Solving the Depth Ambiguity in Single-Perspective Images

MAJED EL HELOU[*], MARJAN SHAHPASKI, AND SABINE SÜSSTRUNK

School of Computer and Communication Sciences, EPFL, Switzerland
[*]*majed.elhelou@epfl.ch*

Abstract: Scene depth estimation is gaining in importance as more and more AR/VR and robot vision applications are developed. Conventional depth-from-defocus techniques can passively provide depth maps from a single image. This is especially advantageous for moving scenes. However, they suffer a depth ambiguity problem where two distinct depth planes can have the same amount of defocus blur in the captured image.

We solve the ambiguity problem and, as a consequence, introduce a passive technique that provides a one-to-one mapping between depth and defocus blur. Our method relies on the fact that the relationship between defocus blur and depth is also wavelength dependent. The depth ambiguity is thus solved by leveraging (multi-) spectral information. Specifically, we analyze the difference in defocus blur of two channels to obtain different scene depth regions. This paper provides the derivation of our solution, a robustness analysis, and validation on consumer lenses.

1. Introduction

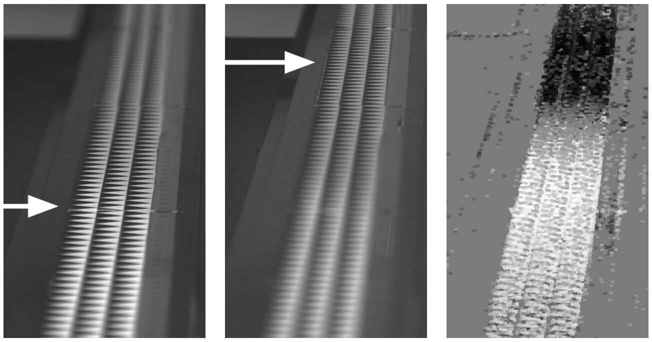
Optical depth estimation, in which depth is estimated using radiation properties, is receiving increased attention due to the proliferation of augmented/virtual reality applications [1–3] and robot vision systems [4, 5]. These applications require depth maps to be generated continuously in real time.

Based on their operating principle, we can classify existing depth-estimation techniques into several categories: stereo vision, photometric stereo, structure from motion, time of flight, structured light, and depth from defocus. Stereo techniques can provide a depth map of the scene, but they require at least two photographs with known relative capture positions, and must solve the correspondence problem [6, 7]. Time-of-flight and structured-light approaches can eliminate the need for multiple captures, but require synthetic illumination that can be problematic in bright or outdoor locations [8]. For these reasons, passive methods requiring only a single capture can prove to be very advantageous.

One such passive technique, depth from defocus, is of particular interest as it allows the recovery of dense depth maps of dynamic scenes while being computationally efficient. The method is based on passive illumination and does not require any additional hardware. Defocus in an image is blur that is dependent on the depth of an object in relation to the focal plane (i.e., the depth plane in the scene where the image is in focus). Hence, a correct estimation of the defocus blur can lead to an exact depth map for the entire scene from a single image.

However, current depth-from-defocus methods have an ambiguity problem in mapping blur magnitude to depth [9–14]. Placing an object away from the focal plane in either direction increases its defocus blur in a similar way if the camera aperture is symmetric. Asymmetrically coded apertures can be used to distinguish the two cases in the depth map computation. However, they are not practical for consumer cameras because the asymmetric aperture deteriorates image sharpness and details [15].

This paper extends our approach in [16] for solving the depth ambiguity in depth-from-defocus methods. Our closed-form solution, which does not require calibration or aperture modifications, makes it possible to obtain an unambiguous, injective mapping between depth and defocus blur.



(a) Green channel (b) NIR channel (c) Blur difference

Fig. 1. The focal plane of the green channel (a) is at a shallower depth than that of the NIR (b). By analyzing the defocus blur difference (c), the depth range is split into two regions that are used to solve the depth ambiguity. A similar shift in focus is present between red and green channels; we show the NIR channel for a better visualization.

We exploit the fact that the blur function of depth is not identical across different wavelength channels. In fact, the position of the focal plane itself depends on the wavelength, as illustrated in Fig. 1 between green and near-infrared (NIR). We leverage the wavelength dependency of defocus blur in parallel with its depth dependency to solve the ambiguity problem by analyzing blur difference. Hence, a one-to-one mapping between defocus blur and depth can be derived using two spectral channels. In this paper, we show results based on both green-red and green-NIR channels.

We analyze the relative blur between two different spectral channels to split the depth range into two subranges that cover the full range. A one-to-one mapping between depth and defocus blur is possible on each of the two subranges, thus providing a full-range one-to-one mapping. Our solution is derived based on a simple lens model, which we also use to derive error bounds. We validate our method with typical consumer cameras using both green and red, as well as the more robust green and near-infrared channels. Our main contributions are:

- ⋆ A simple closed-form solution to the depth ambiguity problem based on the simple lens model, and an error-bound analysis of our solution metric.
- ⋆ The simple-lens derivation is validated experimentally on complex lenses.
- ⋆ A depth ambiguity solution that can be integrated to correct standard depth-from-defocus methods for standard lenses. It is, to the best of our knowledge, the first solution that can be used to correct previous depth-from-defocus methods, as it requires no hardware modifications (in contrast with coded apertures or chromatic-aberration-exaggerated lenses) and no calibration or knowledge of camera settings.

2. Related Work

Depth from defocus is a widely studied problem in the imaging literature. It stands as a good alternative to stereo for depth estimation [17]. Unlike stereo, it does not have to cope with correspondence problems, and only requires a single capture. This has yielded many techniques for estimating image blur [9–12, 18–23] and mapping it to depth. Lin *et al.* propose in [12] to compute multiple aperture-shape filters that they use to distinguish defocus levels in a given image. Other techniques for depth from defocus typically estimate defocus-blur magnitude by analyzing edges, or by re-blurring the images and analyzing the variation [9–11]. All such approaches suffer, however, from their inability to solve the ambiguity problem [9–14]. This depth ambiguity is illustrated in Fig. 2 where locations A and C have equal circles of confusion, or defocus blur (radius r) on the sensor plane despite having different depth. The aforementioned blur estimation techniques cannot distinguish whether the object is at depth A or C. Hence, they are limited to enforce the assumption that the focal plane is at a shallower depth than all of the scene objects (or vice versa). In the general case where the image is focused around the mid-depth range in the scene, objects in front or behind the focal plane may have the same blur which then cannot be resolved into two different depth values. This is due to the symmetry of camera apertures creating a circular rotation-invariant blur as explained in more formal detail in the following sections.

