figure 13(b) shows the result by the proposed method. The blue and the red points are the estimated 3D points from the mirrored image regions. Compared with the reference, this proves the concept of our interference resolution.

Figure 14 shows the result of the *Box* object. The yellow points in (b) correspond to the top-surface labeled by hand and their 3D positions are identical to the points in (a). Compared with the *Prism* object case, Figure 14(b) clearly shows that the estimation fails completely. This is because the top-surface of the object is illuminated by three paths (one direct and two indirect illuminations) that is not assumed in Eq (17).

7.4. Interference by floor

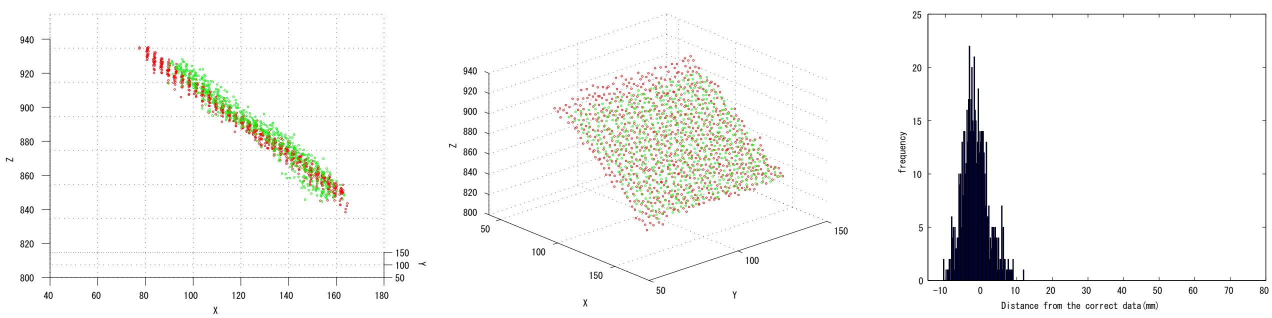
As shown in Figure 15, interferences can be introduced not only by mirrors but also regular surfaces such as the floor. Figures 16 and 17 show the 3D points of the upright surface in Figure 15 by the regular four-bucket sampling and the proposed method.

Here the green points are the reference points captured with a black-out curtain over the white floor surface. The floor geometry and the reflectance is calibrated by capturing the floor with and without the object to obtain the correlations used in Section 6.



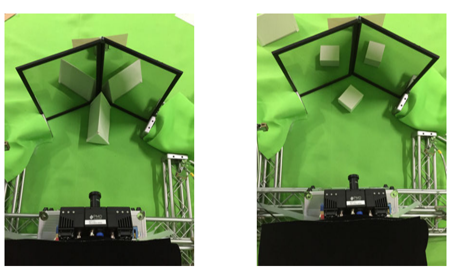
(a) Side-view (b) Top-view (c) Histogram

Figure 10: 3D geometry returned by the ToF camera under interferences (mirrored images). Red: ground truth, blue: under interferences.



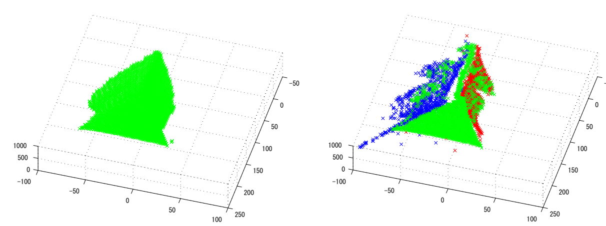
(a) Side-view (b) Top-view (c) Histogram

Figure 11: 3D geometry corrected by the proposed method (mirrored images). Red: ground truth, green: proposed.



(a) *Prism* object (b) *Box* object

Figure 12: Capture targets



(a) Reference (b) Proposed

Figure 13: 3D reconstruction of *prism*

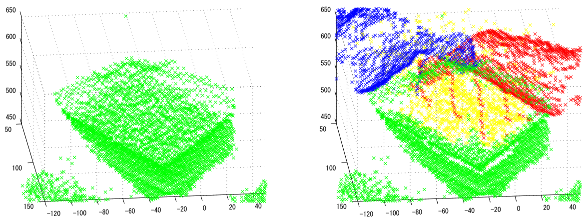
Figure 17 proves that the proposed method can reasonably resolve the interference caused not only by mirrors but also by regular surfaces including the floor.

