SUMMARY This paper proposes a novel Depth From Defocus (DFD) technique based on the property that two images having different focus settings coincide if they are reburred with the opposite focus setting, which is referred to as the "cross-reblurring" property in this paper. Based on the property, the proposed technique estimates the block-wise depth profile for a target object by minimizing the mean squared error between the cross-reburred images. Unlike existing DFD techniques, the proposed technique is free of lens parameters and independent of point spread function models. A compensation technique for a possible pixel disalignment between images is also proposed to improve the depth estimation accuracy. The experimental results and comparisons with the other DFD techniques show the advantages of our technique.

key words: depth from defocus (DFD), depth estimation, multi-focus images, blur, telecentric optics, pixel disalignment

1. Introduction

Depth From Focus (DFF) and Depth From Defocus (DFD) have been known as one of the depth estimation techniques [1]. The typical applications of DFF and DFD include 3-D space recognition in computer vision and 3-D shape reconstruction/modeling for objects. Among various depth estimation techniques such as binocular stereo vision, structure from motion [2], time-of-flight [3], and structured light [4], DFF and DFD are classified into an image-based passive approach because they utilize a set of images for a target object without emitting a controlled energy source. These techniques thus require no dedicated hardware/devices for implementation, which is the principal advantage of the techniques.

DFF and DFD utilize a set of images of a target object with their focus setting varied, and attempt to estimate the best focus setting for each pixel/block in the images, from which the absolute or a relative depth for the object can be estimated. DFF [5]–[8] attempts to find the best focus setting for densely acquired images while DFD [9]–[20] utilizes a small number of images, typically two or three images, to estimate the depth [5].

As reviewed in Sect. 2, DFD techniques proposed in the literature [9]–[22] involve their own disadvantages, which are roughly summarized as:

(a) Limitation of the PSF model
(b) Requirement of camera parameters
(c) Insufficient accuracy in 3-D shape recovery.

To overcome these disadvantages, this paper proposes a novel DFD approach based on the cross-reblurring property elaborated in Sect. 3, and demonstrates its advantages over existing techniques.

A variation of the focus setting potentially causes a pixel disalignment in captured images, which actually leads to a large depth estimation error, as noted in Ref. [15]. This paper also proposes a compensation technique for such a pixel disalignment to improve the estimation accuracy.

The rest of this paper is organized as follows. Section 2 reviews typical DFD techniques proposed in the literature to summarize their advantages and disadvantages. Section 3 elaborates the proposed cross-reblurring-based DFD technique, and a compensation technique for a pixel disalignment is discussed in Sect. 4. Section 5 gives several experimental results, and compares our technique with typical DFD techniques. Finally, Sect. 6 concludes this paper.

2. Review of Typical DFD Techniques

2.1 Image Acquisition in DFD Techniques

Figure 1 depicts an image acquisition system comprising a thin convex lens and an image detector. Light rays radiated from the point object P on the target object pass through the aperture and are refracted by the lens to converge at the point Q. On the image detector, however, the rays form a circular patch with a radius of \( r \), which is the blurred image of the point object P. Under the paraxial approximation, the radius \( r \) can be given as

\[
D = \frac{r}{f} + D_0
\]

Fig. 1 Image acquisition system.
\[ r = a d_i \left( \frac{1}{f_i} - \frac{1}{D} - \frac{1}{d_i} \right), \]  

which indicates that the depth \( D \) can be estimated by finding the corresponding blur radius \( r \) and the parameters \( a, f_i, \) and \( d_i \). The purpose of DFD techniques is to estimate the depth \( D \) to each point on the surface of a target object by using a set of images acquired by varying one of the parameters \( a, d_i, D \), and if possible \( f_i \) to change the focus setting (and hence the blur radius \( r \)). The acquired images are referred to as multi-focus images in this paper.

The point spread function (PSF) \( h(x, y) \) represents the brightness variation of the circular patch in Fig. 1 normalized by the radiated light energy. The PSF is often modeled by the pillbox model

\[ h(x, y) = \begin{cases} \frac{1}{\pi \sigma^2} (x^2 + y^2 \leq \sigma^2) & \text{otherwise} \end{cases}, \]

where \( x \) and \( y \) represent the real-valued coordinates in the spatial domain. Another common PSF model is the Gaussian one given by

\[ h(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \]

where \( \sigma \) is a blur parameter, and is related to \( r \) as

\[ \sigma = k' r \]

with \( k' \) being a lens-specific constant.

The image \( f(x, y) \) on the image detector is a blurred version of the latent all-in-focused image \( f_a(x, y) \), which is represented as

\[ f(x, y) = f_a(x, y) \ast h(x, y), \]

where \( \ast \) denotes the convolution operation in the continuous domain. The discrete image \( f(i, j) \) captured by the image detector is a band-limited and sampled version of the continuous image \( f(x, y) \), and is thus given by

\[ f(i, j) = B[f(x, y)]_{x=L, y=L}, \]

where \( L \) denotes the sampling interval, \( i \) and \( j \) are the integer coordinates, and \( B[\cdot] \) represents a band-limiting operation.

It should be noted that Eq. (6) is a space-variant convolution because the PSF \( h(x, y) \) is dependent on the depth varying on the surface of a target object. To avoid the difficulty in terms of the space-variance, many of DFD techniques proposed thus far commonly assume that the depth is approximately constant in a neighborhood of each point on multi-focus images, which allows a locally space-invariant approach and hence a block-wise depth estimation. Depending on the context in this paper, \( f(i, j) \) is assumed to represent either the whole image or a prescribed small block centered at the point for which the depth is estimated.