Coded apertures can be used to optimize the differentiation of depth-dependent blur relative to depth [15], making the effect of depth variation more detectable through the defocus blur. Early adoption of coded apertures in photography can be found in the works of [24–26]. These approaches are greatly extended in [27]. The proposed method allows the recovery of scene depth information together with an all-in-focus image. The authors design an optimal aperture which produces maximally different blur circles for different scene depths. However, since their designed aperture is symmetric, the method is limited to depth either smaller or greater than the focal distance.

Asymmetric coded apertures can be designed so as to have different blur shapes on the two different sides of the focal plane. A recent technique attempts to solve the ambiguity problem this way by relying on asymmetric coded camera apertures [15], which deteriorate image quality. However, as their method requires camera modifications and affects image sharpness, it is not used in state-of-the-art techniques for depth from defocus [28], which still suffer from the ambiguity problem. Sellent *et al.* also state that the ambiguity cannot be resolved when symmetric camera apertures are used [15]. However, as we show in the rest of the paper, a solution does exist using symmetric-aperture cameras.

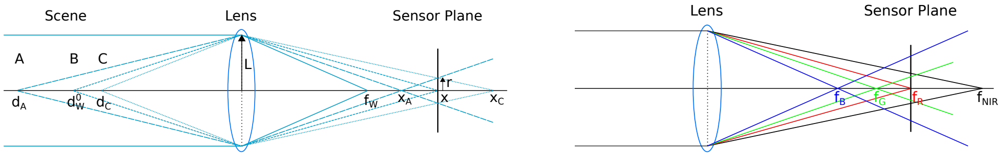
Chromatic aberration, a known physical phenomenon used often for depth-of-field extension [29], has been used by Trouvé *et al.* to solve for depth [30]. They design a specific chromatic-aberration-exaggerated lens which modifies the focal depths of RGB channels. They calibrate their lens by learning the full mapping from every spectral blur triplet to the corresponding depth. Their solution cannot be deployed on typical lenses or on top of previous depth-from-defocus methods. Furthermore, their modified lens results in degraded images due to color spilling caused by axial chromatic aberration. Our solution, on the other hand, requires no hardware modifications, calibration, or learning and can thus be applied on already captured photos with general depth-from-defocus algorithms. As hardware is not modified, our approach does not deteriorate images.

3. Depth Ambiguity

In this section, we first introduce the mathematical framework for our depth disambiguation metric, which we denote by Δ. We then show how it can be applied using synthetic images. Lastly, an analysis of error bounds is provided describing the robustness of our disambiguation to spectral blur estimation errors.

3.1. Mathematical Framework and Solution

We develop our proposed solution in the framework of a simple lens model and extend to more complex camera lens compounds in the next section. The amount of blur on the camera sensor is characterized by the radius of a disk (called circle of confusion), which represents the image of a point light source. This blur is wavelength dependent. The blur radius r_W for a channel W is



(a)

(b)

Fig. 2. Simple lens model physics. (a) Objects A and C have the same circle of confusion (blur radius r) despite being at different depths. (b) Axial chromatic aberration: light is dispersed by the lens due to different refractive indices for different wavelengths.

derived from the simple lens equation:

$$r_W(d) = L\left|1 - \frac{x}{f_W} + \frac{x}{d}\right|, \tag{1}$$

where L is the aperture radius of the simple lens, x is the distance between the center of the lens and the sensor plane, f_W is the focal length of the lens for channel W and d is the depth of the source point relative to the lens (Fig. 2 (a)).

For a given channel W and sensor position x, the focal plane is at scene depth d_W^0, at which the blur radius is minimal: $r_W(d_W^0) \to 0$. d_W^0 is given by:

$$d_W^0 = \frac{x}{\frac{x}{f_W} - 1}. \tag{2}$$

Eq. (2) is valid unless x is smaller than f_W, in which case the image is out of focus for all depth values, since all rays converge behind the sensor plane, and blur becomes a strictly increasing function of depth. By varying the scene depth d in Eq. (1), and by assuming a more realistic scenario where $x > f_W$, we can infer that the blur radius is a convex function of d. Therefore, if and only if the focal plane is set to the closest or furthest source point in the captured scene will the blur radius be, respectively, a strictly increasing or decreasing function of d. However, these two scenarios are special cases, and in practice the blur radius is a convex function of depth with a minimum at the focal plane. This non-injective depth to blur mapping limits current depth-from-defocus algorithms. They must assume that the focal plane is at either one of the two extreme scene depths.

To overcome this limitation, we leverage the differences in blur in different spectral channels. We consider two channels Y and Z of different wavelength and thus different focal length, setting $f_Z > f_Y$ without loss of generality. Noticing that the blur radius is wavelength-dependent, we study the difference $\Delta_{Z,Y}(d) \triangleq r_Z(d) - r_Y(d)$ given by:

$$\Delta_{Z,Y}(d) = \begin{cases} \alpha \triangleq L\left(\frac{x}{f_Y} - \frac{x}{f_Z}\right) & d \leq d_Y^0 \\ 2L\left(1 + \frac{x}{d}\right) - L\left(\frac{x}{f_Y} + \frac{x}{f_Z}\right) & d \in [d_Y^0, d_Z^0] \\ -\alpha = L\left(\frac{x}{f_Z} - \frac{x}{f_Y}\right) & d \geq d_Z^0. \end{cases} \tag{3}$$

For values of scene depth d smaller than d_Y^0 or larger than d_Z^0, $\Delta_{Z,Y}(d)$ is constant with respect to d and only depends on the camera parameters (L and x) and the focal lengths of the two spectral channels Y and Z. Identifying whether a source point is on one side or the other of the focal plane for a given channel is thus equivalent to matching its Δ value to $\pm\alpha$. For instance, if $\Delta_{Z,Y}(d) = -\alpha$ then $d \geq d_Z^0$. Otherwise, the source point is on the side of the focal plane that is closer to the camera ($d \leq d_Z^0$).

