



A theoretical performance model for single image depth from defocus

Pauline Trouvé-Peloux, Frédéric Champagnat, Guy Le Besnerais, Jérôme Idier

► To cite this version:

Pauline Trouvé-Peloux, Frédéric Champagnat, Guy Le Besnerais, Jérôme Idier. A theoretical performance model for single image depth from defocus. *Journal of the Optical Society of America. A Optics, Image Science, and Vision*, 2014, 31 (12), pp.2650-2662. 10.1364/XX.99.099999 . hal-01082597

HAL Id: hal-01082597

<https://hal.science/hal-01082597>

Submitted on 13 Nov 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A theoretical performance model for single image depth from defocus

Pauline Trouvé-Peloux*,¹ Frédéric Champagnat,¹ Guy Le Besnerais,¹ and Jérôme Idier²

¹ONERA- The French Aerospace Lab, Chemin de la Hunière, 91123 Palaiseau, Cedex France

²IRCCyN (UMR CNRS 6597), 1 rue de la Noë, 44321 Nantes Cedex 3, France

*Corresponding author: pauline.trouve@onera.fr

compiled: October 7, 2014

In this paper we present a performance model for depth estimation using Single Image Depth From Defocus (SIDFD). Our model is based on an original expression of the Cramér Rao Bound (CRB) in this context. We show that this model is consistent with the expected behavior of SIDFD. We then study the influence on the performance of the optical parameters of a conventional camera such as the focal length, the aperture and the position of the in-focus plane. We derive an approximate analytical expression of the CRB away from the in-focus plane and we propose an interpretation of the SIDFD performance in this domain. Finally, we illustrate the predictive capacity of our performance model on experimental data comparing several settings of a consumer camera.

OCIS codes: 110.1758, 100.3190, 150.5670

<http://dx.doi.org/10.1364/XX.99.099999>

1. Introduction

This paper deals with performance model of Depth From Defocus (DFD), a passive depth estimation technique using a single lens. DFD relies on the relation between the defocus blur in the image and the object depth with respect to the camera. In early works on DFD, depth was extracted using estimation of the relative defocus blur between two or more images acquired with different optical settings [1, 2]. However, acquisition of several images requires the scene to remain still during the whole acquisition process, which is a strong hypothesis in practice, unless sophisticated imaging systems are used to separate the input beam [3, 4]. More recently, Single Image DFD (SIDFD) alternatives have been proposed [5–9]. Acquisition is then simpler, while the processing becomes more complex, since only a single image of an unknown scene is available to infer the depth map. However, several algorithms have been proposed to deal with this problem [5–9], and SIDFD can now be considered as a candidate approach for designing a RGB-D camera, *i.e.*, a camera, associated to a processing stage, which provides both a color image and a depth map.

Applications of RGB-D cameras range from robot guidance for civilian and military applications, to man-machine interfaces for game consoles or smartphones, and to 3D recording. Classical systems are passive stereoscopic systems using two or more cameras, a recent extension being plenoptic cameras [10], which can be considered as a stereoscopic system with several viewpoints but a limited baseline. Today, most commercial systems are active systems using infrared (IR) light pro-

jection, in the form of laser pulses as in LIDAR scanners and time-of-flight (TOF) cameras, or of projected light patterns such as the Kinect, developed by PrimeSense.

RGB-D cameras based on SIDFD could be used as passive systems using only one aperture compared to stereoscopic systems, and without the microlens array of plenoptic cameras. They could then provide low cost passive RGB-D solutions, as long as their performance in terms of depth estimation accuracy can be assessed. Our work is a first step in this direction. We propose a theoretical model of the performance of SIDFD, based on an original derivation of the Cramér Rao Bound, able to account for the optical parameter of a conventional camera and also for the parameters which affect the processing. Indeed, including the processing in the analysis is essential, as SIDFD puts a high requirement on this part — the counterpart for having a single lens. Note that to improve depth estimation accuracy, unconventional SIDFD camera have been designed [4–6, 8, 11, 12]. We focus here on the performance of conventional cameras. Extending our approach to unconventional DFD cameras is left for future studies.

1.A. Principle of DFD

Fig. 1 illustrates the case of a simplified camera with focal length f with the in-focus plane (IFP) put at some distance z_{IFP} . Let us consider the image of a point source referred to as the Point Spread Function (PSF). If the point source is placed in the camera IFP *i.e.*, at $z = z_{\text{IFP}}$, the corresponding PSF has a limited size imposed by diffraction. On the other hand, if the point

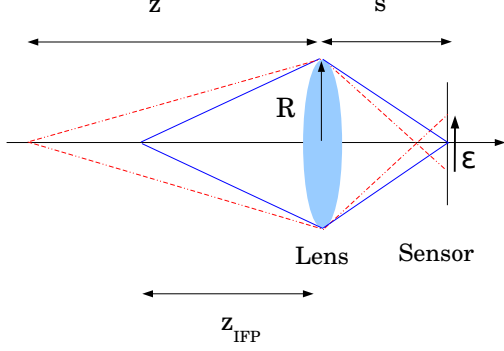


Fig. 1. Geometrical illustration of DFD for a thin lens model. z_{IFP} denotes the distance between the IFP and the lens.

source is placed out of the IFP, *i.e.*, if $z > z_{\text{IFP}}$ or $z < z_{\text{IFP}}$, the resulting PSF is enlarged because of defocus. Using geometrical relations and thin lens approximations, the defocus PSF size is:

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

where R and f are the camera aperture radius and focal length, respectively, and s is the distance between the lens and the sensor. For the developments made in this paper, it is also interesting to express ϵ as a function of the f-number $F_{/\#} = f/(2R)$ and the IFP position z_{IFP} :

$$\epsilon = \frac{f^2}{F_{/\#}} \frac{1}{z_{\text{IFP}} - f} \left(1 - \frac{z_{\text{IFP}}}{z} \right). \quad (2)$$

According to Eq. (1) and (2), for a camera having fixed focal length, aperture and focus, estimation of the defocus blur ϵ gives an estimation of the depth z . To go further, Fig. 2 shows the variations of $|\epsilon|$ with respect to depth for a conventional camera.

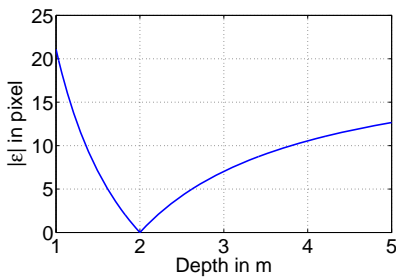


Fig. 2. Defocus blur variation with respect to depth. The parameters of the camera are: focal length $f = 25$ mm, f-number $F_{/\#} = 3$, pixel size $5 \mu\text{m}$ and $s = 25.3\text{mm}$. The IFP distance is $z_{\text{IFP}} = 2\text{m}$.

Several observations can be made from this figure. First, DFD appears feasible as there is an important variation of the blur size with respect to depth within the considered range. One can also note that the blur size $|\epsilon|$ vanishes at $z = z_{\text{IFP}}$, *i.e.*, at the IFP position. As

mentioned before, in this region, the PSF size is actually related to diffraction. In this paper, we assume that the diffracted PSF is smaller than one pixel, in other words, that the camera resolution in the IFP is limited by the sensor. There is a depth range where this assumption is valid, referred to as the *depth of field* (DoF). No blur variation can be measured in the DoF, which is then a "blind region" for DFD. Finally, in a large part of the depth range, two different depths correspond to a single blur size. This ambiguity cannot be resolved with conventional optics having a symmetric aperture. Hence, in this paper, we assume that the object position with respect to the IFP is known (in front or behind) and study the local depth estimation accuracy.

1.B. Related works

Most of the works on DFD performance concern DFD with multiple images. In [13], the performance of DFD is compared with stereoscopic approach performance. The authors show with geometrical arguments that in DFD, the aperture plays the same role as the baseline in stereo. However, as we will show in this paper, an analysis based on geometrical considerations only is not sufficient to explain SIDFD accuracy. They also analyse the optimal change of the focus settings in two images DFD, however their approach is based on a perturbation of the frequency spectrum ratio of the two images used to estimate the depth so it cannot account for the specific processing involved in SIDFD. The influence of the optical parameters in DFD with a conventional camera is studied in [14, 15]. A depth estimation accuracy formula is derived taking into account geometrical defocus blur, diffraction and pixel sampling for one in-focus and one out-of-focus image, a case that is more related to Depth from Focus (DFF) than DFD. Besides, this approach does not account for the influence of the processing on depth estimation. In [16], a Cramér Rao Bound (CRB) is derived to optimize the blur ratio between the two images to maximize the DFD accuracy. The case of SIDFD is only considered to assess the gain brought by multiple image DFD.