8. Discussions

Degenerated case Eq (24) assumes that the 2×2 matrix $\Psi(d_1)$ of Eq (23) is invertible. This assumption holds as long as its determinant is not zero. Since the determinant is given by

$$\begin{aligned}
\det(\Psi(d_1)) &= \cos(\psi_1)\sin(\psi_2) - \cos(\psi_2)\sin(\psi_1) \\
&= \frac{1}{2}\left(\sin(\psi_1 + \psi_2) - \sin(\psi_1 - \psi_2)\right) \\
&\quad - \frac{1}{2}\left(\sin(\psi_2 + \psi_1) - \sin(\psi_2 - \psi_1)\right) \\
&= \sin(\psi_2 - \psi_1),
\end{aligned} \tag{34}$$

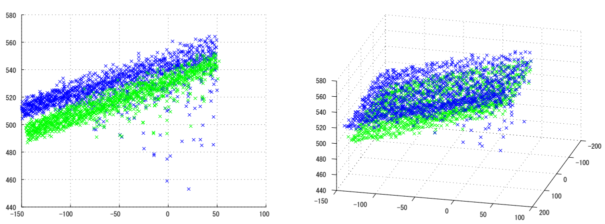
it can be zero if $\psi_2 = \psi_1 + m\pi$ ($m \in \mathbb{N}$). This indicates that the proposed algorithm does not work if the difference of



(a) Reference (b) Proposed

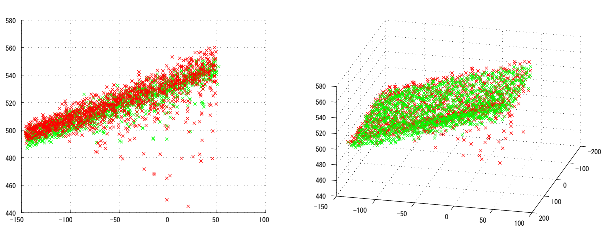
Figure 14: Failure case: 3D reconstruction of *box*

Figure 15: Interference by floor



(a) Side view (b) Top view

Figure 16: Without interference resolution (green: reference, blue: conventional)



(a) Side view (b) Top view

Figure 17: With interference resolution (green: reference, red: proposed)

the direct and indirect path lengths is $\frac{mc}{4f}$.

Surface Reflection Model Section 4 assumed the Lambertian reflection model to obtain Eq (30). This reflection model can be replaced with other models [16, 18] according to the capture targets as long as it depends on the incident and reflected light angles, since the algorithm in Section 5 alternatingly updates the object 3D shape and the normal.

Convergence The alternating algorithm in Section 5 is not proven to converge theoretically. It should depend on the initial estimate of the normal which is computed from the 3D estimate without interference resolutions, we did not observe such non-convergent cases in our experiments.

Smoothness Constraint The algorithm in Section 4 is a pixel-wise depth estimation method without explicit smoothness constraints between neighboring pixels. As done in many 3D reconstruction studies, we can render the problem into a MAP-MRF optimization with Eq (30) as the data term [1].

Simultaneous Estimation The objective function Eq (30) evaluates the depth of a pixel in question while computing its mirrored 3D position as a byproduct. This mirrored 3D point also has a corresponding C-ToF pixel by computing its projection, and we can evaluate the depth on that pixel by using Eq (27). That is, instead of estimating the depth for the real and the mirrored image pixels separately, we can consider estimating a consistent 3D depth for the both pixels simultaneously.

9. Conclusion

This paper proposed a new interference resolution algorithm for correlation-based ToF cameras without involving additional captures nor modifications to the devices.

While it has several points to be further studied as discussed in Section 8, we believe that this paper proves the concept of the single shot interference resolution for the mirror-based multi-view environment qualitatively and quantitatively.