In practice, f_Y and f_Z are close between RGB channels, making the depths d_Y^0 and d_Z^0 close, hence decreasing α towards values close to 0. Note also that more complex lenses are designed to correct color chromatic aberration and minimize the shift between color focal planes [31]. We show in Section 4.1 that the shift is nevertheless detectable even with complex lenses, and that when using NIR the corresponding α value becomes noticeably larger, allowing for a more robust solution (Fig. 5). This is true for two reasons: first, lenses generally do not correct chromatic aberration in NIR, and second, NIR wavelengths are significantly larger than RGB channels' wavelengths, thus causing more aberration.

To compute the Δ metric value at every scene location, it is necessary to estimate the defocus blur radii r_Y and r_Z. However, the estimation of blur does not immediately yield the defocus blur radius because there are additional factors (spherical aberration, field curvature, motion blur, etc.) that add to the blur in an image. For a given image patch $I(d)$ at depth d, captured at a fixed wavelength λ_W, what we observe is a blurred version $I_b(d, \lambda_W)$:

$$I_b(d, \lambda_W) = I(d) * PSF_{eq}(d, \lambda_W, x, y) = I(d) * H_{def}(d, \lambda_W, x, y) * H_0(x, y), \tag{4}$$

where d is the depth of the object captured in pixel coordinates (x, y). We separate the equivalent point spread function PSF_{eq} generally modeled in the literature as a Gaussian kernel [9–11,32,33], into two Gaussian kernels [19]. The first kernel, H_{def}, accounts for the blur due to defocus at depth d in channel W, and H_0 accounts for all other blur factors. The standard deviation of the Gaussian kernel H_{def} is linearly related to the blur radius

$$\sigma_{def}(d, \lambda_W) = k * r_W(d), \tag{5}$$

where k is a constant dependent on the optical system and that can be calculated by calibration [19]. H_0 is assumed to be a function of pixel location and is wavelength-invariant, with the defocus blur being accounted for in H_{def}. H_0 accounts for motion blur, along with all blur effects that are pixel-position dependent, rather than wavelength or depth dependent. The variance σ_{eq}^2 of PSF_{eq} is the sum of the variances of H_{def} and H_0:

$$\sigma_{eq}^2(d, \lambda_W, x, y) = \sigma_{def}^2(d, \lambda_W, x, y) + \sigma_0^2(x, y). \tag{6}$$

Therefore, when the squared blur radius $r_W^2(d)$ is estimated in post-processing, it is $\sigma_{eq}^2(d, \lambda_W, x, y)$ that would be approximated, instead of the desired $\sigma_{def}^2(d, \lambda_W, x, y)$. This generally modifies the estimated value by an offset, $\sigma_0^2(x, y)$ being constant across the image in most cases (static scenes, centered objects or corrected spherical aberration). Hence it does not modify relative depth estimation and is neglected in the literature [10]. We note nonetheless that the pixel-wise subtraction $\sigma_{eq}^2(d, \lambda_Z, x, y) - \sigma_{eq}^2(d, \lambda_Y, x, y)$ could better estimate $\sigma_{def}^2(d, \lambda_Z, x, y) - \sigma_{def}^2(d, \lambda_Y, x, y)$. This is true because $\sigma_0^2(x, y)$ is invariant with respect to wavelength and, the subtraction being pixel-wise, the dependency with respect to (x, y) could be canceled out. The depth ambiguity solution is given by an analysis of the value of our metric Δ, estimated per pixel, which we elaborate on in Section 3.2.

3.2. Synthetic Images Example

We illustrate the framework introduced in 3.1 through a comprehensive simulation. The analysis in this section illustrates the building blocks for the disambiguation solution, and motivates the robustness and error bounds study.

We generate binary synthetic images by adding one or two simple patterns per patch that are randomly selected from a set of 4 patterns (horizontal, vertical or diagonal lines). Fig. 3 (a) shows an example image where the patches are of size 50 × 50 pixels. We then blur the original sharp image according to Eq. (4). To simulate the varying depth, we scan the image column-wise

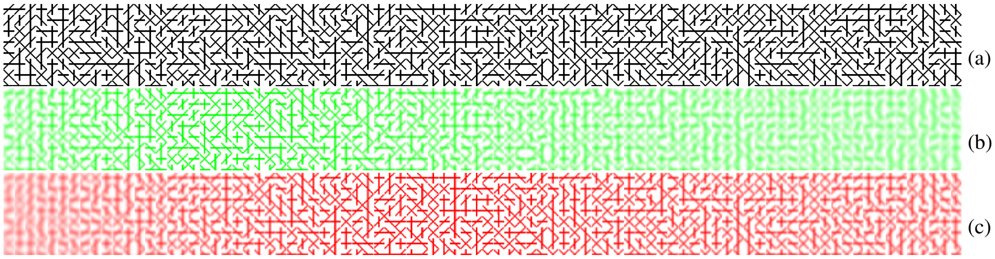


Fig. 3. (a) Sharp image. (b) and (c) Same image blurred as it would be if captured in channels Y (green) and Z (red) respectively, with defocus blur (depth increasing linearly from left to right) and uniform depth-independent blur. Best seen on screen.

(from left to right) and increase the depth linearly with the number of columns. The defocus blur variance σ^2_{def}, corresponding to the squared blur radius, is computed using Eq. (1) for the two channels Y and Z introduced earlier and an empirical constant $\sigma^2_0 = 0.2$ is finally added. The blur is simulated as a 2D convolution with a Gaussian filter.