The present work is also based on the computation of a CRB. Indeed, the CRB is a generic tool of statistics which allows to quantify the minimal estimation error on a parameter [17]. It has been used for several 3D estimation problems, for instance in microscopy to quantify the error in particle position estimation [18, 19] or to estimate the SIDFD performance of a camera with a phase mask [20, 21]. In contrast, this paper focuses on SIDFD with a conventional camera.

Application of the CRB methodology to assess SIDFD performance is difficult because the problem has two unknowns: the scene and the depth map. In this respect, SIDFD is strongly related to blind deconvolution [22]. SIDFD algorithms are thus usually based on some assumption on the unknown scene: a scene with sharp edges is assumed in [7], while learning from a database is proposed in [8], and parametric statistical models of

the scene gradients are used in [5, 6, 9]. To study the performance, it is preferable to use a generic model of the scene, in order to broaden the range of validity of the model. Furthermore, it is also important to minimize the number of parameters allocated to the description of the scene, already knowing that SIDFD is a problem with a large number of parameters for the optical system and the processing. Here, following [9], a Gaussian model of the scene gradients is used. Thanks to an original statistical framework based on improper random vectors, we are able to derive a simple expression of the CRB, while minimizing the number of parameters of the problem. The resulting CRB can be used as an estimate of the expected accuracy of depth estimation. It depends on the parameters of the camera (optics and sensor) and of the numerical processing (signal to noise ratio (SNR), size of the patch used for local depth computation). It can be computed for all depths where the PSF is known, which allows to quantify the variation of performance over some range of use.

Finally, note that criteria related to SIDFD performance have been proposed in [6, 8] for the comparison and optimization of coded apertures. However, these criteria return a global score for an aperture shape. They do not permit to quantify the variation of performance with depth and the authors do not study the influence of other parameters than the shape of the aperture.

1.C. Outline of the paper and main contributions

In Section 2 we describe a SIDFD performance model based on geometrical considerations. We discuss its limitations, that have led us to develop a new statistical performance model. Sections 3 to 7 are dedicated to the derivation of the proposed model and to its application on a conventional camera. They contain the main contributions of the paper:

- Section 3 describes the statistical framework and derive the CRB ;
- the CRB is computed for a particular camera in Section 4, then a discussion on the SIDFD performance for different optical configurations is done in Section 5 ;
- an approximate analytical expression of the CRB for large blurs is given in Section 6: it highlights the influence of optical and processing parameters on the SIDFD accuracy ;
- Section 7 illustrates the predictive capacity of the performance model on experimental data.

We conclude and propose directions of future research in Section 8. Technical developments, in particular about improper random vectors and their use for inference, are deferred to the Appendices.

2. Geometric SIDFD performance model

In order to model depth estimation accuracy, a natural approach is to derive the sensibility of the defocus blur

size with respect to depth using geometric optical arguments. Thus we differentiate Eq. (2) with respect to the depth, which yields:

$$\frac{d\epsilon}{dz} = \frac{f^2 z_{\text{IFP}}}{F_{/\#} z^2 (z_{\text{IFP}} - f)} \simeq \frac{f^2}{F_{/\#} z^2}. \quad (3)$$

Note that here we assume $z_{\text{IFP}} \gg f$, *i.e.*, the camera is used for macroscopic and not microscopic observations. Thus the depth estimation accuracy is:

$$dz \simeq \frac{F_{/\#} z^2}{f^2} d\epsilon. \quad (4)$$

Note that this results is similar to the depth shift formula obtained in [15] using a Taylor expansion of the depth of field expression close to the IFP. Eq. (4) shows that dz depends on $d\epsilon$, which can be interpreted as the smallest blur variation that can be distinguished by processing. One can write $d\epsilon$ as a fraction of the pixel size p_x , *i.e.*, $d\epsilon = \beta p_x$. To get a first order evaluation of SIDFD accuracy, let us consider that $\beta = 1$. Fig. 3 shows the variation of the accuracy dz given by Eq. (4) for an example of camera with optical parameters given in Table 1.

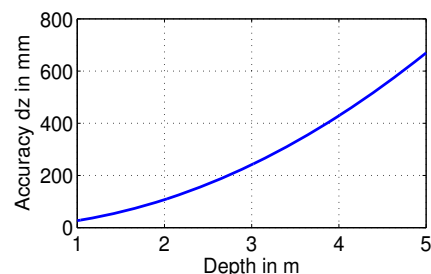


Fig. 3. Geometric accuracy of depth estimation for the imaging system with parameter values given in Table 1.

Focal length	35 mm
$F_{/\#}$	2.8
β	1
p_x	12 μm

Table 1. Optical parameters of a Nikon D200 with a 35mm focal lens, used to plot the geometric performance model of SIDFD in Fig. 3.

According to Fig. 3 and Eq. (4), depth estimation accuracy varies in z^2 , which is usual in passive ranging systems. Moreover, replacing the ratio $f/F_{/\#}$ by the aperture diameter $2R$, one gets

$$dz \simeq \frac{z^2}{2Rf} d\epsilon. \quad (5)$$

Such an expression is similar to that of the depth accuracy in stereoscopy, where $2R$ and $d\epsilon$ would play the role of the baseline and of the minimum measurable disparity, respectively. This is consistent with the performance comparison of stereoscopy and DFD conducted in [13].

Several limitations of the model given by Eq. (4) can be raised. First, accuracy monotonically improves for small depths. However, as shown in Eq. (1) and Fig. 2, when z vanishes, the blur size becomes infinite. This leads to homogeneous images where blur estimation is unlikely to be the most accurate at these depths. Besides, in the DoF, no blur variation can be observed, so it will reduce the depth estimation accuracy, which does not appear in this model. In addition, among the optical and sensor parameters, the focal length, the aperture and the pixel size influence the accuracy but not the position of the in-focus plane. However, a focus change first alters the in-focus plane position, and hence the position of the DoF, but it also modifies the defocus blur size for a given depth. Thus it should influence the accuracy of depth estimation. Finally, expression (5) only takes the optical parameters into account, and not the processing parameters related to blur estimation such as the noise level, or the number of datapoints available for the depth estimation. For all these reasons, a more precise DFD performance model is required.

3. DFD statistical performance model

The proposed performance model for SIDFD is based on a Bayesian framework to model the scene and the data. The scene model is a simple and generic Gaussian prior on the scene gradients. In this framework, we derive a data likelihood where the unknown scene is marginalized out. The latter *marginal* data likelihood is then used to calculate the CRB.

3.A. Data model

A blurred image y from a scene x is usually modelled using the relation:

$$y = h_z * x + n, \quad (6)$$

where n is the acquisition noise and h_z the PSF at depth z . As depth varies in the scene, this relation is assumed to be valid only locally. Using the vector representation on image and scene patches, we have:

$$\mathbf{y} = H_z \mathbf{x} + \mathbf{n}, \quad (7)$$

with \mathbf{y} (respectively \mathbf{x}) collects the N (resp. the M) pixels of the image (resp. scene) patch in the lexicographical order. \mathbf{n} stands for the noise process and H_z is a convolution matrix. The acquisition noise is modeled with a zero mean white Gaussian noise (WGN) of covariance $\sigma_n^2 I_N$, where I_N is the $N \times N$ identity matrix and σ_n^2 the noise variance. The issue in SIDFD is that Eq. (7) has two unknowns: the depth and the scene. Thus, some assumptions on the scene have to be made to estimate the depth.

3.B. Scene model

We choose to model the scene distribution with a Gaussian prior on the scene gradients:

$$p(\mathbf{x}|\sigma_x^2) \propto \exp\left(-\frac{\mathbf{x}^t D^t D \mathbf{x}}{2\sigma_x^2}\right), \quad (8)$$

where D is the vertical concatenation of the convolution matrices relative to horizontal and vertical first order derivative kernels $[-1 \ 1]$ and $[-1 \ 1]^t$ and parameter σ_x is a standard deviation. This simple and generic model, which can be physically interpreted as a $1/f^2$ decrease of the scene spectrum, has previously yielded good results in single image blur identification [9, 12, 22]. In Section 7, we show that it is also valid to predict the DFD performance on real scenes. Note that here the precision matrix $D^t D / \sigma_x^2$, which corresponds to the inverse of the covariance matrix, is singular, since $\mathbf{1}$ is an eigenvector associated to the eigenvalue 0. As a consequence, the covariance matrix of such a random vector is not well defined. According to the definition given in Appendix A.1, \mathbf{x} is an *improper* Gaussian random vector (IGRV). Improper models are not uncommon in statistics, the most famous example being the Brownian motion. A usual trick to ensure the existence of the covariance matrix is to use a slightly modified prior:

$$p(\mathbf{x}|\sigma_x^2) \propto \exp\left(-\frac{\mathbf{x}^t D^t D \mathbf{x}}{2\sigma_x^2} + \epsilon \|\mathbf{x}\|^2\right), \quad (9)$$

see [23] for instance. However, such a modification adds a parameter to the problem, which has to be tuned to get relevant results. Indeed, it is unnecessary here, since the next section shows that we can directly use an improper prior provided some care is taken to manipulate well defined quantities.