Acknowledgment This research is partially supported by NTT Media Intelligence Laboratories and by JSPS Kakenhi Grant Number 26240023.

References

[1] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *IJCV*, 70(1):41–54, 2006.

[2] S. Foix, G. Alenya, and C. Torras. Lock-in time-of-flight (tof) cameras: a survey. *IEEE Sensors Journal*, 11(9):1917–1926, 2011.

[3] K. Forbes, F. Nicolls, G. D. Jager, and A. Voigt. Shape-from-silhouette with two mirrors and an uncalibrated camera. In *Proc. of ECCV*, 2006.

[4] D. Freedman, Y. Smolin, E. Krupka, I. Leichter, and M. Schmidt. Sra: Fast removal of general multipath for tof sensors. In *Proc. of ECCV*, pages 234–249, 2014.

[5] S. Fuchs. Multipath interference compensation in time-of-flight camera images. In *Proc. of ICPR*, pages 3583–3586, 2010.

[6] J. P. Godbaz, M. J. Cree, and A. A. Dorrington. Mixed pixel return separation for a full-field ranger. In *23rd International Conference Image and Vision Computing New Zealand*, pages 1–6, 2008.

[7] S. A. Gudmundsson, H. Aanæs, and R. Larsen. Environmental effects on measurement uncertainties of time-of-flight cameras. In *International Symposium on Signals, Circuits and Systems*, volume 1, pages 1–4, 2007.

[8] M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM Trans. on Graphics*, 34(5):156, 2015.

[9] K. H. Jang, D. H. Lee, and S. K. Jung. A moving planar mirror based approach for cultural reconstruction. *Computer Animation and Virtual Worlds*, 15(3-4):415–423, July 2004.

[10] D. Jiménez, D. Pizarro, M. Mazo, and S. Palazuelos. Modeling and correction of multipath interference in time of flight cameras. *Image and Vision Computing*, 32(1):1–13, 2014.

[11] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar. Coded time of flight cameras: Sparse deconvolution to address multipath interference and recover time profiles. *ACM Trans. on Graphics*, 32(6):167:1–167:10, Nov. 2013.

[12] A. Kirmani, A. Benedetti, and P. A. Chou. Spumic: Simultaneous phase unwrapping and multipath interference cancellation in time-of-flight cameras using spectral methods. In *IEEE International Conference on Multimedia and Expo*, pages 1–6, 2013.

[13] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-flight sensors in computer graphics. In *Proc. of Eurographics (State-of-the-Art Report)*, volume 6, 2009.

[14] R. Lange, P. Seitz, A. Biber, and R. Schwarte. Time-of-flight range imaging with a custom solid state image sensor. In *Industrial Lasers and Inspection (EUROPTO Series)*, pages 180–191. International Society for Optics and Photonics, 1999.

[15] D. Lanman, D. Crispell, and G. Taubin. Surround structured lightning: 3-d scanning with orthographic illumination. In *CVIU*, pages 1107–1117, November 2009.

[16] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. In *Proc. of SIGGRAPH*, pages 759–769, 2003.

[17] N. Naik, A. Kadambi, C. Rhemann, S. Izadi, R. Raskar, and S. Bing Kang. A light transport model for mitigating multipath interference in time-of-flight sensors. In *Proc. of CVPR*, June 2015.

[18] M. Oren and S. K. Nayar. Generalization of lambert's reflectance model. In *Proc. of SIGGRAPH*, pages 239–246. ACM, 1994.

[19] T. Tahara, R. Kawahara, S. Nobuhara, and T. Matsuyama. Interference-free epipole-centered structured light pattern for mirror-based multi-view active stereo. In *Proc. of 3DV*, pages 153–161, 2015.

[20] R. Whyte, A. Bhandari, L. Streeter, M. J. Cree, and A. A. Dorrington. Time frequency duality of time-of-flight range cameras for resolving multi-path interference. In *Proc. of the 29th International Conference on Image and Vision Computing New Zealand*, pages 247–252, 2014.

[21] V. A. C. Zeman. *Constructive Interference for Multi-view Time-of-flight Acquisition*. PhD thesis, 2012.