### 2.2 Review of Typical DFD Techniques

Reference [9] has been known as one of the earliest papers that exploited relative defocus as a cue in the depth estimation problem. This technique assumes the PSF of a lens to be approximated by the Gaussian model (4), and utilizes two multi-focus images \( f_0(x, y) \) and \( f_1(x, y) \) blurred with the blur parameters \( \sigma_0 \) and \( \sigma_1 \), respectively. Taking a ratio of the Fourier Transforms (FTs) of \( f_0(x, y) \) and \( f_1(x, y) \), defined as \( F_0(\omega_x, \omega_y) \) and \( F_1(\omega_x, \omega_y) \), respectively, leads to

\[ \frac{F_0(\omega_x, \omega_y)}{F_1(\omega_x, \omega_y)} = \frac{F_0(\omega_x, \omega_y)H_0(\omega_x, \omega_y)}{F_1(\omega_x, \omega_y)H_1(\omega_x, \omega_y)} = \frac{\sigma_0^2}{\sigma_1^2} \exp\left(-\frac{(\sigma_0^2 - \sigma_1^2)(\omega_x^2 + \omega_y^2)}{2}\right). \]

where \( \omega_x \) and \( \omega_y \) are the angular frequencies in the \( x \) and \( y \) directions, respectively, \( F_a \) is the FT of the all-in-focus image \( f_a(x, y) \), and \( H_0 \) and \( H_1 \) are the FTs of the Gaussian PSFs with the blur parameters \( \sigma_0 \) and \( \sigma_1 \), respectively. Since \( F_a \) (and hence \( f_a \)) have been eliminated in Eq. (8), the technique is a scene/texture-independent one. Taking the logarithm of the both sides leads to

\[ \log\frac{F_0}{F_1} = \log\frac{\sigma_0^2}{\sigma_1^2} - (\sigma_0^2 - \sigma_1^2) \frac{\omega_x^2 + \omega_y^2}{2}, \]

and by applying a linear regression technique with \( (\omega_x^2 + \omega_y^2)/2 \) and \( \log F_0/F_1 \) as its variates, the coefficients \( A = \sigma_0^2 - \sigma_1^2 \) and the intersection \( B = \log \sigma_1^2/\sigma_0^2 \) can be obtained numerically, from which \( \sigma_0 \) and \( \sigma_1 \) can be uniquely determined. The absolute depth \( D \) is then estimated by using Eqs. (2) and (5) together with the predetermined parameters \( a, f_i, d_i, \) and \( k' \).

It should be noted that the technique relies on the Gaussian PSF and its reproductive property. Several DFD techniques based on the same property have been proposed in the literature [10]–[15]. As pointed out in Ref. [16], however, since the Gaussian PSF model may not be an appropriate model for an actual lens, a DFD technique independent of the PSF model is desirable in a real application.

In Ref. [16], the authors investigated a DFD technique based on finding the PSF \( h_2(i, j) \) that relates two multi-focus images \( f_0(i, j) \) and \( f_1(i, j) \) as

\[ f_1(i, j) = f_0(i, j) \otimes h_2(i, j), \]

where \( \otimes \) denotes the convolution operation in the discrete domain. Since the depth is uniquely determined from \( h_2 \), the authors proposed two techniques, i.e., inverse-filtering-based and matrix-based techniques, for estimating \( h_2 \), and
concluded that the latter technique gives more accurate results. In the matrix-based technique, a table containing pairs of \( h_2 \) and its corresponding depth \( d \) is first prepared with a prescribed granularity of \( d \), and the optimal \( h_2 \) that best satisfies Eq. (10) is then searched by a brute-force matching to estimate the depth \( d \). Although this technique allows any kind of PSF models, a search of the optimal \( h_2 \) requires a large computational cost. Furthermore, since the table depends on the focus settings of \( f_0 \) and \( f_1 \), a variation of the lens parameters requires an update of the table. Because of this overhead, the technique is not a parameter-free approach.

In Ref. [17], the authors proposed a DFD technique based on a simple convolution/deconvolution transform called S-transform. The inverse S-transform corresponding to a deconvolution is given by

\[
f(i, j) = f'(i, j) - \frac{M_0}{4} \nabla^2 f'(i, j),
\]

where \( \nabla^2 \) and \( M_0 \) denote the Laplacian operator and the 2nd-order moment of the PSF \( h(i, j) \).

Let \( f_0(i, j) \) and \( f_1(i, j) \) be the multi-focus images defocused with the PSFs \( h_0(i, j) \) and \( h_1(i, j) \), respectively, and let \( M_0 \) and \( M_1 \) denote the 2nd-order moments of \( h_0 \) and \( h_1 \), respectively. An application of the S-transform to \( f_0 \) and \( f_1 \) leads to the relation

\[
f_0(i, j) - \frac{M_0}{4} \nabla^2 f_0(i, j) = f_1(i, j) - \frac{M_1}{4} \nabla^2 f_1(i, j).
\]

By using another relation between \( M_0 \) and \( M_1 \) derived from the optical geometry in Eq. (1), \( M_0 \) and \( M_1 \) can be uniquely determined. Since \( M_0 \) and \( M_1 \) are related to the blur radius \( r_0 \) and \( r_1 \) in \( h_0 \) and \( h_1 \), respectively, the depth \( D \) can be estimated by using Eq. (2) and the predetermined parameters.

Although this technique is computationally very efficient and theoretically elegant, the estimation error cannot be reduced because only a part of information on the PSF, i.e., the 2nd-order moment, is exploited. Another disadvantage of this technique is a noise amplification due to the deconvolution in the S-transform. To avoid the difficulty, the authors in Ref. [18] proposed a revised version of this approach based on “blur equalization” written as

\[
f_0(i, j) * h_1(i, j) = f_1(i, j) * h_0(i, j).
\]

An application of the forward S-transform (corresponding to a convolution) to Eq. (13) leads to

\[
f_0(i, j) + \frac{M_1}{4} \nabla^2 f_0(i, j) = f_1(i, j) + \frac{M_0}{4} \nabla^2 f_1(i, j).
\]

By using Eq. (14) instead of Eq. (12), the depth can be estimated in a similar way as in the previous approach. Since this technique is based on the convolution rather than a deconvolution, the noise amplification problem can be avoided. However, the depth estimation error is still very large with this technique, which will be demonstrated in Sect. 5.