3.C. Marginal likelihood derivation

Derivation of a marginal likelihood from Gaussian quantities is straightforward when covariance matrices exist [24, Section 4.3.2]. Here, the scene \mathbf{x} is modelled as an IGRV and \mathbf{y} , being a filtered and noisy version of \mathbf{x} , is also an IGRV, see Theorem 1 of Appendix A.3. Using Eq. (7), it can be shown that the precision matrix of \mathbf{y} is:

$$Q_\theta = \frac{1}{\sigma_n^2} P_\psi, \quad (10)$$

$$P_\psi = I - H_z (H_z^t H_z + \alpha D^t D)^{-1} H_z^t, \quad (11)$$

where $\alpha = \sigma_n^2 / \sigma_x^2$ can be interpreted as the inverse of a signal to noise ratio (referred to as inverse SNR), $\theta = \{\sigma_n, \psi\}$, and $\psi = \{\alpha, z\}$. Note that matrix $H_z^t H_z + \alpha D^t D$ in (11) is full rank since H_z and $D^t D$ respectively correspond to a band pass and a high pass operator, so that their null spaces have an empty intersection. In appendix A.2, we prove that a likelihood can be derived in the case of an IGRV, according to Proposition 1 below.

Proposition 1 *Let \mathbf{y} be an IGRV with a precision matrix Q_θ , and let us assume that the kernel of Q_θ does not depend on θ . Then, inference about θ can be conducted using the likelihood function:*

$$p(\mathbf{y}|\theta) = \left| \frac{Q_\theta}{2\pi} \right|_+^{\frac{1}{2}} \exp\left(-\frac{1}{2} \mathbf{y}^t Q_\theta \mathbf{y}\right), \quad (12)$$

where $|A|_+$ refers to the product of the non zero eigenvalues of a symmetric matrix A , with the eigenvalues raised to the power of their multiplicity.

To apply this result, we have to study the kernel of Q_θ given by Eqs. (10) and (11). According to Corollary 1 in appendix A.3, one has:

$$\text{Ker}Q_\theta = H_z \text{Ker}D^t D = H_z \text{Ker}D = H_z \mathbf{1}, \quad (13)$$

and $H_z \mathbf{1} \propto \mathbf{1}$ because H_z is a convolution matrix. Thus, $\text{Ker}Q_\theta$ is indeed independent of θ and the likelihood $p(\mathbf{y}|\theta)$ is given by (12).

3.D. Cramér Rao Bound for depth estimation

As shown in Appendix A.4, the Fisher information relative to the estimation of parameter θ from a sample of an IGRV having a precision matrix Q_θ is

$$\text{FI}(\theta) = \frac{1}{2} \text{tr} \left(Q_\theta + \frac{dQ_\theta}{d\theta} Q_\theta + \frac{dQ_\theta}{d\theta} \right), \quad (14)$$

where $+$ denotes the pseudo inverse. Here we assume that the only variable parameter is depth z so that $\theta = \{z\}$. Thus we use the scalar information:

$$\text{FI}(z) = \frac{1}{2} \text{tr} \left(P_z + \frac{dP_z}{dz} P_z + \frac{dP_z}{dz} \right). \quad (15)$$

The CRB, defined as the inverse of the FI,

$$\sigma_{\text{CRB}}^2(z) = \text{FI}^{-1} = 2 \left(\text{tr} \left(P_z + \frac{dP_z}{dz} P_z + \frac{dP_z}{dz} \right) \right)^{-1}, \quad (16)$$

is the lower bound of the variances of all possible unbiased estimates of depth [17]. $\sigma_{\text{CRB}}(z)$ is homogeneous to a depth and can be considered as the accuracy of depth estimation by SIDFD at depth z .

3.E. Computation of the performance model

Not surprisingly, according to Eq. (11) and (16) the SIDFD performance based on CRB depends on the knowledge of the PSF h_z at any depths z . A camera PSF depends on all relevant characteristics of the system such as diffraction, defocus, optical aberrations and sensor integration. Given all these deformations, several papers assume that the shape of the PSF is Gaussian [1, 2, 16]. In [14], it is modeled as the convolution of a Bessel, a pill-box and a square function to model respectively the diffraction, defocus and pixel integration. A finer PSF model including optical aberrations can be computed using Fourier optics principles [25], or, if the lens settings are known, be derived from an optical design software such as **Zemax**. Finally, to get the PSF as close as possible to the actual camera PSF one can calibrate the PSF using high frequency patterns [6, 26].

Given the PSF h_z , the convolution matrix H_z is built for a given patch size and matrix P_z is then computed using Eq. (11) for some value of the inverse SNR α . The

derivative $\frac{dP_z}{dz}$ in Eq. (16) can be approximated with several numerical approaches, here we use the centred finite difference:

$$\frac{dP_z}{dz} \simeq \frac{P_{z+\delta} - P_{z-\delta}}{2\delta}, \quad (17)$$

where δ is a small depth variation with respect to z . Computing $\sigma_{\text{CRB}}(z)$ hence requires the four matrices $\{P_z, P_{z-\delta}, P_{z+\delta}\}$, and thus the knowledge of the PSF at $z - \delta, z$ and $z + \delta$.

4. Single image DFD performance: an example

The proposed performance model is used for a conventional camera with optical parameter values given in Table 1. Such values correspond to a Nikon D200 camera with a 35 mm focal lens, used in the experiments in Section 7. As in several previous references [1, 2, 16], the PSF h_z is simply modelled as a 2D Gaussian function:

$$h_z(u, v) = \frac{1}{2\pi\tau^2(z)} \exp \left(-\frac{u^2 + v^2}{2\tau^2(z)} \right), \quad (18)$$

where the standard deviation $\tau(z)$ is given by $\tau(z) = \rho\epsilon(z)/p_x$, with $\rho = 0.3$, a parameter tuned so that the widths of the simulated PSFs match those obtained using Fourier Optics formula [25] and ϵ given by Eq. (1) or (2). Fig. 4(b) shows the variation of σ_{CRB} with respect to depth for an IFP at $z_{\text{IFP}} = 1.8\text{m}$, with a patch size of 21×21 pixels, $\delta = 1\text{ mm}$ and using two values of the inverse SNR $\alpha = 0.01$ and 0.001 . Fig. 4(a) shows the PSF size variation with respect to depth. The geometric model of Eq. (5) is also plotted in Fig. 4(b) for comparison and a zoom close to the IFP is plotted in Fig. 4(c). The proposed performance model overcomes all the limitations of the geometric model, which have been listed in Section 2. First, it presents a divergence around the IFP, in the DoF where no PSF variation is measurable. Second, it predicts the performance degradation for small depths due to the increasing of the PSF size. Third, it accounts for the noise level, which, as expected, has a major influence on the performance of SIDFD. Fig. 4(c) shows that increasing the SNR slightly improves the best performance and broadens the region of high accuracy, but its main effect is to change the increasing rate of σ_{CRB} when going away from the IFP.

Finally, whereas for the geometric model the best performance is obtained for the smallest depths, the proposed model predicts that the best accuracy on the estimated depth is obtained just after the DoF, *i.e.*, for depths 1.7m and 2.0m in the case illustrated in Fig. 4(b). In other words, the best accuracy is obtained in the regions where the imaging system is slightly out of focus. Such a result is consistent with a study of [18] on localization of particles, where the authors concluded that the best accuracy was obtained from slightly out-of-focus acquisitions.

5. Performance study

In the following we study the impact of tuning two basic camera parameters : the position of the IFP and the aperture size.

