Performance model of depth from defocus with an unconventional camera

Pauline Trouvé-Peloux, Frederic Champagnat, Guy Le Besnerais, Guillaume Druart, Jérôme Idier

Performance model of depth from defocus with an unconventional camera

P. Trouvé-Peloux[1,*], F. Champagnat[1], G. Le Besnerais[1], G. Druart[1], and J. Idier[2]

[1] DTIS, ONERA - Université Paris-Saclay, F-91123 Palaiseau, France
[2] LS2N (UMR CNRS 6004) BP 92101 - 1 rue de la Noë - 44321 Nantes Cedex 3, France
[*] Corresponding author: pauline.trouve@onera.fr

Compiled October 8, 2021

In this paper we present a generic performance model able to evaluate the accuracy of depth estimation using depth from defocus. This model only requires the sensor PSF at a given depth to evaluate the theoretical accuracy of depth estimation. Hence, it can be used for any (un)conventional system, using either one or several images. This model is validated experimentally on two unconventional DFD cameras, using either a coded aperture or a lens with chromatic aberration. Then we use the proposed model for the end-to-end design of a 3D camera using an unconventional lens with chromatic aberration, for the specific use-case of small UAV navigation. This paper is the accepted version of an article published in the *Journal of the Optical Society of America A* under the following DOI : https://doi.org/10.1364/JOSAA.424621

1. INTRODUCTION

Among depth estimation techniques, DFD is a passive monocular approach relying on the relation between depth and defocus blur. Fig. 1 illustrates this relation in the case of a simplified camera. Using geometrical relations and thin lens approximations, the width ϵ of the Point Spread Function (PSF) is:

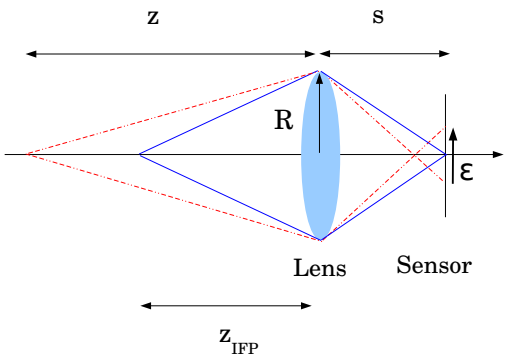
$$\epsilon = 2Rs\left(\frac{1}{f} - \frac{1}{s} - \frac{1}{z}\right), \tag{1}$$

where R and f are the camera aperture radius and focal length, respectively, s is the distance between the lens and the sensor, and z the distance between the lens and the point source.

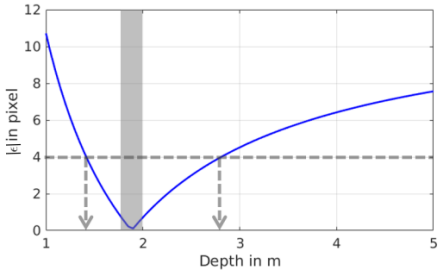
The DFD thus gives access to depth with an extremely simple and compact optical system. However, the simplicity of the experimental set-up comes at the cost of a complex algorithmic problem, since it is necessary to estimate the blur while the observed scene is unknown. In addition, the depth range over which the estimation is accurate is limited. As highlighted in Fig. 1 there a blind zone near the in-focus plane (IFP) and, moreover, the points situated in-front or behind this plane lead to identical blur sizes. To avoid ambiguities, the estimation is then often restricted to the depth domain located beyond the IFP.

In order to overcome these issues, much of the literature on DFD leverages on the acquisition of multiple images and/or use of unconventional optics. Some papers use multiple images with various camera settings [1, 2] or apertures [3]. The processing hence benefits from the fact that all the images are generated from the same scene. However, they require the scene to remain static during settings changes. Many recent works focus on single image DFD with coded aperture that reinforces the depth information contained within the defocus blur [4, 5]. References [6–8] exploit RGB images from a color camera equipped with an unconventional lens leading to a spectral variation of the relationship between blur and depth. This is obtained by changing the aperture shape with the color [6] or by using a lens with chromatic aberration [7, 8]. Some of these unconventional lens used for DFD are illustrated in Fig. 2.

An essential challenge is to characterize the performance of DFD solutions. This is useful for qualifying an existing DFD system, and also for designing such a system based on certain performance requirements. The brief review of the DFD that we have just done implies that such an analysis should highlight the variation in estimation precision with depth or, at least, provide the range of depth where estimation is accurate. It should also be able to take the parameters of unconventional optics into account, offering a tool to optimize them for a given use-case. In this paper, we propose a theoretical performance model meeting these requirements. This model applies to any (un)conventional, multiple or single image DFD system. We evaluate experimentally its reliability on two unconventional DFD systems using either a coded aperture or a lens with chromatic aberration. Finally, we exploit this model to derive an end-to-end methodology for the co-design of a DFD camera with a lens with chromatic aberration.



(a)



(b)

Fig. 1. Principle and limitations of DFD. (a) The image of a point source placed out of the in-focus plane (z_{IFP}) as a geometrical size of ϵ which depends on the depth z. (b) Example of defocus blur variation with respect to depth according to Equation (1). The blind area around the IFP and the depth ambiguities on either side are highlighted in gray on the curve.

2. STATE OF THE ART AND CONTRIBUTIONS

Here we briefly review previous works on performance evaluation of conventional and unconventional DFD systems, and then outline our contributions.

A. Conventional DFD cameras

DFD performance is highly related to sensor parameters. Several papers have proposed models for choosing the best camera settings in multiple image DFD [9–13]. A depth estimation accuracy analytical formula is derived in references [10, 11] taking into account geometrical defocus blur, diffraction and pixel sampling for one in-focus and one out-of-focus image. Hence only the influence of sensor parameters are considered. In references [9, 12, 13] a Cramér Rao Bound (CRB) is derived to optimize the blur ratio between a pair of images of a conventional camera to maximize the DFD accuracy. These models take the influence of processing on precision into account. However, they are derived from a Gaussian or a pill-box PSF model and therefore cannot be used for unconventional optics. Based on the CRB theory, we have proposed a generic performance model of single image DFD which is able to account for various models of sensor PSF, and underlines the relation between blur size and estimation accuracy[14].

B. Unconventional DFD cameras

Performance models dedicated to unconventional DFD camera have been previously proposed [3–5, 15, 16]. In Levin et al. [4] the coded aperture is optimized using the average Kullback Leibler divergence, between each potential depths. In Zhou et al.[3] is defined a cost function that measures the inconsistency between the two defocused images when the estimated blur deviates from the ground truth. Optimal apertures are those for which the inconsistency is maximal, meaning that the coded apertures increase the ability to discriminate defocus blurs. Extending this work, Levin [15] proposes a discrimination score integrated in the frequency domain for any set of coded apertures. In Martinello et al. [5], blurring is interpreted as a projection of the data onto a subspace which is learned on simulated data for each potential depth. The coded aperture is optimized in order to maximise the distance between the kernels of each subspace. In Sellent et al. [16] the coded aperture is optimized by maximization of the difference between the blurred images at different depth levels.

These design approaches consider only a restricted part of the system, here the aperture shape. They only provide a global score on a given camera, that cannot be physically interpreted as a depth estimation accuracy nor provide information on the depth range of operation of the system.