To estimate the blur at a certain location, we first re-blur the image with a unit-variance Gaussian kernel. The gradient magnitude is then estimated for both images (original and re-blurred) at every pixel as the $L2$-norm of $\mathbb{R}^4$ vectors M and M_{reblur}. The four entries of these two vectors are the results of convolving the two images with each of the following kernels:

$$k_1 = \begin{bmatrix} -1/2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}; k_2 = \begin{bmatrix} 0 & 0 & -1/2 \\ 0 & 0 & 0 \\ 1/2 & 0 & 0 \end{bmatrix}; k_3 = \begin{bmatrix} 1/2 & 0 & -1/2 \end{bmatrix}; k_4 = k_3^T. \quad (7)$$

Blur magnitude b is lastly computed according to Eq. (8) at edge locations (x_{edge}, y_{edge}), which are obtained using a Canny edge detector.

$$b(x_{edge}, y_{edge}) = 1 - \frac{\| M(x_{edge}, y_{edge}) \|_2}{\| M_{reblur}(x_{edge}, y_{edge}) \|_2} \quad (8)$$

Fig. 4 shows the results of blur estimation on the synthetic images. The results are plotted as a function of depth by moving horizontally across the images in Fig. 3 (b) and (c), where the blur is constant vertically. We calculate the blur as the median value, over a patch column, of all blur values estimated on edges in that column. The estimate for $\Delta_{Z,Y}(d)$ deviates from its theoretical expression in Eq. (3), at the extreme values of maximum blur where the estimation of blur becomes less accurate. However, to infer depth from defocus blur, it suffices to observe that when $\Delta_{Z,Y}(d) > 0$, $r_Z(d)$ is injective, and when $\Delta_{Z,Y}(d) < 0$, $r_Y(d)$ is injective. Thus, there is no ambiguity in recovering d when the sign of $\Delta_{Z,Y}$ is known. When using both channels to create the inverse mapping, there is an acceptable margin of error where we are allowed to mistake the sign of Δ and still have a correct injective mapping due to redundancy. Indeed, if we define d_n as the distance where $\Delta_{Z,Y}(d_n) = 0$, $r_Z(d)$ is not only injective over $d < d_n$ (green shaded area in Fig. 4), but over the larger range where $d < d^0_Z$ (thick red line in Fig. 4), and similarly for $r_Y(d)$. The robustness topic is explored in more detail in Section 3.3.

3.3. Error Bounds Analysis for the Metric Δ

Two sources of imprecision can be discerned in the practical computation of our disambiguation metric Δ. First, they can be due to the fact that camera acquisition channels are not narrow band, and so the real physical wavelength captured can vary within the channel's range. Second,

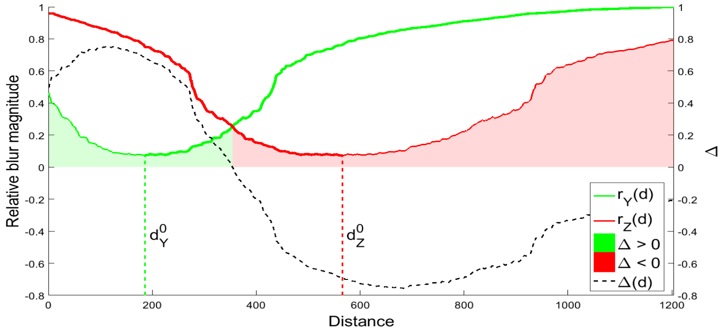


Fig. 4. Blur estimation results across synthetic images of channels Y (Fig. 3 (b)) and Z (Fig. 3 (c)), moving in the direction of increasing depth. The experimental estimate of $\Delta_{Z,Y}(d)$ is plotted in black. The sign of the metric $\Delta_{Z,Y}$ defines the green and red shaded regions that we use to create our one-to-one blur to depth mapping.

they can be caused by inaccuracies in blur estimation. In this section, we model the sources of inaccuracies in Δ and show that there is an acceptable range where inaccuracies do not affect the derivation of the one-to-one mapping between blur and depth.

When analyzing different channels, we have so far assumed a fixed wavelength value (i.e. a narrow-band assumption). In practice, this is not accurate because these channels capture radiation across a certain range of wavelengths imposed by the sensor's color filter array. Looking at the intensity present in a channel W, the captured radiation may have any wavelength $\lambda_W \pm \delta_W$ where λ_W is the central wavelength and δ_W is a wavelength shift that is bound to the limits of channel W. In the extreme case, all radiation captured in the channel W has wavelength $\lambda_W + \delta_W^{max}$ instead of λ_W, where δ_W^{max} is the maximum deviation within band W. δ_W^{max} corresponds to a shift γ_W^{max} in focal length. We also call er_W the algorithmic blur estimation error for channel W, and obtain the predicted blur radius:

$$r'_W(d) = L\left|1 - \frac{x}{f_W + \gamma_W^{max}} + \frac{x}{d}\right| + er_W. \tag{9}$$

The resulting difference $\Delta'_{Z,Y}(d) \triangleq r'_Z(d) - r'_Y(d)$ is then given by:

$$\Delta'_{Z,Y}(d) = \begin{cases} L(\frac{x}{f_Y+\gamma_Y^{max}} - \frac{x}{f_Z+\gamma_Z^{max}}) + E_{Z,Y} & d \leq d_Y^{0'} \\ 2L(1+\frac{x}{d}) - L\left(\frac{x}{f_Y+\gamma_Y^{max}} + \frac{x}{f_Z+\gamma_Z^{max}}\right) + E_{Z,Y} & d \in [d_Y^{0'}, d_Z^{0'}] \\ L(\frac{x}{f_Z+\gamma_Z^{max}} - \frac{x}{f_Y+\gamma_Y^{max}}) + E_{Z,Y} & d \geq d_Z^{0'}, \end{cases} \tag{10}$$

where $E_{Z,Y} \triangleq er_Z - er_Y$, and $d_W^{0'}$ for $W \in \{Y; Z\}$, is given by:

$$d_W^{0'} = \frac{x}{\frac{x}{f_W+\gamma_W^{max}} - 1}. \tag{11}$$

The depth point of interest (d'_n) for our method is the point of intersection where the two channels show equal blur radii, yielding $\Delta'_{Z,Y}(d'_n) = 0$. Depth d'_n, assuming $E_{Z,Y}$ is not substantially large, is given by:

$$d'_n = \frac{2Lx}{L\left(\frac{x}{f_Y+\gamma_Y^{max}} + \frac{x}{f_Z+\gamma_Z^{max}}\right) - E_{Z,Y} - 2L}. \tag{12}$$

As seen in Fig. 4, a mistake in region mapping (defined by the sign of Δ) can cause the loss of the injective mapping only when $d'_n < d^0_Y$ or $d'_n > d^0_Z$. Shifts resulting in $d'_n \in [d^0_Y, d^0_Z]$ preserve a correct one-to-one mapping. This is because on the range $[d^0_Y, d^0_Z]$ both Y and Z channels have an injective mapping (overlap of the thick lines in Fig. 4) and either of them can be used. Therefore, errors between d_n and d'_n can be tolerated and do not affect the disambiguation process as long as $d'_n \in [d^0_Y, d^0_Z]$.