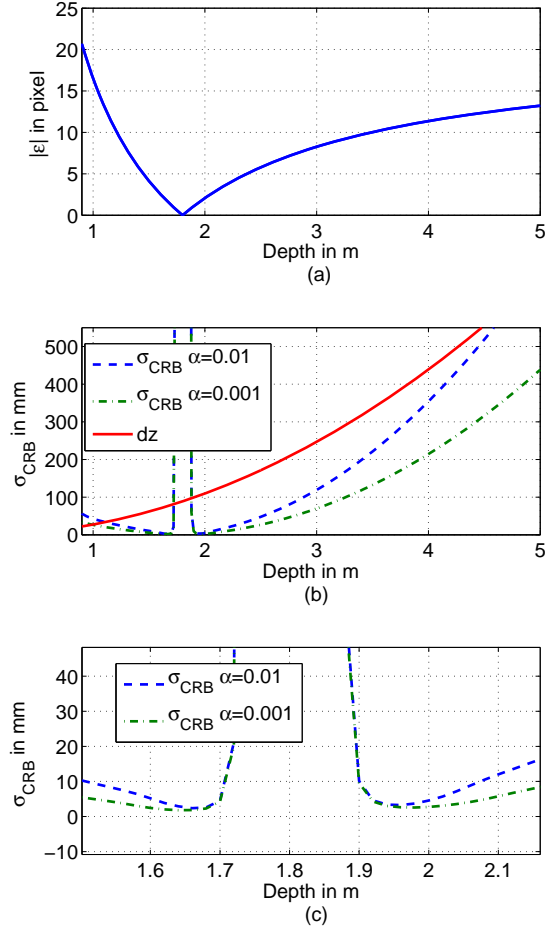


Fig. 4. (a) Geometrical blur size with respect to depth for the imaging system whose parameters are given in Table 1, with the in-focus plane at 1.8m. (b) Corresponding variation of geometrical performance model dz and of σ_{CRB} for two values of the inverse SNR α , with respect to depth. (c) Zoom of the figure (b) close to the IFP.

5.A. Variation of the in-focus plane

We consider again a camera with parameters given by Table 1 and modify the position of the IFP by changing the sensor position s with respect to the lens, all other optical parameters remaining fixed. Two configurations are compared in Fig. 5(a) to (c), the first one with an IFP at $z_{IFP} = 1.5$ m, the second at $z_{IFP} = 1.8$ m. Fig. 5(a) and (b) respectively show the variation of the geometric blur size and the SIDFD performance for both configurations. Fig. 5(c) is a zoom of Fig. 5 (b) close to the IFP.

As expected, both configurations lead to a divergence around the IFP. Note that the blind region (*i.e.*, the region where SIDFD diverges) is larger for $z_{IFP} = 1.8$ m, because the DoF increases with the distance of the IFP from the camera. Fig. 5(b) shows that, in contrast with the conclusions derived from the geometrical model, the SIDFD performance for a given depth highly depends on the IFP position. Let us browse the curves of Fig. 5(b)

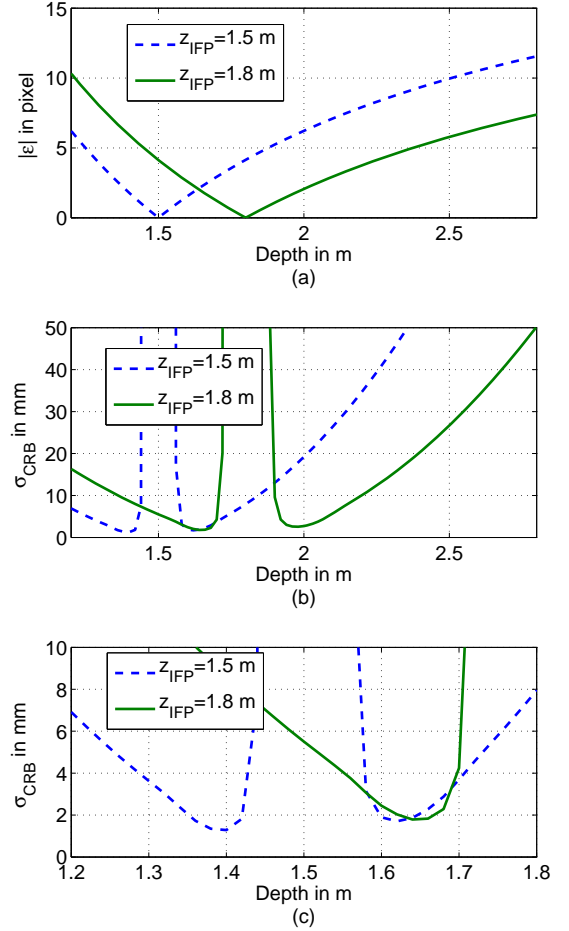


Fig. 5. Influence of the position of the IFP on the SIDFD performance. Geometrical blur size (a) and σ_{CRB} variation (b) with respect to depth for a camera with parameters given in Table 1, for $z_{IFP} = 1.5$ m and $z_{IFP} = 1.8$ m. (c) Zoom of figure (b) around the IFP. The inverse SNR is fixed at $\alpha = 0.001$.

according to increasing depths: for $z < 1.4$ m the best performance is obtained with the IFP at 1.5m; between 1.45 and 1.6m the IFP at 1.5m leads to a divergence of σ_{CRB} and the best choice is $z_{IFP} = 1.8$ m; both choices lead to similar performance around $z = 1.65$ m; IFP at 1.8m diverges between 1.7m and 1.9m; and finally, for $z > 1.9$ m, $z_{IFP} = 1.8$ m yields the best performance.

Two points must be highlighted, as we will see in the following that they are general properties of SIDFD. First, far from the IFP, *i.e.*, for the regions $z < 1.4$ m and $z > 1.8$ m, the best performance is obtained with the IFP closest to the considered depth. By comparing Fig. 5 (b) and (a) it can be noted that, in these regions, the best performance is then obtained with the setting having the sharpest PSF. Second, looking more closely at the region between the two IFP, we note that depth $z = 1.64$ m where the two settings give the same performance is exactly the depth where they lead to the same size of blur (2 pixels).

These facts indicate that SIDFD performance is not only related to depth, as implied by the geometrical model, but also to the blur size.

5.B. Changing the aperture

We now consider the influence of the aperture in Fig. 6, all other parameters being fixed. In particular the inverse SNR α is fixed which implies that the integration time is adapted to aperture size in order to keep a constant input flux. Note that aperture is quite naturally related to SIDFD performance. Indeed, following the geometrical model, one could intuitively expect that diminishing the aperture would lead to a lower SIDFD performance, as it lowers the rate of PSF's variation with depth and also enlarges the DoF, see Fig. 6(a). However,

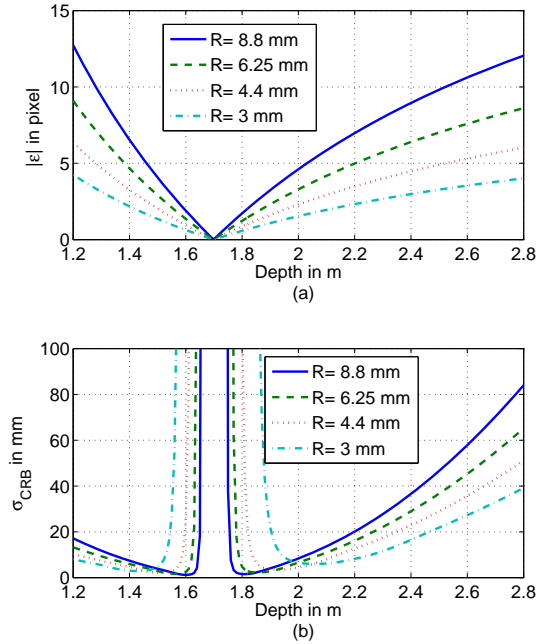


Fig. 6. (a) and (b): geometrical blur size and σ_{CRB} variation with respect to depth, respectively, for a camera with parameters given in Table 1, with various aperture radii. ϵ is calculated using Eq. (1) and inverse SNR $\alpha = 0.001$.

Fig. 6 (b) shows that this reasoning is only true in the vicinity of the DoF, *i.e.*, in the region $1.5\text{m} < z < 2\text{m}$ in the proposed example. Outside this region, the order of the curves is reversed: the smaller the aperture, the better the performance. Once again, we can remark that, far from the DoF, the tuning which leads to the best performance is the one which is associated to the smallest PSF. The next section analyses more deeply this relation between performance and blur size.