C. Contributions

We present a generalisation of the performance model described in reference [14] to the case of an unconventional camera. This model can be used with any single or multiple images, conventional or unconventional camera. It captures an important feature of DFD methods, that is the variation of accuracy with respect to depth itself. The reliability of the model is validated experimentally on two unconventional DFD camera in section 4 and 5. Finally we leverage this model to conduct the end-to-end co-design of a chromatic 3D camera dedicated to the use-case of small UAV navigation in section 6, taking into account jointly image quality and depth estimation performance models. In Section 7 we discuss the conclusions and the perspectives of this work.

3. GENERIC DFD PERFORMANCE MODEL

In this section we describe the proposed generic performance model for DFD. It is based on a Cramér Rao lower bound derived, within a Bayesian framework, from a scene prior and the sensor PSF(s) at a given depth.

A. Image formation model

Defocus blur is a spatially varying blur. The standard "convolution and additive noise" model of observation is therefore applied to image patches where the PSF of the sensor (i.e. here the depth) can be considered constant. A generic formulation of the DFD problem is simply :

$$\boldsymbol{Y} = H_z \boldsymbol{X} + \boldsymbol{N}, \tag{2}$$

where $\boldsymbol{Y}$ (respectively $\boldsymbol{X}$) is a vector that collects pixels of the image (resp. scene) patch(es) in the lexicographical order. $\boldsymbol{N}$ stands for the noise process, which is modeled in this paper as a zero mean white Gaussian noise with variance σ_N^2. H_z is convolution matrix with non zero parameters corresponding to the samples of the PSF(s) describing the DFD system. It has a block Toeplitz structure, each block being itself Toeplitz [Section 4.3.2 17]. Note that as we consider small patches, some care has to be taken concerning boundary hypotheses. In particular the usual periodic model associated with Fourier approaches is not suited here. In the sequel we use "valid" convolutions where the support of the scene is enlarged with respect to the one of the image according to the PSF support [Section 4.3.2 17]. In

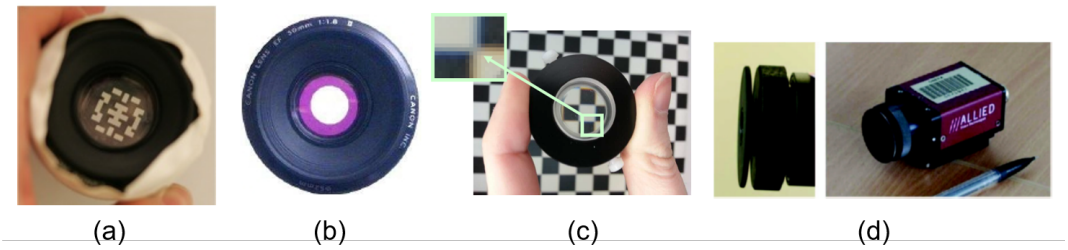


Fig. 2. Examples of unconventional optic devices used for DFD. (a): coded aperture of [4], (b): colored aperture shape of [6], (c): add-on with chromatic aberration of [8] and (d): the 3D camera with chromatic aberration of [7] which is further studied in Section 6.

sections 4.1 and 5.1, a more precise description of quantities $\boldsymbol{Y}$, $\boldsymbol{X}$ and H_z will be given, specific to the considered application.

B. Generic scene prior and data likelihood

In the context of local PSF estimation, a Gaussian prior on the scene is often very effective [18–20]. Hence, we also propose to use a scene Gaussian prior written as:

$$p(\boldsymbol{X};\sigma_X^2) \propto \exp\left(-\frac{\|D\boldsymbol{X}\|^2}{2\sigma_X^2}\right). \tag{3}$$

Matrix D refers to a derivative operator and will be discussed in more details in Sections 4 and 5 depending on the sensor type. The only parameter of the prior is the variance σ_X^2, which corresponds to the variance in pixel level of scene variations between neighbour pixels. As we assume a centered Gaussian random noise of variance σ_N^2 also in pixel level, the likelihood of the observation $\boldsymbol{Y}$ reads:

$$p(\boldsymbol{Y}|\boldsymbol{X};\sigma_N^2) \propto \exp\left(-\frac{\|\boldsymbol{Y}-H_z\boldsymbol{X}\|^2}{2\sigma_N^2}\right). \tag{4}$$

C. Marginalized likelihood and Fisher information

Akin to reference [14], we derive a marginalized data likelihood with respect to the scene [18–20], which is tractable for the Gaussian prior of Equation (3):

$$p(\boldsymbol{Y};\boldsymbol{\theta}) = \left|\frac{Q_{\boldsymbol{\theta}}}{2\pi}\right|_+^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{Y}^t Q_{\boldsymbol{\theta}} \boldsymbol{Y}\right), \tag{5}$$

where $\boldsymbol{\theta} = \{z, \sigma_N^2, \sigma_X^2\}$ and $|Q_\theta|_+$ is the product of the non zero eigenvalues of Q_θ which can be written as:

$$Q_{\boldsymbol{\theta}} = \frac{1}{\sigma_N^2}\left[I - H_z(H_z^tH_z + \alpha D^tD)^{-1}H_z^t\right]. \tag{6}$$

Parameter $\alpha = \sigma_N^2/\sigma_X^2$ can be interpreted as the inverse of a signal to noise ratio. Now by writing $P_\psi = \sigma_N^2 Q_\theta$ and $\psi = \{z, \alpha\}$ one can evaluate the Fisher Information matrix:

$$\mathrm{FI}(\psi) = \frac{1}{2}\mathrm{tr}\left(P_\psi^+ \frac{dP_\psi}{d\psi} P_\psi^+ \frac{dP_\psi}{d\psi}\right). \tag{7}$$

Where $+$ denotes the pseudo-inverse. Details on the derivation of the marginalized likelihood and the Fisher information matrix, specially in the case of D being singular, can be found in reference [14]. The Cramér-Rao bound on the standard deviation of the depth estimation can be deduced from the Fisher information matrix by:

$$\sigma_{\mathrm{CRB}}(\psi) = \mathrm{FI}(\psi)^{-1/2}. \tag{8}$$

It provides a computable performance index with a clear physical interpretation for DFD depth estimation.

D. Computation of the performance model

In this paper our aim is to characterize the depth estimation accuracy in DFD, hence to simplify we assume that the signal to noise ratio (i.e., α) is known and focus only on depth estimation. This amounts to assume that $\psi = \{z\}$. For a depth z and given a small depth variation δ, we compute the PSFs at respectively $\{z, z-\delta, z+\delta\}$ using an optical model such as a simple Gaussian model, Fourier optics, or an optical design software (see Appendix A). Then, the convolution matrices H_z,$H_{z-\delta}$,$H_{z+\delta}$ are derived according to the image formation model which varies with the DFD application. Given a value of α, that sets the signal to noise ratio, the matrices P_z, $P_{z+\delta}$ and $P_{z-\delta}$ are computed using :

$$P_z = I - H_z(H_z^tH_z + \alpha D^tD)^{-1}H_z^t. \tag{9}$$

As in the generic case no analytical formula for the derivative of P_z over z is available, we compute the derivative that appears in Equation (7) using the central finite difference:

$$\frac{dP_z}{dz} \simeq \frac{P_{z+\delta} - P_{z-\delta}}{2\delta}. \tag{10}$$

After the computation of the pseudo-inverse of P_z, the Fisher information is given by Equation (7). Finally, taking the inverse square root of the result according to Equation (8) gives the theoretical minimum standard deviation $\sigma_{\mathrm{CRB}}(z)$ of depth estimation at the current depth z.