In Refs. [19], [20], the authors proposed a sophisticated DFD technique using two multi-focus images \( f_0(i, j) \) and \( f_1(i, j) \), which are captured at \( d_1 = d_0 \) and \( d_1 \), respectively, in Fig. 1. The technique first represents \( d_i \) corresponding to the focused point \( Q \) as

\[
d_i = \frac{d_1 + d_0 + \alpha (d_1 - d_0)}{2} \quad (-1 \leq \alpha \leq 1)
\]

with \( \alpha \) being a parameter. Once \( \alpha \) is estimated, the depth \( D \) can be obtained by using Eq. (15) and the lens law \( D^{-1} + d_i^{-1} = f_i^{-1} \). To estimate \( \alpha \) from the multi-focus images, this technique focuses on the ratio

\[
\frac{M(\omega_x, \omega_y)}{P(\omega_x, \omega_y)} = \frac{H_1(\omega_x, \omega_y) - H_0(\omega_x, \omega_y)}{H_1(\omega_x, \omega_y) + H_0(\omega_x, \omega_y)};
\]

where \( M = F_1 - F_0 \) and \( P = F_1 + F_0 \). Since Eq. (16) is actually a function of the parameter \( \alpha \), the authors approximated it by a cubic polynomial in a certain range of \( (\omega_x, \omega_y) \) with using three 2-D filters \( G_{P1}(\omega_x, \omega_y) \), \( G_{P2}(\omega_x, \omega_y) \), and \( G_{M1}(\omega_x, \omega_y) \):

\[
\frac{M(\omega_x, \omega_y)}{P(\omega_x, \omega_y)} = \frac{G_{P1}(\omega_x, \omega_y)}{G_{M1}(\omega_x, \omega_y)} \alpha + \frac{G_{P2}(\omega_x, \omega_y)}{G_{M1}(\omega_x, \omega_y)} \alpha^3.
\]

Transforming Eq. (17) into the spatial domain leads to

\[
m(i, j) \ast g_{M1}(i, j) = p(i, j) \ast g_{P1}(i, j) \ast \alpha \ast p(i, j) \ast g_{P2}(i, j) \ast \alpha^3,
\]

where \( m = f_1 - f_0 \), \( p = f_1 + f_0 \), and \( \alpha \) are the impulse responses of \( G_{P1} \), \( G_{P2} \), and \( G_{M1} \), respectively. Since Eq. (18) is a cubic equation with respect to \( \alpha \), its solution gives an estimate for \( \alpha \) in Eq. (15), from which \( d_i \) and hence the depth \( D \) can be estimated.

While this technique allows any type of PSF models and gives very accurate depth estimation in a low computational cost, the filters \( G_{P1} \), \( G_{P2} \), and \( G_{M1} \) should be designed for each of the focus setting for \( f_0 \) and \( f_1 \): they should be redesigned every time the distances \( d_0 \) and/or \( d_1 \) or the lens-specific parameters are changed. Therefore, this technique is not a parameter-free approach.

On the other hand, DFD techniques have been extended in conjunction with other imaging or depth-sensing techniques in the literature. Integrations with stereo vision [21] and a coded aperture technique [22] are typical examples of such an approach. However, these extensions are out of focus in this paper: we propose a DFD technique for multi-focus images captured by a monocular camera with a conventional aperture.

3. Proposed Technique

3.1 Image Acquisition with Telecentric Optics

In Fig. 1, a variation of \( d_i \) or \( D \) alters not only the focus setting but also the scaling factor for acquired images, which deteriorates a pixel-to-pixel correspondence between multi-focus images. To avoid such a scaling, some of DFD techniques employ a telecentric optics [19], [20]. Figure 2 depicts an image acquisition system employing an object-side
telescopic lens, where a variation of the lens-object distance $D$ does not affect the size of captured images.

Let $d_0$ represent the distance between the lens and the focal plane in Fig. 2. Then, it is easily verified that the blur radius $r$ in this case is given by

$$r = K |D - d_0| \quad \left( K = \frac{a d_1}{f_1 d_0} \right). \quad (19)$$

Note in Eq. (19) that the lens-specific parameters $a$ and $f_1$ and the depth $D$ are being disentangled in comparison with Eq. (1).

### 3.2 Absolute Depth vs. Relative Depth

Motivated by the simple relation between $r$ and $D$ in Eq. (19), we propose a novel DFD technique based on the cross-reblurring property.

As briefly discussed in Sect. 1, typical applications of DFD techniques include 3-D space recognition in computer vision and 3-D shape reconstruction/modeling. The former requires an estimate for the absolute depth while the latter allows a relative one, i.e., a depth relative to a certain point along with the optical axis, because a 3-D shape can be reconstructed from its relative depth profile. Based on the discussion, this paper proposes a DFD technique that estimates a relative depth, which relaxes a requirement for lens parameters as shown later.

Figure 3 shows the focus settings in acquiring multi-focus images in the proposed technique. Three images $f_0(i, j)$, $f_1(i, j)$ and $f_2(i, j)$ are captured using the telecentric optics in Fig. 2, where the lens and the image detector are moved together by the length $\Delta$ with $d_1$ fixed. Let $z_n$ denote the position of the focal plane for the image $f_n$ ($n = 0, 1, 2$). Note in this case that the distance between the focal planes is also given by $\Delta$ because $d_0$ is constant for the fixed $d_1$. We assume that the object point P for which the depth is estimated is located between $z_0$ and $z_2$.

The problem here is to estimate a relative depth for the object point P at $z = D$ by using $f_0$, $f_1$, and $f_2$. Without loss of generality, we assume $z_0$ as the origin of the $z$ axis, and estimate the relative depth from $z = 0$. For the sake of simplicity, $z$ and $D$ are normalized by the acquisition interval $\Delta$, and are denoted again by $z$ and $D$, which normalizes $z_n = \Delta n$ into $z_n = n$ ($n = 0, 1, 2$) as shown along with the normalized $z$ axis at the bottom of Fig. 3.