6. Asymptotic approximation of the performance for large blurs

In this Section we show that the observations made in the previous section about the relationship between performance and blur size are actually general properties

of SIDFD far from the DoF, in the asymptotic domain where the blur size is large $|\tau(z)| > 1$ and $\alpha < 1$. Indeed, in this domain, an analytical formula for SIDFD accuracy can be obtained under the assumption of a Gaussian PSF and by approximating the convolution operators H and D by circular matrices. The calculations, detailed in Appendix B, lead to a product of two terms:

$$\sigma_{\text{BCR}}(z) \simeq \eta(z)\beta(N, \tau(z), \alpha). \quad (19)$$

The first term, $\eta(z) = \frac{z^2 F/\#}{f^2}$, is the inverse of the defocus blur variation with respect to depth $d\epsilon/dz$ of Eq. (4). It describes the impact of the geometrical configuration of the system on the performance. As already discussed, this term is independent of the position of the IFP and favours a high rate of variation of the defocus blur with respect to depth. The second term writes:

$$\beta(N, \tau(z), \alpha) = \kappa p_x \frac{\tau^2(z)}{\sqrt{N} \rho \left(\ln \frac{\tau^2(z)}{\alpha} - \ln \ln \frac{\tau^2(z)}{\alpha} \right)^{3/2}} \quad (20)$$

with $\kappa = \sqrt{6\pi}$ and $\tau(z) = \rho\epsilon/p_x$. This term accounts for the processing part of the SIDFD system, *i.e.*, the accuracy that can be expected when identifying the blur level by processing a patch of the image. Not surprisingly then, it is an increasing function of the inverse SNR α and varies in $1/\sqrt{(N)}$, where N is the number of pixels in the patch. The dependency of β with respect to the blur size $\tau(z)$ is less easy to infer. It can be shown that the function

$$\gamma(x) = \frac{x}{(\ln x - \ln \ln x)^{3/2}}$$

is positive and monotonically increasing when $x > 1$. As a consequence, in the domain considered here ($|\tau(z)| > 1$ and $\alpha < 1$), β is a monotonically increasing function of $|\tau(z)|$: some examples are plotted in Figure 7. This term then favours small blur sizes to improve performance. Indeed, in practice large blur sizes lead to homogeneous images for which identification of the PSF, or estimation of depth, is inaccurate. Thus the performance of a conventional SIDFD camera results from two factors having contradictory behaviour with respect to the size of blur: η favours configurations having large PSFs, while β leads to select small PSF.

In Section 5, we have numerically observed the impact on σ_{CRB} of tuning two basic camera parameters: position of the IFP and aperture size. It is now interesting to study again the effect of these two parameters to interpret the observations made in Section 5 with the proposed analytical formula.

We now study again the consequences of tuning only the IFP, all other parameters being fixed. As already mentioned, the geometrical term does not depend on the IFP. The variation of performance when tuning z_{IFP} is then completely described by the behavior of β . Actually, the term β depends on the IFP only through

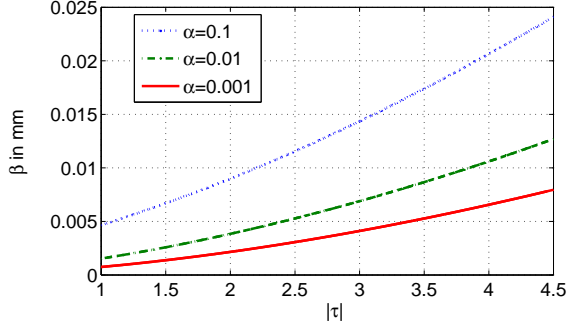


Fig. 7. Variation of β given in Eq. (20) with respect to τ_z , for three values of inverse SNR α , for a camera with parameters given in Table 1.

$\tau(z)$, the relation between both parameters being given by Eq. (2). This leads to two consequences which have been already observed in Section 5.A. First, two configurations with identical blur size have identical performance. Second, far from the DoF, the best performance is obtained with the configuration having the smallest PSF. This lead to put the IFP as close as possible to the range of depth of interest.

Let us now tune the f-number $F/\#$, leaving all other parameters unchanged. Increasing $F/\#$ increases η , but reduces the size of the PSF, hence lowers β . In this situation, the contradiction between both terms can be resolved by rewriting Eq. (19) using (2) as follows:

$$\sigma_{\text{BCR}} \simeq \kappa \frac{z(z - z_{\text{IFP}})}{z_{\text{IFP}} - f} \frac{\tau(z)}{\sqrt{N}(\ln \frac{\tau^2(z)}{\alpha} - \ln \ln \frac{\tau^2(z)}{\alpha})^{3/2}}. \quad (21)$$

The f-number $F/\#$ only affects the fraction on the right part, through $\tau(z)$. As this term is increasing in $\tau(z)$, the best accuracy is obtained for the smallest blur size, and hence for the smallest aperture. This result is consistent with the conclusion of Section 5.B for a depth range far from the IFP.

To summarize, our study in the asymptotic domain $|\tau(z)| > 1$ shows that far from the IFP, the accuracy of SIDFD is governed by the performance of the processing, which depends mainly on the blur size. This explains why, in this domain, the best performance is always obtained with the parameters giving the smallest blur.

7. Experimental validation of the model

So far, our results are theoretical, based on various expressions of the accuracy $\sigma_{\text{CRB}}(z)$ derived from models of the scene and the camera. Our aim in this Section is to present an experimental validation of our theoretical performance model, through empirical evaluation of SIDFD on real textured scenes acquired with a real camera. However, trying to reach experimentally the theoretical performance levels is a very difficult task which requires a carefully controlled experiment. Our objective

here is only to evaluate experimentally the *relative* performance between different configurations and demonstrate that the experiment matches the theoretical prediction, in the sense that the best configuration in the theoretical study leads to the best empirical results.

We first present the SIDFD algorithm used to estimate the depth, which is an extension of the blur identification method proposed in [9]. Then we describe our experimentations. As in Sections 5 and 6, they essentially consist in changing the IFP and the aperture and compare the SIDFD performance, this time empirically.

7.A. SIDFD estimation

7.A.1. Principle

PSF identification following Ref. [9] is conducted over patches of the image where depth is assumed constant. The estimated depth z is selected among a finite set of potential depth values $\{z_1, \dots, z_k, \dots, z_K\}$, each one being related to a unique PSF thanks to a calibration process. The selection criterion is derived within a maximum likelihood framework, as described in the following Section.

7.A.2. Criterion for depth estimation

Using Proposition 1 and the notations of Sec. 3.C the log-likelihood of the data can be written as:

$$\ln p(\mathbf{y}|\psi, \sigma_n^2) = \frac{1}{2} \left(\ln |P_\psi|_+ - (N-1) \ln(2\pi\sigma_n^2) - \frac{\mathbf{y}^T P_\psi \mathbf{y}}{\sigma_n^2} \right), \quad (22)$$

where the notation $\psi = \{z, \alpha\}$ is used for convenience. This expression can be simply maximized over σ_n^2 , the maximum being reached for the value $\hat{\sigma}_n^2 = \mathbf{y}^T P_\psi \mathbf{y} / (N-1)$. Introducing this value in (22) leads to a generalized likelihood that depends only on z and α :

$$\ln p(\mathbf{y}|\psi, \hat{\sigma}_n^2) = \frac{1}{2} \left(\ln |P_\psi|_+ - (N-1) \ln(\mathbf{y}^T P_\psi \mathbf{y}) + b \right), \quad (23)$$

where b is a constant. Maximizing this likelihood is equivalent to minimizing the function:

$$GL(\psi) = GL(z, \alpha) = |P_\psi|_+^{1/(N-1)} \mathbf{y}^T P_\psi \mathbf{y}. \quad (24)$$

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

$$\hat{k}, \hat{\alpha} = \arg \min_{k, \alpha} GL(z_k, \alpha). \quad (25)$$

Parameter $\alpha > 0$ fixes the inverse SNR for the considered patch. k is the index of depth within the finite set of potential depth values $\{z_1, \dots, z_k, \dots, z_K\}$. Details on the implementation of the algorithm are given in [9].

7.B. Experimental evaluation of SIDFD

Experimentations are conducted with D200 Nikon camera with a standard 35 mm lens whose main parameters

are given in Table 1. This camera is used with different IFP positions and different apertures. For each setting, the PSFs are calibrated using the PSF estimation method of [26], using a calibration pattern located in several potential depths behind the IFP, with a step of 5 cm between each position. Depth estimation perfor-

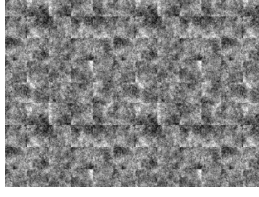


Fig. 8. Texture scene acquired at different depths for experimental DFD accuracy evaluation.

mance by SIDFD is evaluated on a real planar and textured scene shown in Fig. 8. This scene has the spectral content corresponding to the scene model of Eq. (8). It is positioned in a fronto-parallel configuration at various depths in the range defined by the calibration step. For each acquisition, a region of size 200×200 pixels around the image center is extracted and cut into 80 non overlapping 21×21 patches. For each patch, depth is estimated by solving Eq. (25) on the pixels corresponding to the green raw channel. Standard deviation (Std) and bias are evaluated with respect to a ground truth given by a telemeter.

7.B.1. Variation of the IFP

The focus of the camera is manually set at $z_{\text{IFP}} = 1.5\text{m}$ then at $z_{\text{IFP}} = 1.8\text{m}$. In both cases, the f-number is 2.8 and a fixed integration time is used. Empirical performance of SIDFD is represented in Fig. 9.

Fig. 9 shows the bias and the standard deviation. The former is limited with respect to the latter, so we focus on the standard deviation. The standard deviation obtained with the IFP at $z_{\text{IFP}} = 1.8\text{m}$ is significantly lower than with the IFP at $z_{\text{IFP}} = 1.5$ over the whole range of depths (2.1m to 2.5m) which has been evaluated. This result is consistent with our theoretical study, in particular with the conclusions of Section 5.A. Indeed, as predicted by the proposed performance model, depth estimation accuracy varies with the position of the IFP and favors the IFP located closest to the range where depth is estimated.