4. APPLICATION TO DFD WITH A CODED APERTURE

In this section we use the proposed performance model in the case of depth estimation using a single image from a camera with the coded aperture proposed in [4] and illustrated in Figure 2. In order to evaluate the reliability of the proposed model, experimental validations are conducted to compare the theoretical and the experimental performances.

A. Image formation model

In the case of single image DFD, $\boldsymbol{X}$ (resp. $\boldsymbol{Y}$) of equation (2) simply concatenates the M (resp. L) pixels of a scene (resp. an image) patch. H_z is then directly the convolution matrix of size $L \times M$ relative to "valid" convolution with the PSF associated to depth z [Section 4.3.2 17].

Fig. 3. Lens with the coded aperture from Levin et al. [4].

B. Scene prior

As in reference [14], D is defined as the concatenation of the convolution matrices corresponding to the vertical and horizontal first order derivative, i.e., the convolution matrices relative to filters $[-1\ \ 1]$ and $[-1\ \ 1]^T$. This model, which can be physically interpreted as a $1/v^2$ decrease of the scene spectrum, has previously shown good results in single image blur identification [18, 20]. Note that matrix D is singular, as $D\mathbf{1} = 0$, with $\mathbf{1}$ corresponding to a homogeneous patch of pixels equal to 1. In such a case, the scene prior is said to be improper. However depth inference can still be derived from such a prior (see reference [14]).

C. Comparison of coded aperture and conventional aperture DFD performance

C.1. Camera settings

We consider here two identical lens of focal 35 mm used with a Nikon D200 camera, having pixels of size 6μm. The coded aperture shown on the left image of Fig. 2 is inserted within the aperture of one of the lens. For the other lens, the f-number is fixed at 3.2 so that the aperture has the same size as the coded aperture. For both lenses the camera in-focus plane is set at 1.5 m. To avoid unwanted effects of demosaicking on defocus blur, we consider here a subsampled image extracted by taking one of the two green pixels from the raw data. The processed image resolution is then divided by two with respect to the full sensor resolution. Figure 3 shows the lens with the coded aperture.

C.2. Theoretical performance

To simulate the camera PSF, we use a spatial integration over the sensor pixel of the optical PSF given by Fourier Optics as described in Appendix A. As we process the green channel extracted from the raw data, in the PSF simulation the pixel size is simply assumed to be twice as large as the actual pixel size. We also neglect chromatic aberration and simulate the PSF at the wavelength 532 nm only. Figure 13 shows the variation of σ_{CRB} calculated using simulated PSF, with and without the coded aperture, for a patch size of 25×25 pixels and $\alpha = 0.001$. In the region from 1.8 to 2.2 m, the performances of the two configurations are similar, with an increase of the theoretical standard deviation. Then after approximately 2.2 m, the coded aperture clearly shows better performance than the conventional one, which confirms the results of a performance gain using the coded aperture rather than a conventional one [4]. We can see here that this gain is getting more significant as the defocus blur increases.

It should be emphasized that the influence of the coded aperture on the DFD performance varies with the depth of observation. This is in favour of a performance model depending on the depth, rather than a global score.

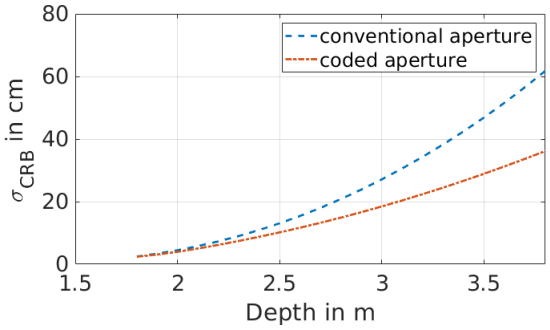


Fig. 4. Theoretical depth estimation accuracy of a camera with a conventional or the coded aperture proposed in [4].

C.3. Experimental performance

Experimental setup Experimentally, the PSFs of the lenses with and without the coded aperture are calibrated on axis from 1.6 m to 4 m with a step of 5 cm using the calibration pattern and codes from Delbracio et al. [21]. This method relies on acquisition of a known high frequency texture and estimation of the PSF using an inverse problem algorithm. True depth value is given by a telemeter. To evaluate depth estimation, a textured planar scene is put at different distances from the lens and an image is acquired. This scene, shown in image (b) of Figure 13, is made of a collection of patches generated to follow the Gaussian prior of Equation (3). Experimentally, the signal to noise ratio is maintained comparable for both lenses by increasing the integration time for the coded aperture case.

Depth estimation results At each scene position, depth is estimated using the algorithm presented in Appendix B on patch size of 25×25 pixels, within a central crop of the image. Standard deviation and mean value of the depth estimation are calculated on 17k patches with 50% overlapping extracted from this crop. Figure 5 shows experimental error bars of depth estimation with respect to the true depth, measured with a telemeter: (a) with the coded aperture and (b) with the conventional aperture. Figure 5 (c) shows root mean square error (RMSE), for both configurations with respect to depth.

Without the coded aperture the depth estimation seems correct near the in-focus plane, but it shows a significant increase of the standard deviation after 2.5 m, then bias becomes predominant after 3 m. For the coded aperture, the bias remains negligible on all the studied depth range, with a regular increase of the standard deviation, as expected in the theoretical performance model. The comparison of theoretical and experimental performance curves shows that the theoretical model allows a reliable comparison, depth by depth, of the influence of the different forms of apertures on the DFD. However, the theoretical standard deviations σ_{CRB} appear lower than the empirical ones. First, it should be noted that the theoretical model depends on the choice of the noise standard deviation σ_N which may be too optimistic here. Besides, bias is not considered in the derivation of the CRB, while it does contribute to the experimental error. Ultimately this difference comes from the fact that the model is a simplification of reality: the scene does not exactly follow the prior distribution, the acquisition depends on non-modeled phenomena (aberrations, variable illumination), the PSF are estimated by calibration, etc. Nevertheless, we show here that the theoretical model can provide a useful prediction of the relative performance of two DFD systems over a range of depths.