4. Experiments

In this section, we prove the correctness of our disambiguation solution with complex lenses. We then test our disambiguation results based on off-the-shelf blur estimators and provide proof-of-concept depth-from-defocus examples.

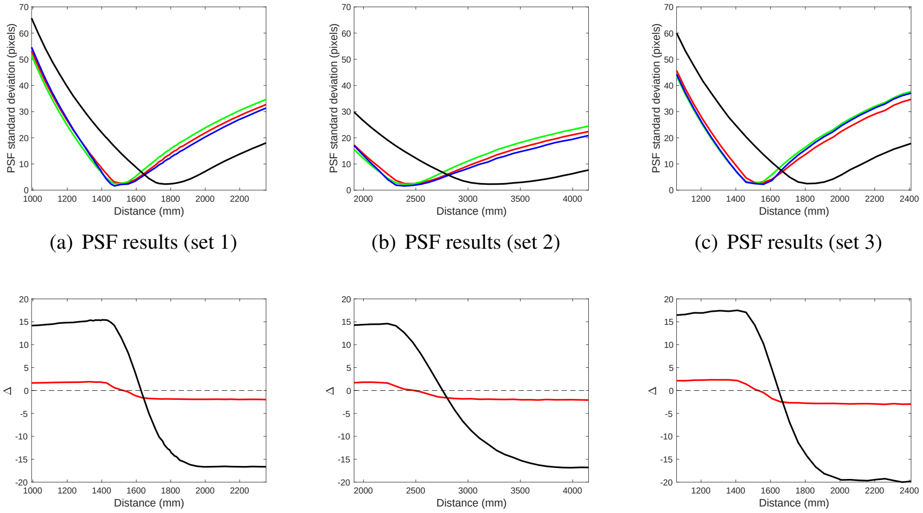
4.1. Experiments with Complex Lenses

We validate our theoretical assumptions, made through the adoption of the simple lens model equations, on more complex lenses. The objective of this section is to show that the relationships derived between defocus blur, depth and wavelength can extend to a typical lens, and thus demonstrate the generalization of our proposed metric Δ.

The defocus blur being directly linked to the PSF, we estimate the latter in our following experiments. To that end, we use a flat target consisting of a sharp 5° vertical slanted edge. We place the camera at varying distances from the slanted edge and capture an image at each location. Next, we compute the edge spread functions for each of the RGBN (RGB and NIR) channels by following the ISO 12233 standard [34]. Then, the line spread function (LSF) is derived through simple differentiation across the horizontal dimension. In accordance with the assumption of a Gaussian PSF, we fit a single Gaussian curve to the LSF. The standard deviation of the resulting Gaussian curve is finally retained per channel and per depth to represent the amount of depth-dependent and wavelength-dependent blur. With the symmetry imposed by the Gaussian PSF assumption, it is sufficient to estimate a single LSF to approximate the two-dimensional PSF. We repeat this experiment after focusing the camera at a different depth, and a third time using a different lens. Results are plotted in Fig. 5, for the three experiment sets acquired under the following settings:

- **Set 1**: Canon EF 50mm f/2.5 lens with f-stop of 2.5. Color is focused at depth 1513mm and we acquire 51 images over the range [993, 2353]mm.
- **Set 2**: Canon EF 50mm f/2.5 lens with f-stop of 2.5. Color is focused at depth 2390mm and we acquire 29 images over the range [1910, 4152]mm.
- **Set 3**: Canon EF 50mm f/1.8 II lens with f-stop of 2.5. Color is focused at depth 1509mm and we acquire 28 images over the range [1059, 2409]mm.

As expected, the amount of blur in a given channel increases monotonically as the camera is moved closer to the slanted edge, or away from it, relative to the depth where the channel was in focus. However, for our proposed method to generalize, the main property of our Δ metric must still hold. Essentially, this means that its value must change signs only once with increasing depth, and exactly in the depth range delimited by the two focal planes corresponding to the chosen channels. It is advantageous to compute Δ between a color channel and NIR due to the larger difference in blur magnitude that allows for a less robust blur estimation. This is due to the larger wavelength separation, and due to the uncorrected chromatic aberration in NIR. However, we also validate our assumptions even between spectrally adjacent channels, namely green and red. The Δ plots are shown in the bottom row of Fig. 5, where $\Delta_{R,G}$ and $\Delta_{NIR,G}$ are plotted in red and black, respectively. The general properties of Δ are preserved compared to those derived



(a) PSF results (set 1)

(b) PSF results (set 2)

(c) PSF results (set 3)

(d) Δ metric (set 1)

(e) Δ metric (set 2)

(f) Δ metric (set 3)

Fig. 5. Top row: blur magnitude as a function of depth for each of the RGBN channels, repeated over three image sets. The plots are in the corresponding colors and in black for NIR. Blur magnitude is estimated as the standard deviation of a Gaussian curve fitted to the edge spread function, which in turn is computed from a 5° slanted edge. Bottom row: the corresponding Δ plots in red for $\Delta_{R,G}$ and in black for $\Delta_{NIR,G}$.

through the simple lens model, with a larger shift between NIR and green. The focal planes of red, green and blue are very close to each other due to the chromatic aberration correction of our lenses. This correction is optimized for the visible channels, with the NIR rarely taken into account (only in superachromatic lenses [35, p. 105]). Despite this correction, and the closely-spaced red and green focal planes, $\Delta_{R,G}$ still crosses zero between the two, as desired.

4.2. Accuracy in Controlled Experiments

We further extend our experiments to test our proposed method on objects placed in the captured scene, and using off-the-shelf blur estimators. Although blur estimation is outside the scope of this paper, the objective of this section is to assess the applicability of our method in the absence of sharp slanted edges, or when a precise PSF estimation is not available. In other words, we test the method's robustness when simple blur estimation techniques are used.