7.B.2. Aperture size variation

Two f-numbers of 4 and 2.8 (corresponding respectively to aperture radius of 4.4 mm and 6.25 mm) are selected manually on the camera. The depth range is 1.7 to 2.1 m with the IFP put at $z_{\text{IFP}} = 1.5\text{m}$. Here the integration time is adapted so that the input flux is the same for the two apertures, in order to modify only the size of the PSF and not the value of the inverse SNR α .

Fig. 10 shows that both configurations have a limited bias below the calibration step, while they have significantly different standard deviations: the lowest aperture

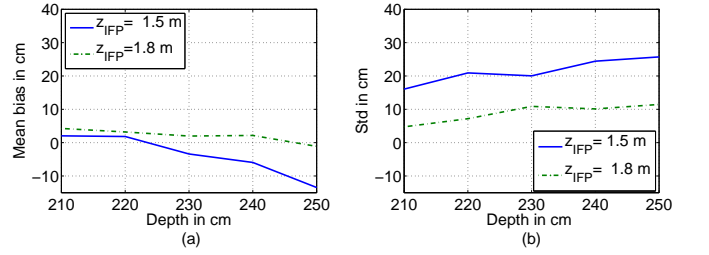


Fig. 9. Experimental performance of SIDFD for two different IFP positions as a function of depth. Left: bias in cm, right: standard deviation in cm.

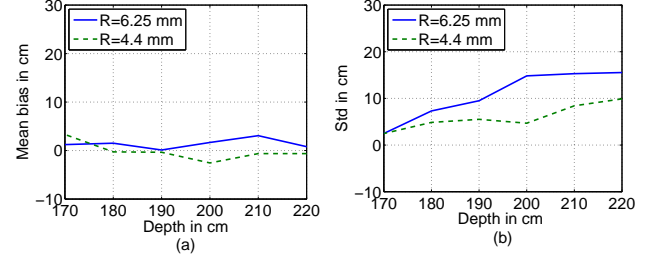


Fig. 10. Depth estimation experimental mean error (a) and standard deviation (b) obtained for two different aperture radii, with the same IFP position.

radius leads to the best accuracy. This experimental observation is consistent with our theoretical performance study, in particular Fig. 6 which predicts that, far from the IFP, the best configuration is the one with the smallest aperture radius *i.e.*, the highest f-number.

8. Conclusion

We have presented a new performance model for estimation of depth from defocus using a single image (SIDFD). It is essentially a CRB derived within a Bayesian statistical framework using a generic model of the scene and can be computed for every depth in the range where the PSF of the camera is known.

Several results about SIDFD can be drawn from this model. For SIDFD with a conventional camera, one can distinguish three behaviors in the performance curve depending on the considered range. The depth of field, *i.e.*, the region close to the IFP where the PSF size is below the pixel size, is a blind region where depth is not measurable. Far from the IFP, the depth estimation accuracy is limited by the performance of the processing and governed mainly by the size of the PSF: the smaller the PSF, the better the performance. Finally, the best performance is obtained near the DoF where the camera is slightly out of focus. In these regions, one has the best resolution in depth, but this high accuracy is restricted to a very narrow range of depths.

As it accounts for both camera settings (focal length, f-number, focus) and processing parameters (SNR, patch size) the proposed model can be used for the joint design of a SIDFD system *i.e.*, the tuning of optical and

processing parameters to optimize the performance in a given range. In the case of a conventional camera used in a range of depths located beyond the DoF, we have studied the influence of two parameters: the focus and the aperture. Our model predicts that the best performance will be reached by moving the IFP closer to the range of interest (in the limit of letting this range out of focus) and by reducing the f-number. These predictions have been confirmed by an empirical evaluation conducted on real images using a consumer camera. It should be emphasized that these conclusions stem from the study of the dependence of the SIDFD performance with respect to the size of the PSF, which is an original feature of our work.

An interesting perspective is to apply this model to the design of SIDFD systems with unconventional cameras, for instance cameras having coded apertures, or a chromatic lens [6, 11, 12].

Appendix A: Fisher Information using improper Gaussian prior

Inference using finite dimensional IGRVs is not much documented in the statistical literature, a notable exception is Besag and Kooperberg's work [27]. However, the latter addresses the problem of improper precision matrix estimation from second order increment statistics, and not the issue of generalizing the likelihood function. Definition of a likelihood with improper prior are given in [28, 29] but their justifications applies only to their application that is the estimation of the regularization parameter. We propose in the following a generalization to more general parameters estimation, in particular to parameters related to the PSF.

A.1. Definition and properties of IGRVs

One difficulty of dealing with IGRVs is that classical results about covariance propagation with linear transforms cannot be used because IGRVs have no covariance, indeed they do not have a probability density function, because they do not live in a probability space, but rather in an unbounded measure space [30]. This leads to the following definition of an IGRV.

Definition 1 *The random vector \mathbf{X} is an improper Gaussian random vector (IGRV) with precision matrix Q if its measure has a density of the form $p_{\mathbf{X}}(\mathbf{x}) \propto \exp(-\frac{1}{2}\mathbf{x}^t Q \mathbf{x})$, with Q a non negative definite (NND) singular matrix.*

Inference on an IGRV requires the use of several lemmas and propositions that are introduced hereafter. The following lemma deals with the propagation of precision matrices with respect to linear transforms.

Lemma 1 *Let \mathbf{X} be an IGRV with a $M \times M$ precision matrix Q , and A be a regular matrix. Then $\mathbf{Y} = A\mathbf{X}$ is an IGRV with precision matrix $A^{-t}QA^{-1}$.*

Proof Let B be any measurable set on \mathbb{R}^M , by a change of variable $\mathbf{y} = A\mathbf{x}$ under the integral we have:

$$\begin{aligned} \int_B \exp\left(-\frac{1}{2}\mathbf{x}^t Q \mathbf{x}\right) d\mathbf{x} \\ = \int_{A(B)} \exp\left(-\frac{1}{2}\mathbf{y}^t A^{-t}QA^{-1}\mathbf{y}\right) \frac{d\mathbf{y}}{|A|} = \infty. \end{aligned} \quad (26)$$

This means that \mathbf{Y} is also an IGRV, with $A^{-t}QA^{-1}$ as precision matrix.

The following lemma relies on Lemma 1. It shows that an IGRV can be split into two parts, a purely proper one and purely improper one. In other words, the space of IGRVs can be considered as the set of Gaussian random vectors (RVs) with regular covariance plus the limit of the latter when covariance becomes infinite. Conversely, deterministic behaviour (*i.e.*, singular covariances) are not allowed in IGRV spaces.

Lemma 2 *Let \mathbf{Y} be an IGRV of size N with precision matrix Q and n be the rank of Q , let A_p (resp. A_i) be any $n \times N$ (resp. $(N - n) \times N$) matrix whose rows span the range (resp. the kernel) of Q . Then $\mathbf{Y}_i = A_i\mathbf{Y}$ and $\mathbf{Y}_p = A_p\mathbf{Y}$ are independent RVs such that \mathbf{Y}_i is a Lebesgue measure and \mathbf{Y}_p is a Gaussian RV with regular precision matrix $(A_p^+)^t Q A_p^+$ ($^+$ denotes the pseudo inverse).*

Proof Let $A = (A_p^t A_i^t)^t$. Since Q is NND, $A_p^t A_i = 0$, so $A^{-1} = (A_p^+ A_i^+)^t$. Using Lemma 1 for $(\mathbf{Y}_p^t \mathbf{Y}_i^t)^t = A\mathbf{Y}$:

$$p_{\mathbf{Y}_p, \mathbf{Y}_i}(\mathbf{y}_p, \mathbf{y}_i) \propto \frac{1}{|A|} \exp\left(-\frac{1}{2}\mathbf{y}_p^t (A_p^+)^t Q A_p^+ \mathbf{y}_p\right).$$

Such a joint density is separable with respect to each variable (since it is a constant function of \mathbf{y}_p), so \mathbf{Y}_p and \mathbf{Y}_i are independent. Finally $(A_p^+)^t Q A_p^+$ is regular because A_p spans the range of Q .