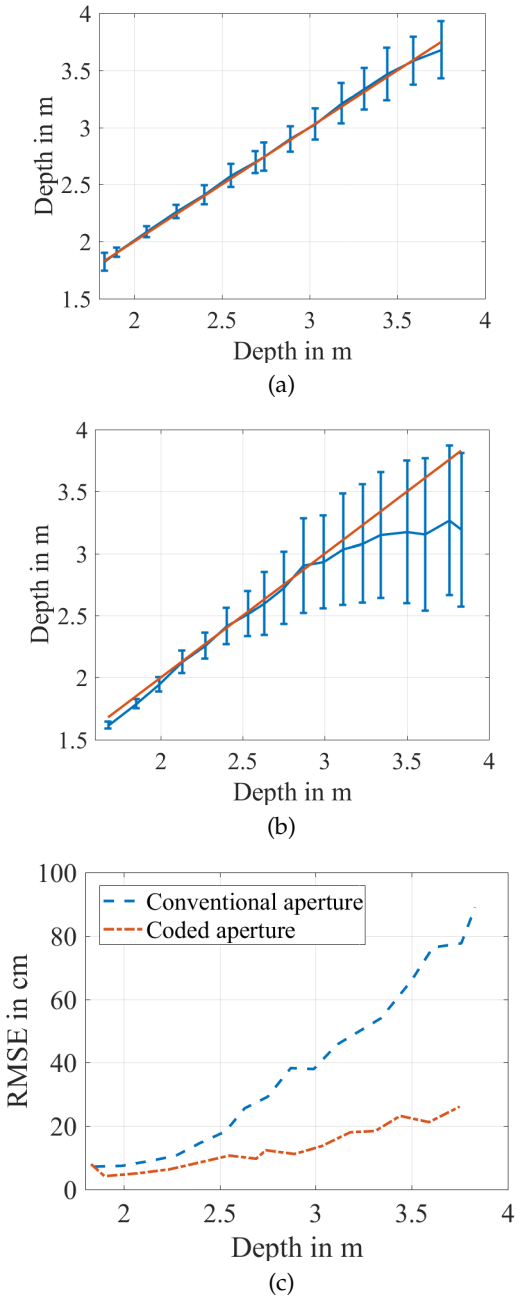


Fig. 5. Depth estimation experimental error bars using (a) a coded aperture and (b) a conventional camera having the same focal length, f-number and focus.(c) RMSE with and without coded aperture.

5. APPLICATION TO DFD WITH A LENS WITH CHROMATIC ABERRATION

In this section, we use the proposed performance model to evaluate the theoretical performance of a camera with a lens having chromatic aberration and a color sensor. We compare the performance of two focus setting of such a camera.

A. Image formation model with a color sensor

Assuming a color sensor, in a single acquisition are produced three sub-images corresponding to the red, green and blue pixels of the sensor. As we consider a lens with chromatic aberration, there is a different PSF for each color channel — this is precisely the benefit of chromatic DFD where each depth is encoded by a triplet of PSFs [7]. At the same time, each color channel sees a different spectral bandwidth of the scene. When dealing with colored data the components in the RGB decomposition are usually correlated. Hence a separable scene prior on RGB components is not suited. Following [7, 22] we propose to use the luminance (L) and the red-green (C_1) and blue-yellow chrominance (C_2) decomposition instead of the RGB decomposition using the transform:

$$\begin{bmatrix} \boldsymbol{x}_R \\ \boldsymbol{x}_G \\ \boldsymbol{x}_B \end{bmatrix} = T \begin{bmatrix} \boldsymbol{x}_L \\ \boldsymbol{x}_{C_1} \\ \boldsymbol{x}_{C_2} \end{bmatrix} \tag{11}$$

$$\text{with } T = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{bmatrix} \otimes \mathbf{I}_M, \tag{12}$$

where $\otimes$ stands for the Kronecker product and $\mathbf{I}_M$ is the identity matrix of size $M \times M$. According to reference [22], the three components of the luminance/chrominance (LC) decomposition can be assumed uncorrelated. Let $\boldsymbol{X}^t = [\boldsymbol{x}_L, \boldsymbol{x}_{C_1}, \boldsymbol{x}_{C_2}]$ be vector of length $3M$ made of concatenation of the luminance and chrominances vectors, and $\boldsymbol{Y}^t = [\boldsymbol{y}_R, \boldsymbol{y}_G, \boldsymbol{y}_B]$, the vector of length $3N$ that concatenates the R,G and B color vectors. Then the observation model (2) applies with matrix H_z of size $3L \times 3M$ defined as:

$$H_z = \begin{bmatrix} H_R(z) & 0 & 0 \\ 0 & H_G(z) & 0 \\ 0 & 0 & H_B(z) \end{bmatrix} T. \tag{13}$$

where each H_C matrix corresponds to the $N \times M$ convolution matrix associated to the PSF of channel C. Note that most of the color sensors are actually made of a set of R,G,B color filters regularly organized in front of the sensor pixels according to the Bayer pattern. Hence, the RGB color channels extracted from the raw data are actually subsampled versions of the full sensor data. To model this sampling, two approaches are possible. First, one can remove adequate lines from the convolution matrices H_R, H_V and H_B corresponding to the missing pixels. The second approach is to model the system as a 3CCD sensor and to double the size of the sensor pixel size. This reduces the PSF size with a factor of 2. As the size of the convolution matrix depends on the PSF size, to limit the computational cost, we use the second approach here.

B. Scene prior

Assuming that the luminance and chrominance decompositions are uncorrelated, we propose to use the generic Gaussian prior defined in (3) with D such as:

$$D = \begin{bmatrix} \sqrt{\mu_c} D_0 & 0 & 0 \\ 0 & D_0 & 0 \\ 0 & 0 & D_0 \end{bmatrix}. \tag{14}$$

D_0 is the vertical concatenation of the convolution matrices relative to the horizontal and vertical first order derivation operator, and μ is the ratio of the luminance and the chrominance

variances, fixed at 0.04 as in [7]. Note that D has three zero eigenvalues, one for each component.

C. Performance comparison of two settings of a chromatic DFD camera

C.1. Camera settings

We consider the camera introduced in reference [7], see image (d) in Figure 2, and study the influence of the in-focus plane (IFP) position. The camera lens has chromatic aberration. Its focal length for the green channel is of 25 mm with an f-number of 4 and the amount of longitudinal chromatic aberration is of 200 μm. It is mounted on a Stingray F-504 color camera of pixels 3.45 μm.

C.2. Theoretical performance

We consider two different camera settings: the IFP of the green channel put at 2.8 m (denoted IFPG2.8 in the following) which corresponds to the original setting of [7], or put at 3.8 m (denoted IFPG3.8). Change of focus are obtained by modifying the position of the sensor and PSFs are then directly extracted from the optical design software Zemax available from our previous work [7].

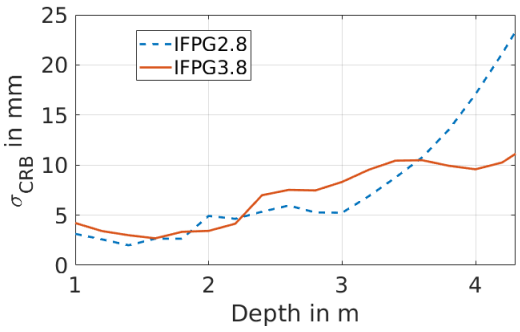


Fig. 6. Theoretical depth estimation results using two different focus of the same chromatic camera and the same illumination conditions.

Figure 6 shows the theoretical performance of both settings. The influence of a change of focus is variable with the depth range that is considered. Both configurations show similar performance before 2.5 m, then in the (2.5, 3.5) m interval the IFPG2.8 setting has a slightly better performance. After 3.5 m IFPG3.8 takes over, while the σ_{CRB} of IFPG2.8 begins to increase sharply.

C.3. Experimental performance

Experimental setup Experimentally, for the two focus settings, the camera PSF(s) are calibrated using the method proposed in Delbracio et al. [21] using a known high frequency pattern put at different distance from the camera. Then as in Section C.3, a fronto parallel textured pattern (corresponding to the target (a) of Figure 13) is put on a tripod at various depths given by a telemeter from the camera.

Depth estimation results Depth is estimated within a central crop of the acquired images using the DFD algorithm presented in Appendix B and patch size of 21 $\times$ 21 pixels. Standard deviation and mean value of the depth estimation are calculated on the depth estimation results from which we remove 5% of outliers. According to Figures 7(a) to (c) both configurations show similar RMSE before 3 m, with a slightly better performance for IFPG2.8. After 3.5 m the performance of the original focus setting IFPG2.8 the performance of depth estimation starts to degrade after 3.5 m, while it is still acceptable on the whole tested depth range for IFPG3.8. These experimental results are consistent in relative value with the theoretical performance shown in Figure 6.