For these tests, we used the same camera and lens combination as in Section 4.1, image set 1. The experimental setup consists of a ruler with a high-frequency triangular pattern printed on top, and a test object surrounded by four black circular markers. By using the triangular pattern on the ruler, we can visually see where the RGBN channels are in focus. The objects are then placed either closer than the color focal plane or deeper than the NIR focal plane. These two cases correspond to a positive or negative $\Delta_{NIR,G}$ (or $\Delta_{R,G}$) value respectively. As discussed in Section 3.2, when the object is located between the two focal planes, we have two options for obtaining a one-to-one blur to depth mapping. Therefore, we restrict our tests to the two regions outside the focal planes, and create four test sets. For each of the four test sets, we capture different objects in RGBN and use object segmentation to separate them from the background.

We compute $\Delta_{R,G}$ and $\Delta_{NIR,G}$ for every object in our dataset by subtracting the blur estimates obtained using the following methods: Hu [11], Zucker [9], Crété [36] and El Helou [20]. The blur estimation techniques perform best on edge locations, therefore, instead of interpolating

Table 1. Percentage accuracy results of the sign (positive or negative) of the metrics $\Delta_{R,G}$ (left) and $\Delta_{NIR,G}$ (right), determining the disambiguation regions. The results are on our four test sets and are based on four off-the-shelf blur estimators.

	Zucker [9]	**Hu** [11]	**Crété** [36]	**El Helou** [20]
Set A	69.2/71.2	61.4/81.7	80.4/93.2	63.4/79.5
Set B	81.4/74.3	63.2/77.3	70.9/88.6	68.8/75.8
Set C	66.9/88.4	68.0/90.6	89.1/96.5	73.5/85.1
Set D	86.9/80.0	65.1/89.0	71.0/95.0	77.1/89.3

the results to the rest of the image, which does not provide additional information, we assess accuracy solely over edge locations. For the methods that provide per-pixel blur maps (Hu [11] and Zucker [9]), we compute the percentage of correctly labeled pixels. For the methods that compute a single blur value per image (Crété [36] and El Helou [20]), we divide the object images into non-overlapping 32×32 pixel patches, and compute the percentage of correctly labeled patches. Finally, we compute the average accuracy in estimating the sign of $\Delta_{R,G}$ and $\Delta_{NIR,G}$ across all images in each set, and report the results in Table 1.

For sets A and B, we place 9 different objects once in front of the color focal plane (A) and once behind that of the NIR (B), respectively, and capture them one by one with an f-stop of 8. A correct labeling for set A is a positive Δ, since the green channel's focal plane is closer to the object's plane, which renders the object less blurry than in red or NIR. The opposite is true for set B. Note that the same labeling is true for $\Delta_{R,G}$ and $\Delta_{NIR,G}$, as the focal plane of the red channel is even closer than that of NIR, yet behind that of the green channel. We see that all algorithms achieve an accuracy higher than 60% with $\Delta_{R,G}$. Accuracies higher than 50% make the blur detection algorithm applicable to real-world scenes, where a majority vote across an object patch would yield a correct classification. Once the depth ambiguity is resolved using the sign of Δ, a one-to-one blur-to-depth mapping is trivially obtained.

Sets C and D are similar to the first two but the camera aperture is at f-stop 2.5, and we capture a set of 16 different objects. The increased aperture increases the blur intensity away from the focal plane. The magnified blur helps the blur estimation algorithms perform better in predicting the sign of Δ, thus the increased precision across all methods. Note how the smaller error margins of $\Delta_{R,G}$ compared with those of $\Delta_{NIR,G}$ (illustrated in the bottom row of Fig. 5) make it less robust to inaccurate blur estimation.

4.3. Proof-of-Concept Examples for Depth from Defocus

We evaluate our disambiguation solution on a set of 5.5 Mega Pixel images taken by a Canon Rebel T1i camera. We use the green and NIR spectral channels to make up for the blur estimation inaccuracies. We compute blur maps for these spectral channels using the algorithm presented in [11] with re-blur variance values of 10 and 12. The value of $\Delta_{NIR,G}$ is computed pixel-wise by subtracting the blur maps of the NIR and green channels, respectively. Because of inaccuracies in blur estimation, which is outside the scope of this paper, we make final decisions over patches of size 50×50 pixels to improve robustness. Instead of comparing Δ to 0, we compare it to $\pm t$ where t is a threshold chosen to be around half the amplitude of maximum blur difference in the image (reflecting an approximation for the theoretical $\pm\alpha$ defined in Eq. (3)). This is because blur cannot be well estimated with simple blur estimators in extremely low spatial frequency regions and results in $\Delta \approx 0$. The decision of whether $\Delta > t$, $\Delta < -t$ or $\Delta \in [-t, t]$ is a majority vote over pixel patches. We run the results through a ternary state cellular automaton, with equal

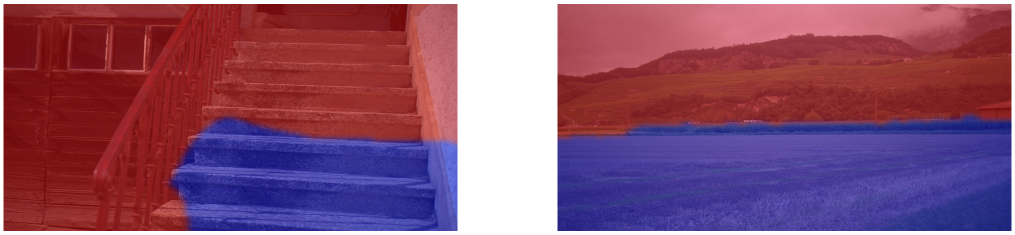
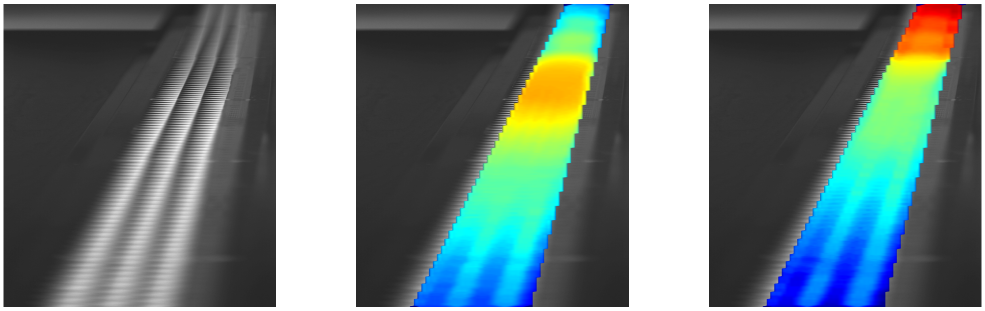


Fig. 6. Examples with the sign of $\Delta_{NIR,G}$ overlaid transparently. Blue corresponds to $\Delta_{NIR,G} > 0$ and red to $\Delta_{NIR,G} < 0$.