A.2. Likelihood of an improper Gaussian distribution

We are now ready to derive a likelihood for an IGRV with structured precision matrix parametrized by a parameter θ , written Q_θ . Clearly, Lemma 2 offers a simple solution: project the data on the range of Q_θ to yield a proper RV. This solution is valid only if the projection does not depend on θ . The latter remark leads to the critical requirement in Proposition 1 that the range of Q_θ (or equivalently the kernel of Q_θ) be independent on θ :

Proposition 1 *Let \mathbf{Y} be an IGRV with structured precision matrix Q_θ , and assume that the kernel of Q_θ does not depend on θ . Then, inference about θ can be conducted with the likelihood:*

$$L(\theta|\mathbf{y}) = \left| \frac{Q_\theta}{2\pi} \right|_+^{\frac{1}{2}} \exp\left(-\frac{1}{2}\mathbf{x}^t Q_\theta \mathbf{x}\right), \quad (27)$$

where $|\cdot|_+$ is the product of the positive eigenvalues of a matrix, with the eigenvalues raised to the power of their multiplicity.

Proof The proof has two parts. The first one provides a first likelihood expression (28), which is further shown to admit the simpler form (27) in the second part.

First part. Since Q_θ is NND, the range of Q_θ is orthogonal to its kernel. Thus, since the kernel of Q_θ does not depend on θ , its range is also independent on θ . Let A_p (resp. A_i) be any matrix whose rows span the range (resp. the kernel) of Q_θ , so that A_p (resp. A_i) is independent on θ . Using Lemma 2, we define a one-to-one linear mapping between \mathbf{Y} and $[\mathbf{Y}_p^t \mathbf{Y}_i^t]^t$, with \mathbf{Y}_p and \mathbf{Y}_i two independent RVs, \mathbf{Y}_i is a Lebesgue measure, independent on θ , and \mathbf{Y}_p is a Gaussian RV with regular precision matrix $(A_p^+)^t Q_\theta A_p^+$. Therefore, all information about θ lies in \mathbf{Y}_p : according to the statistical terminology, \mathbf{Y}_p plays the role of a *sufficient statistics*. So the likelihood of θ given the sample \mathbf{y} is defined as the likelihood for a Gaussian RV given the sample $\mathbf{y}_p = A_p \mathbf{y}$:

$$\begin{aligned} L(\theta|\mathbf{y}) &= p_{\mathbf{Y}_p}(A_p \mathbf{y}) \\ &= \left| \frac{(A_p^+)^t Q_\theta A_p^+}{2\pi} \right|^{\frac{1}{2}} \exp \left(-\frac{1}{2} \mathbf{y}^t A_p^t (A_p^+)^t Q_\theta A_p^+ A_p \mathbf{y} \right). \end{aligned} \quad (28)$$

The term inside the exponential simplifies to

$$\mathbf{y}^t A_p^t (A_p^+)^t Q_\theta A_p^+ A_p \mathbf{y} = \mathbf{y}^t Q_\theta \mathbf{y}, \quad (29)$$

because $A_p^+ A_p$ is an orthogonal projection onto the range of Q_θ . But even so, Eq. (28) is still unpractical due to the determinant: it depends on a particular choice of a basis of the range of Q_θ .

Second part. One can select A_p such that its rows define an orthonormal basis of the range of Q_θ , and we show hereafter that

$$|A_p^+ Q_\theta A_p^+| = |Q_\theta|_+. \quad (30)$$

First, because of orthonormality, $A_p^+ = A_p^t$, so that we have to show that $|A_p Q_\theta A_p^t| = |Q_\theta|_+$. Let $Q_\theta = B_\theta^t \Lambda_\theta B_\theta$ denotes an eigenvalue/eigenvector decomposition including only non zero eigenvalues. Then,

$$A_p Q_\theta A_p^t = (A_p B_\theta^t) \Lambda_\theta (A_p B_\theta^t)^t. \quad (31)$$

Using the following lemma 3, $A_p B_\theta^t$ is an orthonormal matrix, therefore $|A_p B_\theta^t|^2 = 1$. It follows that $|A_p Q_\theta A_p^t| = |\Lambda_\theta| = |Q_\theta|_+$. Putting together (28), (29) and (30) yields (27).

Lemma 3 *Let A and B be two $n \times N$ matrices such that $AA^t = I$ and $BB^t = I$ and $\text{Ker } A = \text{Ker } B$. Then $AB^t BA^t = I$.*

Proof The rows of A define an orthonormal set. One can complete this set with $N - n$ orthonormal vectors to form an orthonormal basis. Let C be the $(N - n) \times N$ matrix of these vectors. Then $\tilde{A} \triangleq [A^t C^t]^t$ is a $N \times N$ unitary matrix. $\tilde{B} \triangleq [B^t C^t]^t$ is also a

$N \times N$ unitary matrix because $BB^t = I$ and $BC^t = 0$ ($\text{Ker}(A) = \text{Ker}(B)$). Let us expand $\tilde{A}\tilde{B}^t$:

$$\tilde{A}\tilde{B}^t = \begin{bmatrix} AB^t & AC^t \\ CB^t & CC^t \end{bmatrix} = \begin{bmatrix} AB^t & 0 \\ 0 & I \end{bmatrix} \quad (32)$$

As the product of unitary matrices $\tilde{A}\tilde{B}^t$ is unitary:

$$\tilde{A}\tilde{B}^t \tilde{B}\tilde{A}^t = \begin{bmatrix} AB^t BA^t & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \quad (33)$$

So $AB^t BA^t = I$ and thus AB^t is an orthonormal matrix.

A.3. Precision matrix of a filtered and noisy IGRV

In the present paper, the data are a filtered and noisy version of an IGRV. The following theorem shows that in this case the data also follows an IGRV model.

Theorem 1 *Let \mathbf{X} be an IGRV with $M \times M$ precision matrix Q_X . Let H be a $N \times M$ matrix and \mathbf{N} be a IGRV independent of \mathbf{X} , with $N \times N$ precision matrix Q_N . Assume that $H^t Q_N H + Q_X$ is a regular matrix, then $\mathbf{Y} = H\mathbf{X} + \mathbf{N}$ defines an IGRV with precision matrix:*

$$Q_Y = Q_N - Q_N H (H^t Q_N H + Q_X)^{-1} H^t Q_N. \quad (34)$$

Proof Let us start by $[\mathbf{X}, \mathbf{N}]$, whose density reads:

$$p_{\mathbf{X}, \mathbf{N}}(\mathbf{x}, \mathbf{n}) \propto \exp - \frac{1}{2} (\mathbf{x}^t Q_X \mathbf{x} + \mathbf{n}^t Q_N \mathbf{n}). \quad (35)$$

Then, performing the linear transform $[\mathbf{X}, \mathbf{N}] \rightarrow [\mathbf{X}, \mathbf{Y}]$ with $\mathbf{Y} = H\mathbf{X} + \mathbf{N}$

$$p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) \propto \exp - \frac{1}{2} \mathbf{x}^t Q_X \mathbf{x} + (\mathbf{y} - H\mathbf{x})^t Q_N (\mathbf{y} - H\mathbf{x}). \quad (36)$$

Then performing the linear transform $[\mathbf{X}, \mathbf{Y}] \rightarrow [\tilde{\mathbf{X}}, \mathbf{Y}]$ with $\tilde{\mathbf{X}} = \mathbf{X} - Q_X^{-1} H^t Q_N \mathbf{Y}$, $Q_{\tilde{\mathbf{X}}} = H^t Q_N H + Q_X$ and $Q_Y = Q_N - Q_N H (H^t Q_N H + Q_X)^{-1} H^t Q_N$:

$$p_{\tilde{\mathbf{X}}, \mathbf{Y}}(\tilde{\mathbf{x}}, \mathbf{y}) \propto \exp - \frac{1}{2} \tilde{\mathbf{x}}^t Q_{\tilde{\mathbf{X}}} \tilde{\mathbf{x}} - \frac{1}{2} \mathbf{y}^t Q_Y \mathbf{y}. \quad (37)$$

Therefore, $\tilde{\mathbf{X}}$ and \mathbf{Y} are independent, and Q_Y is the precision matrix of \mathbf{Y} . While $\tilde{\mathbf{X}}$ is proper with covariance $Q_{\tilde{\mathbf{X}}}^{-1}$, \mathbf{Y} is improper. The kernel of Q_Y has a simple connection with that of Q_X :

Corollary 1 *Under the conditions of Theorem 1, and the condition that Q_N is regular,*

$$\text{Ker } Q_Y = H \text{Ker } Q_X. \quad (38)$$

Proof It follows from (36) and (37) that

$$\mathbf{x}^t Q_X \mathbf{x} + (\mathbf{y} - H\mathbf{x})^t Q_N (\mathbf{y} - H\mathbf{x}) = \tilde{\mathbf{x}}^t Q_{\tilde{\mathbf{X}}} \tilde{\mathbf{x}} + \mathbf{y}^t Q_Y \mathbf{y}, \quad (39)$$

where $\tilde{\mathbf{X}} = \mathbf{x} - Q_X^{-1} H^t Q_N \mathbf{y}$. Let us first show that $\text{Ker} Q_Y \subset H \text{Ker} Q_X$. Let $\mathbf{y} \in \text{Ker} Q_Y$, and define $\mathbf{x} = Q_X^{-1} H^t Q_N \mathbf{y}$. For such \mathbf{x} and \mathbf{y} , the r.h.s of (39) is null, so that both terms of the l.h.s of (39) are null. So $\mathbf{x} \in \text{Ker} Q_X$ and $\mathbf{y} = H\mathbf{x}$, i.e., $\mathbf{y} \in H \text{Ker} Q_X$. The converse inclusion $H \text{Ker} Q_X \subset \text{Ker} Q_Y$ follows the same route, but in the opposite way. Let $\mathbf{y} \in H \text{Ker} Q_X$, there is at least one \mathbf{x} , such that $\mathbf{y} = H\mathbf{x}$ and $Q_X \mathbf{x} = 0$. For such \mathbf{x} and \mathbf{y} , the l.h.s of (39) is null, so that both terms of the r.h.s of (39) are null. In particular $\mathbf{y}^t Q_Y \mathbf{y} = 0$ means that $\mathbf{y} \in \text{Ker} Q_Y$.