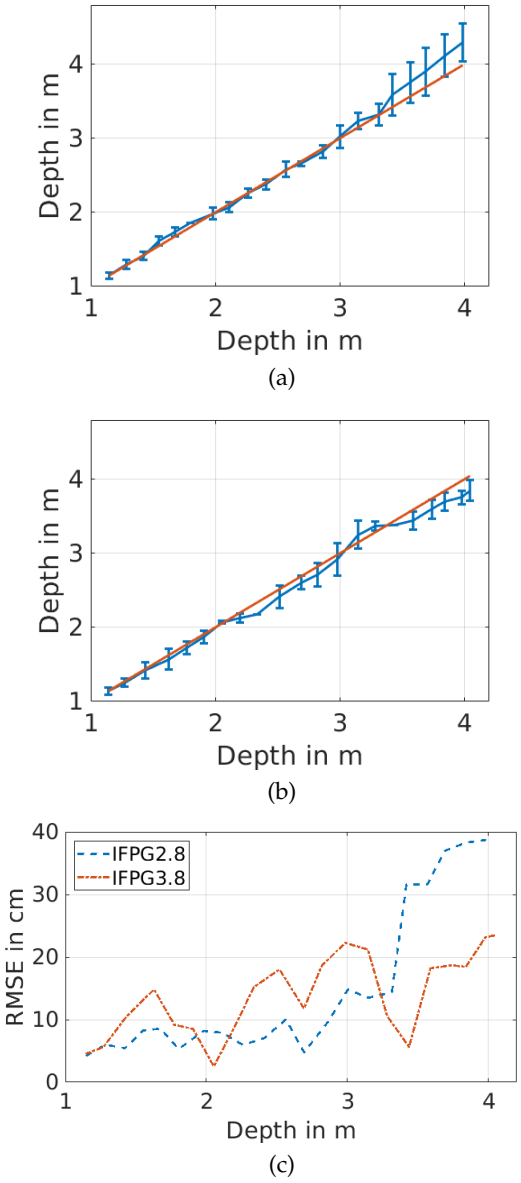


Fig. 7. Experimental depth estimation results using (a) the original focus of the green channel at 2.8 m (IFPG2.8) of [7],(b) focus at 3.8 m (IFPG3.8). (c) Comparison of experimental RMSE for both settings.

As for the coded aperture case, this experiment validates the ability of the proposed theoretical model to predict the best settings for a chromatic DFD system.

6. APPLICATION IN THE END-TO-END DESIGN OF A CHROMATIC DFD CAMERA

In this section, the performance model is applied to the end-to-end co-design of a DFD camera (this work has been partially presented in [23]). The camera is designed for the specific use-

case of a small UAV navigation. We first present the proposed co-design methodology before describing the chosen use-case and then the results of each design step. In contrast with previous co-design works, we use two different performance models here, one for depth estimation, and the other for image quality, and discuss the trade off between these requirements. Finally we present experimental results of the co-designed camera.

A. Co-design proposed methodology

Starting from scratch, we propose the following template procedure for the co-design of the camera. First, we define the requirements on the camera from the use-case, then we choose an optical DFD concept and algorithm. Then, we define design criteria and conduct a preliminary system specification. Finally, we fine-tune of the system parameters.

B. Use-case description

We consider the design of a compact 3D camera for the payload of a small UAV with autonomous flight capabilities in close range outdoor or indoor environments. The imaging system should allow the UAV to detect, recognize and avoid objects in front of it including thin objects such as electric wires or posts. Hence, we aim at an accurate depth estimation close to the image axis, while maintaining a reasonable image quality for recognition tasks. Taking into account the speed and evasive capabilities of the UAV, the depth estimation range is fixed at (1-5) m with a required depth accuracy of 10 cm, and a field of view of 25°. To simplify the co-design, we fix the sensor type before the lens optimization. We choose a Stingray F-504 color sensor which has a pixel size of 3.45μm with a resolution of 2046×2452 pixels.

C. Requirements from the use-case

As the sensor parameters and the field of view are fixed, the lens focal length is then of 25 mm. We choose a f-number of 3 in order to have sufficient light intensity to use the camera for indoor and outdoor scenes without having too strong optical design constraints. To be able to identify thin obstacles, the depth map spatial X-Y resolution is fixed to approximately 2 cm at 3 m. Thus the depth map spatial resolution has to be around 160 μm in the image plane. This resolution limits the patch size to 46×46 pixels on the sensor.

D. Choice of an optical concept and algorithm

As we have chosen a color camera, in particular to facilitate recognition tasks, we turn towards chromatic DFD since it enlarges the depth estimation range thanks to the one-to-one relation between depth and defocus blur triplet. On the other hand, chromatic aberration reduces image quality, and a restoration process has to be included in the image processing to make recognition tasks tractable. The proposed vision payload is then made of a lens with chromatic aberration and two image onboard processing softwares: one for depth estimation and one for image restoration. The DFD algorithm is the one presented in Appendix B in the case of a color sensor. As for image quality restoration, we propose to use a high frequency transfer guided by the estimated depth map, as in reference [7].

E. Joint performance models

We define two performance models for the camera, one that evaluates the depth estimation accuracy, and the other the image quality.

E.1. Depth estimation accuracy

In order to optimize depth estimation for some given depth range D_r we propose a design criterion named C_1 based on the mean value over D_r of the σ_{CRB} described in Equation (8):

$$C_1(D_r) = < \sigma_{\text{CRB}}(z) >_{z \in D_r} . \tag{15}$$

E.2. Image quality

To manage the high frequency transfer, one needs to have at least a sharp channel at each depth. This property can be related to the depth of field (DOF), whose formal definition can vary with the chosen optical model. In the case of a camera with chromatic aberration, there is a different DOF for each color channel. Thus we define an image quality criterion that measures the union of these DOFs inside the sought camera depth range D_r:

$$C_2(D_r) = \text{GDOF} = D_r \bigcap \left(\bigcup_{c=R,G,B} \text{DOF}_c \right) . \tag{16}$$

This criterion C_2 can be interpreted as a generalized depth of field (GDOF) of the camera after image restoration, a quantity illustrated in Figure 8.

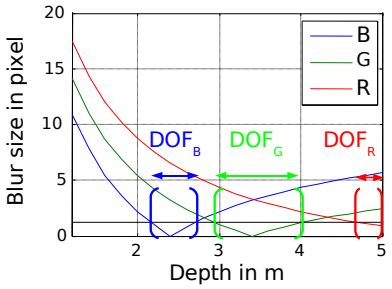


Fig. 8. Illustration of the image quality criterion generalized depth of field (GDOF): the union of the R,G,B depth of fields (DOF). Note that in this example, the GDOF does not stand within a single continuous depth range.