Let $r_z$ represent the blur radius for the object point P on the image $f_z$ ($z = 0, 1, 2$). Since $r_z$ is evaluated on discrete images, let us normalize $r_z$ as well by the sampling interval $L$. Then, $r_z$ in [pixel] is given by modifying Eq. (19) into

$$r_z = K |D - z| \quad \left( K = \frac{a d_1}{f_1 d_0} \frac{\Delta}{L} \right) \quad (z = 0, 1, 2). \quad (20)$$

### 3.3 Proposed DFD Technique

Figure 4 illustrates the cross-reblurring approach exploited in this paper. We assume that the acquired image $f_z(i, j)$ is a sampled version of the latent all-in-focus continuous image $f_z(x, y)$ blurred with the blur radius $r_z$ ($z = 0, 1, 2$). By letting $h_z(x, y)$ denote the continuous PSF corresponding to $r_z$, the blurring and sampling process is represented as

$$f_z(i, j) = B \left[ f_z(x, y) * h_z(x, y) \right]_{x = i L, y = j L}$$

for $z = 0, 1, 2$. Note that $r_z$ and $h_z$ are unknown but being uniquely determined by the normalized depth $D$ to be estimated.

Now let $d$ be an estimate for the true depth $D$. In this case, each candidate of $d$ uniquely determines its own blur radius $r'_z$ on $f_z$ and hence the corresponding PSF $g_z(i, j)$ for $z = 0, 1, 2$. Then, reblurring $f_0$ by $g_1$ and $f_1$ by $g_0$ in a cross-coupled manner yields

$$\begin{cases} f_0(i, j) = f_0(i, j) \otimes g_1(i, j) \\ f_0(i, j) = f_1(i, j) \otimes g_0(i, j) \end{cases}$$

\quad (22)
as depicted in Fig. 4. If \( d \) is equal to the true depth \( D \) and the blurring with \( g_z(i, j) \) in the discrete domain well simulates one in the continuous domain, the blurred images \( f_{01} \) and \( f_{10} \) coincide, which was also noted in Ref. [18]. This property implies that the true depth \( D \) can be estimated by finding \( d \) in such a way that the corresponding \( g_1 \) and \( g_0 \) may render \( f_{01} \) and \( f_{10} \) to coincide each other.

Since \( g_1 \) and \( g_0 \) also depend on the coefficient \( K \) in Eq. (20), the idea here is to estimate the depth and the coefficient simultaneously by minimizing the error criterion

\[
E(d, k) = \sum_{(i, j) \in \text{BLK}} e_0^2(i, j, d, k) + \sum_{(i, j) \in \text{BLK}} e_2^2(i, j, d, k)
\]

(23)

where BLK represents a prescribed small block centered at the pixel for which the depth is estimated. \( d \) and \( k \) that minimize Eq. (23) give the best estimates for \( D \) and \( K \).

Unfortunately, however, a minimization of Eq. (23) cannot determine \( d \) and \( k \) uniquely because they are related as in Eq. (20); \( r_z \) rather than \( d \) and \( k \) can be uniquely determined in this case. To overcome the ambiguity, the error criterion (23) is extended by introducing another pair of multi-focus images \( f_1 \) and \( f_2 \) in Fig. 3:

\[
E(d, k) = \sum_{(i, j) \in \text{BLK}} e_0^2(i, j, d, k) + \sum_{(i, j) \in \text{BLK}} e_2^2(i, j, d, k)
\]

(24)

The second term makes it possible to uniquely determine the optimal \( d \) and \( k \). Eq. (24) is actually minimized by using Gauss-Newton (GN) minimization technique, whose details are discussed in Sect. 3.4.

Before leaving this section, the following brief comments are made for the proposed approach. As reviewed in Sect. 2, Xian and Subbarao proposed a similar approach in Ref. [18] referred to as “blur equalization”. They applied the property to the S-transform while the proposed technique directly formulates the property as in Eqs. (23) and (24), which clearly indicates the difference between the two techniques.

The proposed technique can be realized in the frequency domain as well by taking the DFT of the error criterion. In fact, Eq. (23) can be represented in the frequency domain as

\[
E(d, k) = \sum (F_0G_1 - F_1G_0)^2,
\]

(25)

corresponds to a circular convolution in a DFT block, which is different from an ordinary convolution for aperiodic signals/images. The circular convolution neglects blurred components entering from surrounding blocks, which causes an error in the criteria (25). To avoid such an error, the proposed technique attempts to minimize Eqs. (23) and (24) in the spatial domain at the cost of computational efficiency.

3.4 Computational Details of the Proposed Technique

This subsection discusses computational details for minimizing the error criteria by GN technique. Equation (23) is focused on for simplicity: Eq. (24) can be minimized by applying the same technique to the second term in Eq. (24).

3.4.1 Minimization of Eq. (23) based on GN technique

Let \( (d^{(n)}, k^{(n)}) \) and \( E^{(n)} \) be the estimates for the optimal \( (d, k) \) and the error in Eq. (23), respectively, in the \( n \)-th GN iteration. By expanding \( e_0 \) with respect to \( d \) and \( k \), and neglecting higher-order terms, Eq. (23) leads to

\[
E(d^{(n)} + \Delta d, k^{(n)} + \Delta k) = E(d^{(n)}, k^{(n)}) + \frac{\partial E}{\partial d} \Delta d + \frac{\partial E}{\partial k} \Delta k + \frac{1}{2} \left( \left( \frac{\partial E}{\partial d} \right)^2 + \left( \frac{\partial E}{\partial k} \right)^2 \right),
\]

(26)

where \( \Delta d \) and \( \Delta k \) are the correction terms for \( d^{(n)} \) and \( k^{(n)} \). \( \Delta d \) and \( \Delta k \) minimizing Eq. (26) can be obtained by solving

\[
\frac{\partial E}{\partial \Delta d} = \frac{\partial E}{\partial \Delta k} = 0
\]

(27)

for \( \Delta d \) and \( \Delta k \), which is written in a matrix form as

\[
\begin{bmatrix}
\sum_{(i, j) \in \text{BLK}} \frac{\partial e_0}{\partial d} \frac{\partial e_0}{\partial d} \sum_{(i, j) \in \text{BLK}} \frac{\partial e_0}{\partial d} \\
\sum_{(i, j) \in \text{BLK}} \frac{\partial e_0}{\partial d} \frac{\partial e_0}{\partial d} \sum_{(i, j) \in \text{BLK}} \frac{\partial e_0}{\partial d} \\
\end{bmatrix}
\begin{bmatrix}
\Delta d \\
\Delta k
\end{bmatrix}
= 
\begin{bmatrix}
\sum_{(i, j) \in \text{BLK}} e_0(i, j, d^{(n)}, k^{(n)}) \frac{\partial e_0}{\partial d} \\
\sum_{(i, j) \in \text{BLK}} e_0(i, j, d^{(n)}, k^{(n)}) \frac{\partial e_0}{\partial d}
\end{bmatrix}
\]