(a) G channel

(b) Depth [11]

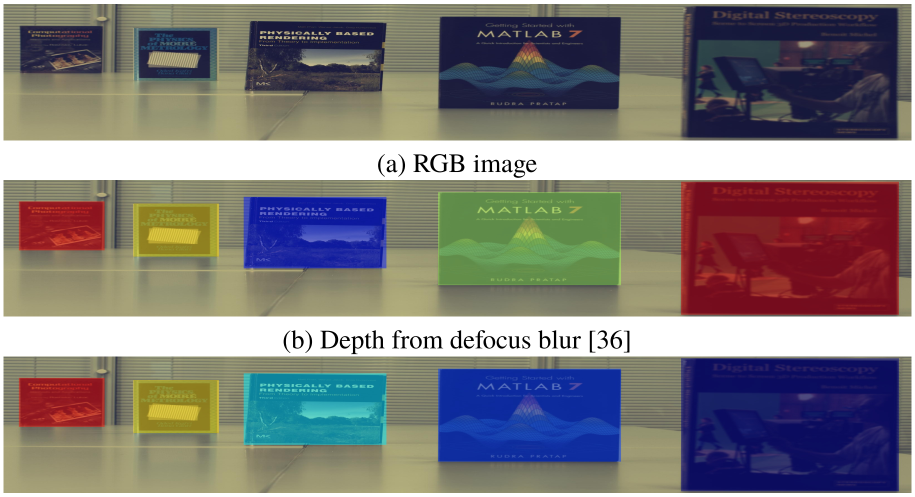
(c) Corrected depth

Fig. 7. (a) Ruler with continuously increasing depth. (b) The corresponding depth map by [11] cannot resolve the depth ambiguity: depth is incorrectly assumed proportional to defocus blur. (c) Depth map corrected using our $\Delta_{NIR,G}$ solution is proportional to blur on one side of the focal plane, and inversely related to it on the other side, yielding a correct depth from defocus, (depth values increase from blue to red).

influence from adjacent neighbors and interpolate all locations where $\Delta \in [-t, t]$ to either $\Delta > t$ or $\Delta < -t$. Two resulting maps are overlaid in Fig. 6. In almost all of our images, as is generally the case, we see that the focal planes are not at the extreme depths but rather around mid-range. This means that algorithms mapping defocus blur to depth are wrong over an entire subrange of scene depth. In our results, the edge separating the two Δ regions is not perfectly smooth. However, there is an acceptable margin of error between the two cases of positive or negative Δ that does not lead to ambiguity mistakes as explained in Section 3.3.

Finally, we provide two proof-of-concept depth from defocus examples that use our solution for the depth ambiguity problem. Fig. 7 shows an image of a ruler, shot horizontally with continuously increasing depth. From left to right are the original green image, the image with depth estimation [11], and the depth estimation with our correction. Using $\Delta_{NIR,G}$, we can distinguish the two ambiguous regions and deduce two injective mappings, Fig. 1 (c). Closer than d^0_{NIR} (namely $\Delta_{NIR,G} > 0$), NIR blur is inversely related to depth, and further from d^0_G (namely $\Delta_{NIR,G} < 0$), green blur is proportional to depth. Lastly, we construct the correct depth map by combining the maps obtained with green and NIR. Fig. 8 shows a similar example with a different blur estimator [36]. The color image is focused on the middle book, and, without our ambiguity correction, the depth of the closest two books is incorrectly estimated. In this example, we only use $\Delta_{R,G}$ to correct the mapping.

We limit these proof-of-concept examples of generalized depth-from-defocus to two images. A more thorough evaluation only tests the performance of the underlying blur estimator. As shown in Section 4.1, given an accurate enough blur estimation, our disambiguation solution is always correct. Section 4.2 shows how this accuracy drops depending on the blur estimator used,



(a) RGB image

(b) Depth from defocus blur [36]

(c) Corrected depth using $\Delta_{R,G}$

Fig. 8. (a) Five books placed with decreasing depth from left to right. (b) Depth from defocus blur based on [36], generated using the G channel and on every book. As in Fig. 7 (b), depth ambiguity cannot be resolved. (c) The depth map corrected using our $\Delta_{R,G}$ solution, (depth values increase from blue to red).

and this section provides visual proof-of-concept results with off-the-shelf blur estimators.

5. Conclusion

In this paper, we introduced a novel metric that resolves depth ambiguity. This metric is the pixel-wise difference of defocus blur between two spectral channels. Knowing the sign of the metric, a one-to-one mapping between defocus blur and depth can be obtained, irrespective of where the camera is focused during capture. This solution does not require any additional hardware, modified lenses, calibration, or learning, and does not affect image quality.

There are two main implications of our solution for depth ambiguity. First, depth from defocus can now be applied to previously captured photographs, irrespective of focus distance. Second, when an image is taken solely for depth-from-defocus estimation, it can be focused around mid-depth. This means that the maximum blur magnitude will be reduced compared to when the image is focused on the closest or deepest point in a scene. Knowing that blur estimation becomes less precise with large blur magnitudes, this permits more accurate blur estimation for depth from defocus as well as better quality images.