F. Preliminary system specification

One needs to have a rough idea of the amount of chromatic aberration that is required for the system. To explore efficiently a large domain of possible systems we use a simple model of a lens having chromatic aberration in this preliminary design. The PSF for each color channel are Gaussian, with standard deviation $\sigma = \rho\epsilon$, with ϵ given by Equation (1) and $\rho = 0.3$[14]. In this case, the DOF is simply defined as the depth range where $|\epsilon| \leq t_{px}$. We simulate the PSFs associated to various chromatic imaging systems, having a focal length of 25 mm at the green channel, an f-number of 3, and a pixel size of 6.9 μm, which corresponds to twice the original pixel size, in order to simply take the Bayer pattern into account. Each system has a different triplet of RGB in-focus planes obtained by variations of the R and B focal lengths and sensor position. We calculate the criteria C_1 and C_2 in the depth range from 1 to 5 m for each potential system. We use a patch size of 23×23 pixels and $\alpha = 0.001$.

However, maximisation of C_2 or minimisation of C_1 do not lead to the same in-focus planes. Hence, a trade-off has to be found. Here the critical purpose is obstacle avoidance, hence the accuracy of the depth estimation is our prime criterion. Therefore we choose to reorder the triplets according to increasing σ_{CRB} and select the triplets having a value of C_1 less than 10%

above the minimal value of C_1. We select the triplet having the maximum value of C_2 among these selected triplets. This corresponds to a longitudinal chromatic aberration $df = f_R - f_B$ around 130 μm and the green channel focused around 3.4 m.

G. Fine-tuning of the system parameters

The first order lens parameter optimization gives the approximate optimal position of the RGB in-focus planes and the required amount of longitudinal chromatic aberration. According to these constraints, a first architecture is designed using the optical software `Zemax`. This architecture shown in Figure 9 is inspired from a Double Gauss reference architecture, where doublets become single lenses to reduce the number of lenses.

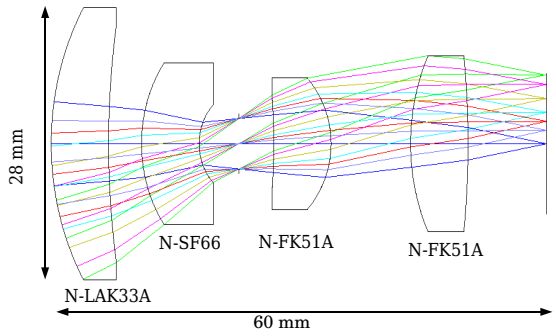


Fig. 9. Architecture of the co-designed lens.

The symmetrical lens position with respect to the aperture avoids odd order aberrations such as lateral chromatic aberrations. In contrast to Section F, we now deal with the real physical lens parameters such as lens curvature radius, thickness or glass type. We optimize these parameters starting from the initial architecture. The spectral bandwidths of each color channel are taken into account during the optimization using three wavelengths for each color. To slightly diversify the positions of the R,G,B in-focus planes a new optimization of the lens is conducted with imposed chromatic aberration amounts (df) near the optimized value obtained in Section F. For each value of chromatic aberration, and for different sensor plane positions near the original position, the criterion C_1 and C_2 are then evaluated. To do so, we use the optical polychromatic PSF simulated by `Zemax` for each configuration, integrated according to a sensor of pixel size of 6.9 μm, to get a simplified model of the Bayer pattern. Here, the optical design software can directly provide the FTM of the system at any wavelength, so we evaluate the C_2 criterion as the depth range where 50% of the PSF encircled energy is below one pixel. Note that this is not a severe threshold as we only require a reasonable image quality here. Figure 10 presents the variation of C_1 and C_2 obtained for each system.

As for the case of preliminary system optimization with simulated Gaussian PSFs, we observe that the setting maximizing the generalized depth of field does not fit with the setting that minimizes the depth estimation standard deviation. We finally choose a trade-off having a longitudinal chromatic aberration of 100μm, with RGB in-focus planes respectively at 2.2, 2.6 and 3.8 m, that corresponds to the black rectangle in Figure 10. Figure 11 shows the theoretical performance of the optimized system in the depth range 1 to 5 m.

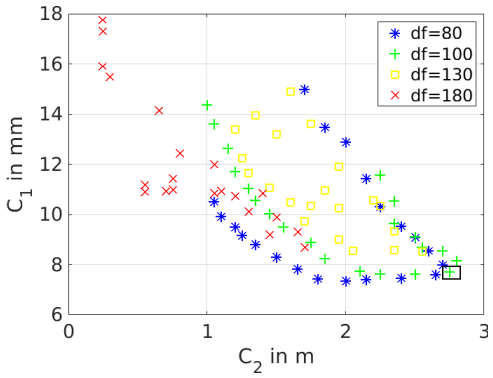


Fig. 10. C_1 and C_2 scores obtained for systems simulated using Zemax having various amount of chromatic aberration (df is in μm corresponds to the focal length difference between the red and the blue color channel).

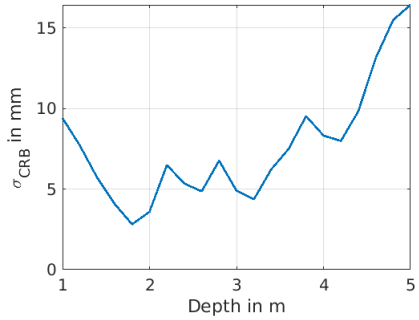


Fig. 11. Theoretical performance of the codesigned lens.

H. Experimental validation

H.1. Experimental setup

We have realized the co-designed lens, according to the specifications obtained in Section G. Figure 12 shows a picture of the codesigned camera. In the following, we evaluate its experimental performance. Note that for this experimental performance validation, the camera is not embedded on a real UAV but simply attached to a tripod. The PSFs of each channel of the co-designed lens are calibrated from 1 to 5 m with a step of 5 cm, with a ground truth given by a telemeter using the method of Delbracio et al.[21]. Acquisitions are made of colored textured plane scenes put at different distances from the lens. Because of the PSF variation with field angle, PSF calibration is also carried out off-axis for 9 image regions where the PSF is assumed to be constant.

Fig. 12. Co-designed 3D camera.

H.2. On axis depth estimation accuracy

For each scene and at each distance, depth is estimated with the DFD algorithm of Appendix B on image raw patches of size 46×46 pixels, from which we extract three R,G,B patches of size 23×23 pixels, inside a centred region of size 240×240 pixels, where the PSF is supposed to be constant and with a patch overlapping of 50%. Figure 13(a) to (d) show four of the scenes used in the experiment and Figure 14 shows experimental results analyses: errorbar for the target (c) of Figure 13 and RMSE for each of the targets.

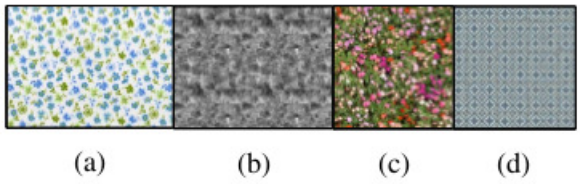


Fig. 13. Experimental targets.

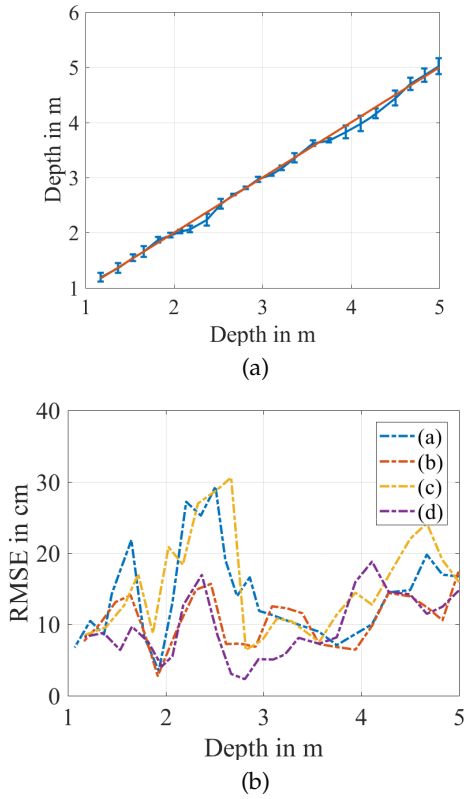


Fig. 14. Experimental depth estimation results using the CAM3D prototype. (a) Error bar for the target (d) of figure 13. (b) RMSE for the four targets.