(28)

Based on the solution of Eq. (28), \( d^{(n)} \) and \( k^{(n)} \) are updated as

\[
\begin{cases}
(d^{(n+1)} = d^{(n)} + \Delta d \\
k^{(n+1)} = k^{(n)} + \Delta k
\end{cases}
\]

(29)

These steps are iterated until a stopping criteria is satisfied.

\[
\left| \frac{E^{(n+1)} - E^{(n)}}{E^{(n)}} \right| < T_E
\]

(30)

is utilized for the criteria in this paper, where \( T_E \) is a prescribed threshold value.
3.4.2 Calculation Details for Eqs. (23) and (28)

The PSFs \( g_0 \) and \( g_1 \) in Eqs. (23) and (28) can be obtained by first calculating the blur radius \( r_2^{(n)} \) for \( d^{(n)} \) and \( k^{(n)} \) by

\[
r_2^{(n)} = k^{(n)} \left| d^{(n)} - z \right| \quad (z = 0, 1)
\]

and by substituting \( r_2^{(n)} \) into the PSF model, e.g., Eq. (3) or (4). \( e_01 \) is then calculated by performing the two convolution operations in Eq. (23). On the other hand, the partial derivative \( \partial e_01/\partial d \) required in Eq. (28) can be written as

\[
\frac{\partial e_01}{\partial d} = \frac{\partial (f_0 \otimes g_1 - f_1 \otimes g_0)}{\partial d} = \left( f_0 \otimes \frac{\partial g_1}{\partial r} - f_1 \otimes \frac{\partial g_0}{\partial r} \right).
\]

Since a PSF is an even function in \( r \) and thus \( r = k|d - z| \) can be replaced with \( r = k(d - z) \), Eq. (32) leads to

\[
\frac{\partial e_01}{\partial d} = k \left( f_0 \otimes \frac{\partial g_1}{\partial r} - f_1 \otimes \frac{\partial g_0}{\partial r} \right).
\]

In a similar way, \( \partial e_01/\partial k \) is deduced as

\[
\frac{\partial e_01}{\partial k} = \left( f_0 \ast (d - 1) \frac{\partial g_1}{\partial r} - f_1 \ast d \frac{\partial g_0}{\partial r} \right).
\]

All the elements in the linear system (28) can be determined by evaluating these formulae with the current estimates \( d = d^{(n)} \) and \( k = k^{(n)} \).

3.4.3 Application to Pillbox PSF Model

The proposed technique can be directly applied to the Gaussian PSF model because \( g_0 \) and its derivatives can be readily obtained by discretizing the Gaussian PSF (4). On the other hand, it is not the case for the discretized pillbox model

\[
h(i, j) = \begin{cases} \frac{1}{\pi r^2} & (i^2 + j^2 \leq r^2) \\ 0 & \text{(otherwise)} \end{cases}
\]

because of the following limitations:

1) The original model (3) cannot be simulated by Eq. (35) for a small value of \( r \). In particular, the model results in the discrete delta function for \( r \) less than 1.0.
2) Equation (35) is not differentiable with respect to \( r \).

Rewriting Eq. (22) using \( f_0(x, y) \) leads to

\[
\begin{align*}
f_{01} &= B \left[ f_0(x, y) \ast h_0(x, y) \right]_{x=1, y=-L} \otimes g_1(i, j) \\
f_{10} &= B \left[ f_0(x, y) \ast h_1(x, y) \right]_{x=1, y=-L} \otimes g_0(i, j).
\end{align*}
\]

The limitation 1) implies that \( f_{01} \) and \( f_{10} \) do not coincide even for the true depth \( d = D \) if at least either \( h_0 \) or \( h_1 \) (and hence the corresponding \( g_0 \) or \( g_1 \)) becomes very narrow. These limitations can be attributed to the direct discretization of Eq. (3) into Eq. (35) without considering the band-limiting process preceded by the discretization. Since the band limitation in Eq. (36) removes the frequency components \( |\omega_k|, |\omega_k| \geq \pi/L \), the same components can be eliminated from the FTs of \( h_0(x, y) \) and \( h_1(x, y) \). The discussion above suggests that \( g_0(i, j) \) simulating \( h_c(x, y) \) in the discrete domain can be obtained by taking the inverse DFT of a band-limited and sampled version of the FT of the pillbox model (3).

The FT of the pillbox model (3) [23] and its derivative with respect to \( r \) are given by

\[
\begin{align*}
H(\omega_x, \omega_y) &= 2J_1(\rho \omega) \left( \omega = \sqrt{\omega_x^2 + \omega_y^2} \right) \\
\frac{\partial}{\partial r} H(\omega_x, \omega_y) &= \frac{2 \omega J_0(\rho \omega)}{\rho \omega}.
\end{align*}
\]

where \( J_1(x) \) and \( J_0(x) \) are the 1st and 0th order Bessel functions, respectively. Eq. (37) is then sampled as

\[
G(n_x, n_y) = \begin{cases} H\left(\frac{2 \pi n_x}{N}, \frac{2 \pi n_y}{N}\right) & (n_x, n_y = 0, 1, \ldots, \left\lfloor \frac{N-1}{2} \right\rfloor) \\
G(N - 1 - n_x, N - 1 - n_y) & (n_x, n_y = \left\lfloor \frac{N-1}{2} \right\rfloor + 1, \ldots, N - 1)
\end{cases}
\]

so that \( G(n_x, n_y) \) is symmetric with respect to \( n_x = n_y = \left\lfloor \frac{N-1}{2} \right\rfloor \), where \( N \) denotes the block size of the 2-D DFT. \( g_0(i, j) \) and \( g_1(i, j) \) are then obtained by taking the inverse DFT of Eq. (39). Since the band limitation in Eq. (39) smoothes out the pillbox model into a continuous surface, the limitations 1) and 2) can be avoided. The derivatives of \( g_0 \) and \( g_1 \) with respect to \( r \) can also be calculated in a similar way for Eq. (38) based on the commutativity between the DFT operation and differentiation with respect to a parameter in the DFT.