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. A. Wilson and H. Benko, "Combining multiple depth cameras and projectors for interactions on, above and between surfaces," in *ACM symposium on User interface software and technology,* (2010), pp. 273–282.
2. A. Canessa, M. Chessa, A. Gibaldi, S. Sabatini, and F. Solari, "Calibrated depth and color cameras for accurate 3d interaction in a stereoscopic augmented reality environment," J. Vis. Commun. Image Represent. **25**, 227–237 (2014).
3. E. Marchand, H. Uchiyama, and F. Spindler, "Pose estimation for augmented reality: a hands-on survey," IEEE Transactions on Vis. Comput. Graph. **22**, 2633–2651 (2016).
4. J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," in *European Conference on Computer Vision (ECCV),* (2014), pp. 834–849.

5. M. Mancini, G. Costante, P. Valigi, and T. Ciarfuglia, "Fast robust monocular depth estimation for obstacle detection with fully convolutional networks," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* (2016), pp. 4296–4303.
6. Q. Yang, "Local smoothness enforced cost volume regularization for fast stereo correspondence," IEEE Signal Process. Lett. **22**, 1429–1433 (2015).
7. M. Nia, E. Wong, K. Sim, and T. Rakgowa, "Stereo correspondence matching with clifford phase correlation," in *IEEE International Symposium on Robotics and Intelligent Sensors (IRIS),* (2015), pp. 193–195.
8. B. Park, Y. Keh, D. Lee, Y. Kim, S. Kim, K. Sung, J. Lee, D. Jang, and Y. Yoon, "Outdoor operation of structured light in mobile phone," in *ICCV Workshop,* (2017), pp. 2392–2398.
9. J. Elder and S. Zucker, "Local scale control for edge detection and blur estimation," IEEE Transactions on Pattern Analysis Mach. Intell. pp. 699–716 (1998).
10. S. Zhuo and T. Sim, "Defocus map estimation from a single image," Pattern Recognit. pp. 1852–1858 (2011).
11. H. Hu and G. Haan, "Low cost robust blur estimator," in *IEEE International Conference on Image Processing (ICIP),* (2006), pp. 617–620.
12. J. Lin, X. Ji, W. Xu, and Q. Dai, "Absolute depth estimation from a single defocused image," IEEE Transactions on Image Process. pp. 4545–4550 (2013).
13. F. Mannan and M. S. Langer, "What is a good model for depth from defocus?" in *IEEE Conference on Computer and Robot Vision (CRV),* (2016), pp. 273–280.
14. Y. Tai and M. S. Brown, "Single image defocus map estimation using local contrast prior," in *IEEE International Conference on Image Processing (ICIP),* (2009), pp. 1797–1800.
15. A. Sellent and P. Favaro, "Which side of the focal plane are you on?" in *IEEE International Conference on Computational Photography (ICCP),* (2014), pp. 1–8.
16. M. El Helou, M. Shahpaski, and S. Süsstrunk, "Closed-form solution to disambiguate defocus blur in single-perspective images," in *Mathematics in Imaging,* (Optical Society of America, 2019), pp. MM1D–2.
17. Y. Xiong and S. A. Shafer, "Depth from focusing and defocusing," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR),* (1993), pp. 68–73.
18. Y. Tai and M. Brown, "Single image defocus map estimation using local contrast prior," in *IEEE International Conference on Image Processing (ICIP),* (2009), pp. 1797–1800.
19. J. Garcia, J. Sanchez, X. Orriols, and X. Binefa, "Chromatic aberration and depth extraction," in *IEEE International Conference on Pattern Recognition (ICPR),* (2000), pp. 762–765.
20. M. El Helou, Z. Sadeghipoor, and S. Süsstrunk, "Correlation-based deblurring leveraging multispectral chromatic aberration in color and near-infrared joint acquisition," in *IEEE International Conference on Image Processing (ICIP),* (2017).
21. M. Chiang and T. E. Boult, "Local blur estimation and super-resolution," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR),* (1997), pp. 821–826.
22. A. Chakrabarti, T. Zickler, and W. T. Freeman, "Analyzing spatially-varying blur," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR),* (2010), pp. 2512–2519.
23. N. Joshi, R. Szeliski, and D. J. Kriegman, "Psf estimation using sharp edge prediction," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR),* (2008), pp. 1–8.
24. P. Peebles and R. Dicke, "Origin of the globular star clusters," The Astrophys. J. **154**, 891 (1968).
25. J. Ables, "Fourier transform photography: a new method for x-ray astronomy," Publ. Astron. Soc. Aust. **1**, 172–173 (1968).
26. E. E. Fenimore and T. Cannon, "Coded aperture imaging with uniformly redundant arrays," Appl. optics **17**, 337–347 (1978).
27. A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," ACM Transactions on Graph. **26**, 70 (2007).
28. S. Mahmoudpour and M. Kim, "Superpixel-based depth map estimation using defocus blur," in *IEEE International Conference on Image Processing (ICIP),* (2016), pp. 2613–2617.
29. O. Cossairt and S. Nayar, "Spectral focal sweep: Extended depth of field from chromatic aberrations," in *IEEE International Conference on Computational Photography (ICCP),* (2010), pp. 1–8.
30. P. Trouvé, F. Champagnat, G. Le Besnerais, J. Sabater, T. Avignon, and J. Idier, "Passive depth estimation using chromatic aberration and a depth from defocus approach," Appl. optics **52**, 7152–7164 (2013).
31. M. El Helou, F. Dümbgen, and S. Siisstrunk, "AAM: An assessment metric of axial chromatic aberration," in *IEEE International Conference on Image Processing (ICIP),* (2018), pp. 2486–2490.
32. A. Pentland, T. Darrell, M. Turk, and W. Huang, "A simple, real-time range camera," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR),* (1989), pp. 256–261.
33. A. Pentland, "A new sense for depth of field," IEEE Transactions on Pattern Analysis Mach. Intell. pp. 523–531 (1987).
34. "ISO 12233:2014, photography - electronic still-picture cameras. resolution and spatial frequency responses." .
35. E. Allen and S. Triantaphillidou, *The Manual of Photography and Digital Imaging* (CRC Press, 2012).
36. F. Crété, T. Dolmiere, P. Ladret, and M. Nicolas, "The blur effect: perception and estimation with a new no-reference perceptual blur metric," in *Electronic Imaging,* (2007), pp. 64920I–64920I.