For each scene, bias is comparable to the PSF calibration step (5 cm) and standard deviation is on the order of 10 cm. These results show that the obtained 3D camera matches our performance requirements for depth estimation in the specified depth range.

H.3. Depth map

Figure 15 shows an example of depth map obtained with our camera compared to the depth map given by the Kinect camera.

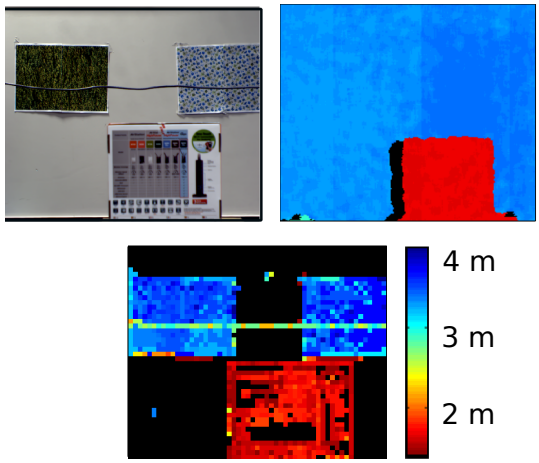


Fig. 15. From left to right: RGB image, Kinect camera and codesigned camera depth maps. Black label is for homogeneous regions.

To take into account off-axis PSF variation, the image is separated into 9 regions where the PSF is assumed to be constant. Depth estimation is then conducted on these regions with corresponding set of calibrated PSF(s). On textured regions, both 3D camera give the same depth levels. In contrast to the Kinect camera, which is an active system, we do not estimate depth on homogeneous regions (black label), because they are insensitive to defocus. On the other hand, the wire is visible in our depth map whereas it does not appear with the Kinect. This can be particularly interesting for the use-case of autonomous drone navigation.

H.4. Restored image

Figure 16 shows example of the restored image corresponding to the image of Figure 15, using a high frequency transfer. As in [7, 8], the weights of the transfer is defined with respect to the estimated distance from the depth maps. The zoom on the image allows to see the improvement of image quality due to the restoration process.

7. CONCLUSION

In this paper we have proposed a generic performance model for DFD camera leveraging from the calculation of the Cramér Rao lower bound with generic prior on image and scene. This model captures an important feature of DFD methods: the variation of the accuracy with the depth itself. The model can be used to compare the performance of DFD systems based on single or multi-image acquisition and use of conventional or unconventional optics, as far as its PSF(s) is(are) known for the depth under investigation. The proposed model has been applied on two DFD unconventional optics: coded aperture and lens with chromatic aberration. In both cases the relative performance comparison given by the theoretical model is confirmed experimentally.

Then we have used the proposed model for the end-to-end design of a chromatic camera. We used a coarse to fine approach where the camera PSF is firstly modeled using a simple Gaussian model to get a rough camera parameters estimation, and then modeled using an optical design software to conduct a fine-tuning of the parameters of the camera. The joint design is conducted using two performance models, one for depth estima-

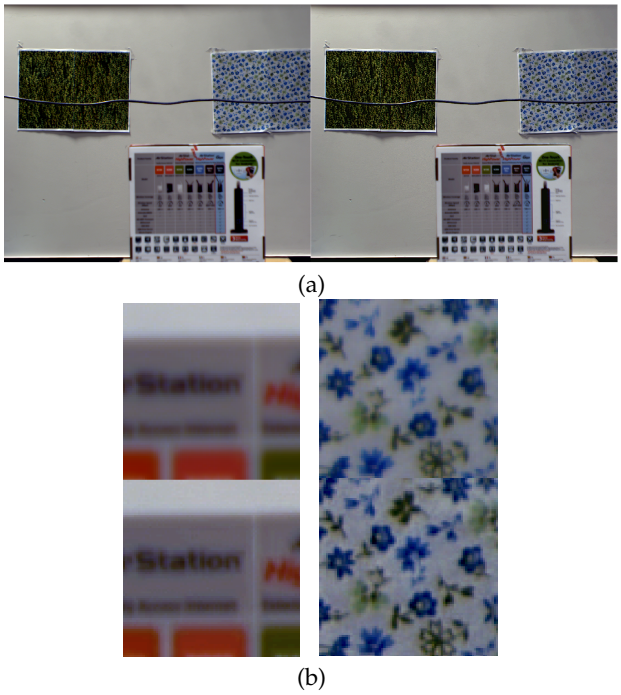


Fig. 16. Results of image restoration using a high frequency transfer between the color channels. (a) Left : original image. Right : Restored image.(b) Zooms. Top line : Original image, Bottom line : Restored image.

tion accuracy and one for image quality, requiring the choice of a trade-off in the design. A prototype of the codesigned camera has been built and its depth estimation accuracy was empirically assessed on the order of 10 cm from 1 to 5 m range, matching the requirements of the use-case of UAV navigation. Future works involve embedding such camera on a small UAV to assess the performance on real usage condition.

In this paper, image processing has been conducted using unsupervised methods for both depth estimation and image restoration. However, efficient processing based on neural networks, as in [24] for image restoration or [25] for DFD could be considered to improve the results. Prediction of the camera theoretical performances using such methods is a new challenge. End-to-end design of lens and neural networks has been recently investigated for various applications such as depth of field extension [24], HDR [26], and depth estimation [27] using unconventional optics. These works benefit from the neural network optimization framework to optimize jointly optical and processing parameters, without an explicit definition of a performance model. In future works, we intend to study the potential interaction of analytical performance models such as the one proposed in this paper with end-to-end imaging system optimization using tools from the deep learning domain.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

A. PSF SIMULATION

The proposed performance model relies on the knowledge of the PSF at any depth. The PSF can be modeled by several ways such as parametric models, Fourier optics or optical design software. We briefly present here models that are used in the paper.

A. Parametric model

In the DFD literature [1, 2, 9, 12, 13], the PSF is usually modeled using Gaussian with standard deviation $\sigma = \rho\epsilon$ with ϵ given by Equation (1). Parameter ρ is usually empirically chosen, here as in [7] we fix it at 0.3. This model is reasonable near the in-focus plane, where diffraction, defocus and sampling effect can be modeled with a Gaussian model of the blur. It is less appropriate for large defocus where geometrical optic effects are more predominant.