3.4.4 Initial guess for \( d^{(0)} \) and \( k^{(0)} \)

A simple technique based on the DFF approach [6], [24] is utilized for obtaining the initial values \( d^{(0)} \) and \( k^{(0)} \). The technique first evaluates

\[
M(z) = \sum_{(i,j)\in B_{\text{LK}}} (f_i(i, j) \otimes h_{BPF}(i, j))^2
\]

for \( z = 0, 1, 2 \) as a focus measure for the multi-focus image \( f_i \), where \( h_{BPF} \) is a circularly-symmetric 2-D band-pass filter (BPF). Under the conditions that the PSF yielding \( f_i \) is close to the Gaussian model and that the passband of the BPF is very narrow, \( M(z) \) can be approximated by

\[
M(z) \approx \exp \left( -k^2 (d - z)^2 \Omega_c^2 \right) \quad (z = 0, 1, 2),
\]

where \( \Omega_c \) denotes the radius of the passband of the BPF [24]. \( d \) and \( k \) in Eq. (41) can be determined by using a linear regression technique for \( \log M(z) \) \( (z = 0, 1, 2) \). \( d \) is then utilized for \( d^{(0)} \) while \( k^{(0)} \) is estimated as \( k^{(0)} = c \cdot k \), where the constant \( c \) is determined in such a way that the main lobes of the Gaussian and the actual PSF models are as close as
possible. \( c \approx 1.73 \) is obtained for the pillbox model.

4. Compensation for Pixel Disalignment

Figure 5 illustrates an actual acquisition system for multi-focus images, where the telecentric lens and the camera are fixed on the computer-controlled linear stage, and are moved forward and backward against the target object to vary the focus setting. If the optical axis of the lens and the motion axis of the stage are completely aligned, ideal multi-focus images can be captured. However, such an ideal alignment cannot be expected in an actual application: a slight difference between the axes causes a pixel disalignment between multi-focus images, which results in a large estimation error, as noted in Ref. [15].

In Ref. [15], the author assumed that one of a pair of multi-focus images is a scaled, rotated, and translated version of the other, and estimated the depth together with the multi-focus images is a scaled, rotated, and translated version can be neglected in Fig. 5, only a possible translation of the other, and estimated the depth together with the multi-focus images. To estimate \( S_x \) and \( S_y \) together with the depth, the error criterion (24) is modified into

\[
E(d, k, s_x, s_y) = \sum_{(i,j) \in \text{BLK}} e^2(i, j, d, k, s_x, s_y) + \sum_{(i,j) \in \text{BLK}} e^2_{\text{fg}}(i, j, d, k, s_x, s_y)
\]

(42)

where

\[
e_{01}(i, j, d, k, s_x, s_y) = f_0(i + \frac{s_x}{2}, j + \frac{s_y}{2}) \otimes g_1(i, j)
\]

\[
- f_1(i - \frac{s_x}{2}, j - \frac{s_y}{2}) \otimes g_0(i, j)
\]

\[
e_{12}(i, j, d, k, s_x, s_y) = f_1(i + \frac{s_x}{2}, j + \frac{s_y}{2}) \otimes g_2(i, j)
\]

\[
- f_2(i - \frac{s_x}{2}, j - \frac{s_y}{2}) \otimes g_1(i, j)
\]

and is minimized with respect to \( d, k, s_x, s_y \) by GN approach.

The bilinear interpolation is utilized for interpolating pixels corresponding to the real-valued displacements \( \pm s_x/2 \) and \( \pm s_y/2 \) in Eq. (42). The partial derivatives \( \partial f_2 / \partial s_x \) and \( \partial f_2 / \partial s_y \) required in the GN technique can be readily obtained from the bilinear interpolation formula, e.g.,

\[
\frac{\partial f_0}{\partial s_x} = \frac{1}{2} \left( f_0(i + 1, j) - f_0(i, j) \right) \left( 1 - \frac{s_y}{2} \right)
\]

\[
+ \frac{1}{2} \left( f_0(i + 1, j + 1) - f_0(i, j + 1) \right) \frac{s_y}{2}.
\]

(43)

By using Eq. (43), the linear equation with respect to \( \Delta d, \Delta k, \Delta s_x, \) and \( \Delta s_y \) can be derived in a similar way as in Eq. (28). For the initial values \( s_x^{(0)} \) and \( s_y^{(0)} \), a small positive value, e.g., 1.0 \( \times 10^{-2} \), is utilized.

5. Experimental Results

This section gives several experimental results by using synthetic and real test images to demonstrate the advantage of our technique. Note that the term “proposed technique” in this section indicates one with the pixel disalignment compensation, except otherwise indicated. The size of BLK is fixed to 15 \( \times 15 \), and \( T_E = 0.1 \) and \( N = 11 \) are utilized in the proposed technique.

5.1 Comparison of the Pillbox and Gaussian PSF Models

Figure 6 shows real multi-focus images captured by the system in Fig. 5 with a slanted planar metal plate as their target object. A \( \times 0.25 \) telecentric lens (Edmund optics SilverTL) and a monochrome CCD camera with 1280 \( \times 960 \) pixels in 8-bit depth (Imaging Source DMK 23U445) were utilized in Fig. 5. Since an actual lens suffers from a non-ideality known as field curvature, 3-D shapes reconstructed by DFD techniques are inevitably distorted by the non-ideality. To avoid the distortion, a partial image was cropped from each of the captured multi-focus images so that the field curvature can be neglected. Figures 6(a)–(c) are the images of 240 \( \times 240 \) pixels cropped from the original full-sized images.