B. Fourier optics

When assuming a diffraction limited optical system and in the Fresnel approximation, the amplitude PSF can be modeled as the Fraunhaufer diffraction pattern of the exit pupil [28]. In the case of lens with aberrations, the complex amplitude transmittance at the exit pupil plane at the point of coordinate (u, v) reads :

$$P(u,v) = A(u,v) \exp \frac{2i\pi}{\lambda} W(u,v), \tag{17}$$

where $A(u,v)$ corresponds to the aperture shape at pupil coordinates u, v. W is the path-length error between the aberration free spherical reference wavefront and the actual wavefront. In particular when only defocus error is considered, W reads:

$$W(u,v) = \frac{1}{2}(u^2+v^2)\left(\frac{1}{s}+\frac{1}{z}-\frac{1}{f}\right), \tag{18}$$

where R and f are the camera aperture radius and focal length, respectively, s is the distance between the lens and the sensor, and z the distance between the lens and the point source, as shown in Figure 1. The optical intensity PSF is then the square of the modulus of the Fourier Transform of P. Note that other aberrations can be introduced within W. Finally this optical PSF is then be integrated over the pixel to get the intensity PSF at the sensor resolution.

C. Optical design software

Any optical design software can extract the PSF from the complete lens parameters and ray tracing. This approach requires to know all the parameters of the lenses (radius of curvature, glass type, thickness...) but it provides a finer PSF model than parametric or Fourier optics models that assume thin lens. Moreover, residual aberrations that may appear due to element misalignement or mismanufacturing can be modeled using such a software.

B. GENERIC DFD ALGORITHM

In this paper we use a DFD algorithm previously published in reference [20] for single monochrome image and in reference [7] for color image. Here, we briefly describe it using the generic formalism of Section 3 for image and scene prior. It is based on a maximum likelihood approach using the likelihood defined in Equation (4). Note that this likelihood depends on three parameters, including the depth. To reduce this number of parameters, a maximisation of the marginal likelihood over σ_N^2 can be conducted, the maximum being reached for the value $\sigma_N^2 = \boldsymbol{Y}^t P_\psi \boldsymbol{Y}/(L-m)$ (see [17], [Section 3.8.2]). Introducing this value in Equation (4) leads to a generalized likelihood that depends only on ψ :

$$p(\boldsymbol{Y};\psi) \propto |P_\psi|_+^2 (\boldsymbol{Y}^t P_\psi \boldsymbol{Y})^{-(L-m)/2}, \tag{19}$$

where N is the length of $\boldsymbol{Y}$ and m the number of zero-eigenvalues of P_ψ. Maximizing this generalized likelihood is equivalent to minimizing the function:

$$GL(\psi) = |P_\psi|_+^{-1/(L-m)} \boldsymbol{Y}^t P_\psi \boldsymbol{Y}. \quad \textbf{(20)}$$

Finally, for each patch, the DFD problem reduces to the optimization of a cost function over two parameters:

$$\widehat{k}, \widehat{\alpha} = \arg\min GL(z_k, \alpha). \quad \textbf{(21)}$$

Parameter $\alpha > 0$ fixes the inverse SNR for the considered patch. k is the index of depth within the finite set of K potential depth values z_1, ..., z_k, ..., z_K . Details on the implementation of Equation (20) can be found in reference [20].

REFERENCES

1. A. Pentland, "A new sense for depth of field," IEEE Transactions on Pattern Analysis Mach. Intell. **4** (1987).
2. M. Subbarao, "Parallel depth recovery by changing camera parameters," in *Internation Conference on Computer Vision,* (1988).
3. C. Zhou, S. Lin, and S. K. Nayar, "Coded aperture pairs for depth from defocus and defocus deblurring," Int. journal computer vision **93**, 53–72 (2011).
4. A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," ACM Trans. Graph. **26**, 70–es (2007).
5. M. Martinello and P. Favaro, "Single image blind deconvolution with higher-order texture statistics," in *Video Processing and Computational Video,* (Springer, 2011), pp. 124–151.
6. A. Chakrabarti and T. Zickler, "Depth and deblurring from a spectrally varying depth of field," in *European Conference on Computer Vision,* (2012).
7. 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. Opt. **52**, 7152–7164 (2013).
8. P. Trouvé-Peloux, J. Sabater, A. Bernard-Brunel, F. Champagnat, G. Le Besnerais, and T. Avignon, "Turning a conventional camera into a 3d camera with an add-on," Appl. optics **57**, 2553–2563 (2018).
9. A. Rajagopalan and S. Chaudhuri, "Performance analysis of maximum likelihood estimator for recovery of depth from defocused images and optimal selection of camera parameters," Int. J. Comput. Vis. **30**, 175–190 (1998).
10. I. Blayvas, R. Kimmel, and E. Rivlin, "Role of optics in the accuracy of depth-from-defocus systems," J. Opt. Soc. Am. A **24**, 967–972 (2007).
11. R. Blendowske, "Role of optics in the accuracy of depth-from-defocus systems: comment," J. Opt. Soc. Am. A **24**, 3242–3244 (2007).
12. S.-W. Shih, P.-S. Kao, and W.-S. Guo, "An error bound of relative image blur analysis," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 4 (IEEE, 2004), pp. 100–103.
13. F. Mannan and M. S. Langer, "Optimal camera parameters for depth from defocus," in *2015 International Conference on 3D Vision,* (IEEE, 2015), pp. 326–334.
14. P. Trouvé-Peloux, F. Champagnat, G. Le Besnerais, and J. Idier, "Theoretical performance model for single image depth from defocus," J. Opt. Soc. Am. A **31**, 2650–2662 (2014).
15. A. Levin, "Analyzing depth from coded aperture sets," in *Computer Vision – ECCV 2010,* K. Daniilidis, P. Maragos, and N. Paragios, eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010), pp. 214–227.
16. A. Sellent and P. Favaro, "Optimized aperture shapes for depth estimation," Pattern Recognit. Lett. **40**, 96–103 (2014).
17. J. Idier, *Bayesian approach to inverse problems* (John Wiley & Sons, 2013).
18. A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *2009 IEEE Conference on Computer Vision and Pattern Recognition,* (IEEE, 2009), pp. 1964–1971.
19. A. Chakrabarti, T. Zickler, and W. T. Freeman, "Analyzing spatially-varying blur," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* (IEEE, 2010), pp. 2512–2519.
20. P. Trouvé, F. Champagnat, G. L. Besnerais, and J. Idier, "Single image local blur identification," in *IEEE International Conference on Image Processing,* (2011).
21. M. Delbracio, P. Musé, A. Almansa, and J.-M. Morel, "The non-parametric sub-pixel local point spread function estimation is a well posed problem," International Journal of Computer Vision pp. 1–20 (2011).
22. L. Condat, "A generic variational approach for demosaicking from an arbitrary color filter array," (IEEE, 2009).
23. P. Trouvé, F. Champagnat, G. Le Besnerais, G. Druart, and J. Idier, "Design of a chromatic 3d camera with an end-to-end performance model approach," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops,* (2013).
24. S. Elmalem, R. Giryes, and E. Marom, "Learned phase coded aperture for the benefit of depth of field extension," Opt. Express **26**, 15316–15331 (2018).
25. M. Carvalho, B. Le Saux, P. Trouvé-Peloux, A. Almansa, and F. Champagnat, "Deep depth from defocus: how can defocus blur improve 3d estimation using dense neural networks?" in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops,* (2018), pp. 0–0.
26. C. A. Metzler, H. Ikoma, Y. Peng, and G. Wetzstein, "Deep optics for single-shot high-dynamic-range imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* (2020), pp. 1375–1385.
27. J. Chang and G. Wetzstein, "Deep optics for monocular depth estimation and 3d object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision,* (2019), pp. 10193–10202.
28. J. W. Goodman, *Introduction to Fourier optics* (Roberts and Company Publishers, 2005).