The proposed technique with the either pillbox or Gaussian PSF model was applied to Fig. 6 to reconstruct the shape of the target object. Figure 7 depicts the results, where (a) and (b) are with the pillbox and Gaussian models, respectively. A comparison based on a numerical error criteria is impossible in this example because the ground-truth of the shape is unknown. As the target object is planar in this experiment, the root mean squared error (RMSE) is calculated between the reconstructed plane and its approximate plane.
Fig. 7 Shapes reconstructed for Fig. 6 with the pillbox and Gaussian PSF models.

determined by a least-square fitting, which are also shown in Fig. 7.

In Fig. 7, the pillbox model (a) gives the better result than the Gaussian one (b). In particular, the shape with the Gaussian model is slightly curved rather than a straight line in the side view (d), which is caused by a mismatch of the PSFs between the actual lens and the reblurring simulation. These results indicate that the Gaussian model is not always an appropriate PSF model for an actual lens, as pointed out in Ref. [16]. The proposed technique is thus advantageous because it allows any kind of PSF models in contrast to Refs. [9]–[15].

5.2 Comparison of the Proposed Techniques with and without the Disalignment Compensation

The proposed techniques with and without the pixel disalignment compensation were compared by using synthetic test images in order to numerically evaluate the effectiveness of our approach.

Figure 8 shows the object models (a)–(d) and the surface textures (e)–(g), from which the synthetic test images were generated. Specifically, an all-in-focus image (320 × 240 pixels) was first generated for a combination of each model and each texture in Fig. 8, which leads to 12 images in total. Then, each of the all-in-focus images was defocused to generate the multi-focus images $f_0, f_1, \text{ and } f_2$ by simulating the acquisition system in Fig. 3. The pillbox PSF model was utilized in this simulation with the parameter $K$ in Eq. (20) varied from 1.0 to 3.0 in a step of 0.1 for each of the all-in-focus images. For simulating the pixel disalignment, a four-times upsampled image was first generated and then downsampled into the original size on each of the sampling grids corresponding to $(s_x, s_y) = (0, 0), (0.25, 0.25), \text{ and } (0.5, 0.5)$. Gaussian noise with zero mean and standard deviation $\sigma = 0.5$ was added to each of the test images as an observation noise followed by a quantization into 8-bit depth.

The proposed techniques with and without the disalignment compensation were then applied to each set of the test images with $(s_x, s_y) = (0, 0), (0.25, 0.25), \text{ and } (0.5, 0.5)$, and the RMSE between the model and the reconstructed shape was evaluated. Table 1 summarizes the RMSEs averaged over the models, textures, and $K$s for each disalignment $(s_x, s_y)$. The table indicates that a small disalignment less than a pixel causes a large estimation error, and that the compensation technique effectively reduces the error in particular for a large disalignment.

5.3 Comparisons with Typical DFD Techniques

The proposed technique was compared with typical DFD techniques in Refs. [18], [19]. These techniques were first modified so that they can be applied to the synthetic and real test images in the previous subsections.

In the S-transform-based technique [18], the depth can be estimated from $M_z(z = 0, 1)$ in Eq. (14) and the lens parameters. In our acquisition system and computer-simulation with a telecentric optics, although the only parameter necessary in this approach is $K$ in Eq. (20), a direct substitution of $K$ for each of the synthetic test images leads to very poor estimation results because of the limitation 1) in Sect. 3.4.3.

To improve the technique, a similar approach as in Eq. (24) was applied, where the error terms $e_{0,1}$ and $e_{1,2}$ in Eq. (24) were replaced with the S-transform-based criterion
Table 2 Summary of the DFD techniques being compared.

|                | PSF model | Required images | Lens parameters | Estimation accuracy | Computation time |
|----------------|-----------|-----------------|-----------------|--------------------|-----------------|
| Proposed       | Arbitrary | 3               | Unnecessary     | High               | 1               |
| Ref. [18]      | Degenerated into 2nd-order moment | 2               | Necessary       | Low                | 0.0047          |
| Ref. [19]      | Arbitrary | 2               | Necessary       | High               | 0.0052          |

![Fig. 9](image_url) Comparison of RMSEs by the three techniques.

with $M_z = r_z^2 = K^2(D - z)^2$ for the pillbox model. The estimates for $D$ and $K$ can be obtained by applying GN technique in a similar way as in Sect. 3.4.1. Note that this modification makes it possible to compare the S-transform-based and our techniques directly because they require the same number of multi-focus images, i.e., $f_0$, $f_1$ and $f_2$.

On the other hand, the technique in [19] requires a set of the filters $G_{P1}$, $G_{P2}$, and $G_M$ and a pre-filter for each value of $K$ of the test images. We thus designed these filters in $7 \times 7$ taps with $K$ varied from $K = 1.0$ to $3.0$ in a step of 0.1. For the synthetic test images, the filters corresponding to $K$ of each image set were utilized while for the real images, the best filters giving the minimum RMSE were selected in the bank of the filters. In order to make a fair comparison, this technique, requiring only two multi-focus images, was applied to either of the pairs $f_0$ and $f_1$ or $f_1$ and $f_2$ selectively depending on the focus measure $M(z)$ in Eq. (40): the pair $f_0$ and $f_1$ is selected if $M(0) > M(2)$ indicating that $f_0$ is more focused than $f_2$, otherwise the pair $f_1$ and $f_2$ is utilized.

These three techniques were applied to each of the synthetic test images with no pixel disalignment. In Fig. 9, the RMSEs averaged over the models and textures are plotted for each technique. Figure 10 demonstrates the shapes reconstructed by the three techniques for the model 2 with $K = 2.0$. On the other hand, Fig. 11 shows the shapes reconstructed for the real test images in Fig. 6 by the two techniques being compared.

We can confirm from Figs. 9 and 10 and Figs. 7 (a) and 11 that the proposed technique gives the best results from the viewpoint of the subjective and objective accuracy.

Finally, Table 2 summarizes the advantages and disadvantages of the three techniques together with their normalized execution times. The proposed technique requires an intensive computational cost because of the convolution operations in each GN iteration. However, it offers several advantages including the parameter-freeness, PSF-independency, and higher estimation accuracy at the cost of computational efficiency.

6. Conclusions

This paper proposed a novel DFD technique based on the cross-reblurring property that two images having different focus settings coincide if they are reblurred with the opposite focus setting. By minimizing the mean squared error between the cross-reblurred images, block-wise relative depths can be estimated without requiring lens parameters in the case an object-side telecentric optics is employed. Unlike many of existing DFD techniques, our technique allows any kind of PSF models. A compensation technique for a pixel disalignment was also investigated to reduce its impact on the estimation accuracy.

The experimental results given in this paper demonstrates that the proposed technique gives higher estimation accuracy over typical DFD techniques at the cost of computational efficiency.

An actual lens inevitably suffers from several non-idealities including field curvature and astigmatism, which
severely degrades the depth estimation accuracy in DFD techniques. Compensation techniques for these nonidealities are thus necessary for further improving the estimation accuracy, which is left for future work.

References

[1] H.H. Buelthoff and A.L. Yuille, “Shape-from-X: Phychophysics and computation,” Proc. SPIE, vol.1383, no.1, pp.235–246, April 1991.
[2] R. Szeliski and S.B. Kang, “Recovering 3D shape and motion from image streams using nonlinear least squares,” Journal of Visual Communication and Image Representation, vol.5, no.1, pp.10–28, March 1994.
[3] Y. Cui, S. Schuon, D. Chan, S. Thrun, and C. Theobalt, “3D shape scanning with a time-of-flight camera,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.1173–1180, San Francisco, United States, June 2010.
[4] M. Gupta, A. Agrawal, A. Veeraraghavan, and S.G. Narasimhan, “Structured light 3D scanning in the presence of global illumination,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.713–720, Colorado Springs, United States, June 2011.
[5] Y. Xiong and S.A. Shafer, “Depth from focusing and defocusing,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.68–73, New York, USA, June 1993.
[6] S.K. Nayer, “Shape from focus system,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.302–308, Champaign, United States, June 1992.
[7] M. Subbarao and J. Tyan, “Optimal focus measure for passive autofocus and depth-from-focus,” Proc. SPIE 2598, Videometrics IV, Philadelphia, United States, Sept. 1995.
[8] M. Subbarao and C. Tiao, “Accurate recovery of three-dimensional shape from image focus,” IEEE Trans. Pattern Anal. Mach. Intell., vol.17, no.3, pp.266–274, March 1995.
[9] A.P. Pentland, “A new sense for depth of field,” IEEE Trans. Pattern Anal. Mach. Intell., vol.PAMI-9, no.4, pp.523–531, July 1987.
[10] A.N. Rajagopalan and S. Chaudhuri, “Space-variant approach to recovery of depth from defocused images,” Computer Vision and Image Understanding, vol.68, no.3, pp.309–329, Dec. 1997.
[11] S. Chaudhuri and A.N. Rajagopalan, Depth from defocus: A real aperture imaging approach, Springer-Verlag, New York, 1999.
[12] M. Asif, A.S. Malik, and T.-S. Choi, “3D shape recovery from image defocus using wavelet analysis,” Proc. IEEE International Conference on Image Processing 2005, Genova, Italy, Nov. 2005.
[13] K.S. Pradeep and A.N. Rajagopalan, “Improving shape from focus using defocus cue,” IEEE Trans. Image Process., vol.16, no.7, pp.1920–1925, July 2007.
[14] V.P. Namboodiri, S. Chaudhuri, and S. Hadap, “Regularized depth from defocus,” Proc. IEEE International Conference on Image Processing 2008, San Diego, United States, Oct. 2008.
[15] R. Ben-Ari, “A unified approach for registration and depth in depth from defocus,” IEEE Trans. Pattern Anal. Mach. Intell., vol.36, no.6, pp.1041–1055, June 2014.
[16] J. Ens and P. Lawrence, “An investigation of methods for determining depth from focus,” IEEE Trans. Pattern Anal. Mach. Intell., vol.15, no.2, pp.97–108, Feb. 1993.
[17] M. Subbarao and G. Surya, “Depth from defocus: A spatial domain approach,” International Journal of Computer Vision, vol.13, no.3, pp.271–294, 1994.
[18] T. Xian and M. Subbarao, “Depth-from-defocus: Blur equalization technique,” Proc. SPIE 6382, Two- and Three-Dimensional Methods for Inspection and Metrology IV, Boston, United States, Oct. 2006.
[19] M. Watanabe and S.K. Nayer, “Minimal operator set for passive depth from defocus,” Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.431–438, San Francisco, United States, June 1996.
[20] M. Watanabe and S.K. Nayer, “Rational filters for passive depth from defocus,” International Journal of Computer Vision, vol.27, no.3, pp.203–225, 1998.
[21] A.N. Rajagopalan, S. Chaudhuri, and U. Madenagudi, “Depth estimation and image restoration using defocused stereo pairs,” IEEE Trans. Pattern Anal. Mach. Intell., vol.26, no.11, pp.1521–1525, Nov. 2004.
[22] C. Zhou, S. Lin, and S.K. Nayer, “Coded aperture pairs for depth from defocus and defocus deblurring,” International Journal of Computer Vision, vol.93, no.1, pp.53–72, 2011.
[23] R.N. Bracewell, The Fourier transform and its applications, International edition, Electrical and Electronic Engineering Series, McGraw-Hill, Singapore, 1986.
[24] T. Yoshida, “3-D shape measurement with shape from focus/ defocus technique and its applications,” IEICE Technical Report, vol.115, no.348, pp.61–66, Dec. 2015 (in Japanese).

Kazumi Takemura received the B.E. from University of Fukui, Japan, in 2016. She is currently a master course student of Graduate school of Engineering, University of Fukui, where he is doing a research on 3D imaging.

Toshiyuki Yoshida received the B.E., M.E., and D.E. in electronics from Tokyo Institute of Technology, Japan, in 1986, 1988, and 1991, respectively. He was an associate professor in Tokyo Institute of Technology in 1994. He has moved to University of Fukui in 2003, where he is currently a professor in Dept. of Information Science. His research interests include image/video and signal processing and 3-D imaging. He is a co-author of several books, e.g., “Image Coding” (in Japanese, Corona Publishing) and “Image Information Coding” (in Japanese, Ohmsha).