A.4. Fisher information of an improper Gaussian distribution

For a zero-mean Gaussian statistical model parametrized by regular precision matrices Q_θ , it can be shown using standard Gaussian identities that the Fisher information is:

$$\text{FI}(\theta) = \frac{1}{2} \text{tr} \left(Q_\theta^{-1} \frac{dQ_\theta}{d\theta} Q_\theta^{-1} \frac{dQ_\theta}{d\theta} \right) \quad (40)$$

The proof of Eq. (40) can be found in [31], the main difference is that we parametrize the data density using the precision matrix instead of the covariance matrix. Using the results of proposition 1, the latter identity generalizes to improper Gaussian distributions. It amounts to replacing the inverse of Q_θ by its pseudo-inverse:

$$\text{FI}(\theta) = \frac{1}{2} \text{tr} \left(Q_\theta^+ \frac{dQ_\theta}{d\theta} Q_\theta^+ \frac{dQ_\theta}{d\theta} \right) \quad (41)$$

Proof Let A be any $n \times N$ matrix whose rows span the range of Q_θ , then $\mathbf{Y} = A\mathbf{X}$ is a Gaussian vector with regular precision matrix $A^{+t} Q_\theta A^+$. Using Eq. (40) with this precision matrix, $\text{FI}(\theta) = \frac{1}{2} \text{tr}(K^2)$ with:

$$K = (A^{+t} Q_\theta A^+)^{-1} A^{+t} \frac{dQ_\theta}{d\theta} A^+. \quad (42)$$

Note that we have used the fact that A does not depend on θ to remove the derivative operator around A . Using $\text{tr}(BC) = \text{tr}(CB)$ it follows that:

$$\begin{aligned} \text{tr}(K^2) &= \\ \text{tr} \left(A^+ (A^{+t} Q_\theta A^+)^{-1} A^{+t} \frac{dQ_\theta}{d\theta} A^+ (A^{+t} Q_\theta A^+)^{-1} A^{+t} \frac{dQ_\theta}{d\theta} \right). \end{aligned} \quad (43)$$

It can be shown that the Fisher information does not depend on a particular choice of matrix A . Moreover, if we choose A with orthonormal rows we have $A^+ = A^t$ and:

$$A^+ ((A^+)^t Q_\theta A^+)^{-1} (A^+)^t = A^t (A Q_\theta A^t)^{-1} A = Q_\theta^+. \quad (44)$$

Thus, Eq. (41) holds.

Appendix B: Analytical formula of the performance model for Gaussian PSF

In this section we show that an analytical formula for σ_{CRB} can be derived, assuming that the matrices H and D are circulant, that the PSF can be modeled with a Gaussian model and that the defocus blur is large.

Let first recall that according to Eq. (11) and (14) the σ_{CRB} is calculated using:

$$\text{FI}(\theta) = \frac{1}{2} \text{tr} \left(Q_\theta + \frac{dQ_\theta}{d\theta} Q_\theta + \frac{dQ_\theta}{d\theta} \right), \quad (45)$$

with $Q_\theta = \frac{1}{\sigma_n^2} P_\psi$ and

$$P_\psi = I - H_z (H_z^t H_z + \alpha D^t D)^{-1} H_z^t. \quad (46)$$

Assuming that the matrices H and D are $\sqrt{N} \times \sqrt{N}$ circulant, they can be diagonalized [24, Section 4.3.2]:

$$H_z = W^* \tilde{H}_z W \quad (47)$$

$$D = \begin{bmatrix} W^* \tilde{L}_1 W \\ W^* \tilde{L}_2 W \end{bmatrix}, \quad (48)$$

where W is the unitary matrix associated to Fourier decomposition, \tilde{H}_z , \tilde{L}_1 , \tilde{L}_2 are diagonal matrices, whose diagonal coefficients are the 2D DFT of respectively the PSF matrix h , $[-1 \ 1]$ and $[-1 \ 1]^t$. Using this decomposition, one can write: $P_\psi = W^* \tilde{P}_\psi W$ with

$$\tilde{P}_\psi(m) = \frac{\alpha |\tilde{D}|^2(m)}{|\tilde{H}_z|^2(m) + \alpha |\tilde{D}|^2(m)}, \quad (49)$$

where $|\tilde{D}|^2 = |\tilde{L}_1|^2 + |\tilde{L}_2|^2$ and the shortened notation $M(m)$ syands for $M(m, m)$ for any matrix M . Replacing Eq. (49) into Eq. (45) leads to:

$$\text{FI}(\psi) = \frac{1}{2} \sum_{m=2}^N \left(\frac{1}{|\tilde{H}_z|^2(m) + \alpha |\tilde{D}|^2(m)} \frac{d(|\tilde{H}_z|^2)}{dz}(m) \right)^2, \quad (50)$$

where by convention $|\tilde{D}|^2(1) = 0$. Let $|\tilde{h}|^2$ be the squared modulus of 2D TFD of h and $|\tilde{d}|^2$ the sum of the squared modulus of the TFD of $[-1 \ 1]$ and $[-1 \ 1]^t$. The lexicographical representations of $|\tilde{h}|^2$ and $|\tilde{d}|^2$ correspond to the diagonal values of respectively $|\tilde{H}_z|^2$ and $|\tilde{D}|^2$. We propose to use the following approximations:

$$|\tilde{h}|^2(\mu, \nu, z) \simeq e^{-4\pi^2 \tau(z)^2 (\mu^2 + \nu^2)}, \quad (51)$$

$$|\tilde{d}|^2(\mu, \nu) \simeq 4 (\sin^2(\pi\mu) + \sin^2(\pi\nu)). \quad (52)$$

where μ and ν are reduced frequencies and $\tau(z) = \rho\epsilon(z)/p_x$. Derivation of $|\tilde{h}|^2$ with respect to z leads to $\sigma_{\text{CRB}}(z) = \text{FI}(z)^{-1/2} = A(z)/B(z)$ with:

$$A(z) = \frac{\sqrt{2} z^2 p_x}{16\pi^2 R s} \simeq \frac{\sqrt{2} z^2 p_x F/\#}{8\pi^2 f^2} \quad (53)$$

$$B(z) = \rho\tau(z) \sqrt{\sum_{\mu, \nu} \left((\mu^2 + \nu^2) \frac{|\tilde{h}|^2(\mu, \nu, z)}{|\tilde{h}|^2(\mu, \nu, z) + \alpha |\tilde{d}|^2(\mu, \nu)} \right)^2}. \quad (54)$$

Calculation of B amounts to analytically calculate:

$$S = \sum_{\mu, \nu} \frac{(\mu^2 + \nu^2)^2 |\tilde{h}|^4(\mu, \nu, z)}{(|\tilde{h}|^2(\mu, \nu, z) + \alpha |\tilde{d}|^2(\mu, \nu))^2}. \quad (55)$$

We approximate the sum S with :

$$I_{nt} = \iint_{\mu\nu} \frac{(\mu^2 + \nu^2)^2 |\tilde{h}|^4(\mu, \nu, z)}{(|\tilde{h}|^2(\mu, \nu, z) + \alpha |\tilde{d}|^2(\mu, \nu))^2} N d\mu d\nu \quad (56)$$

$$= \iint_{\mu\nu} (\mu^2 + \nu^2)^2 g(\mu, \nu) N d\mu d\nu, \quad (57)$$

$$\text{with } g(\mu, \nu) = \frac{|\tilde{h}|^4(\mu, \nu, z)}{(|\tilde{h}|^2(\mu, \nu, z) + \alpha |\tilde{d}|^2(\mu, \nu))^2}. \quad (58)$$

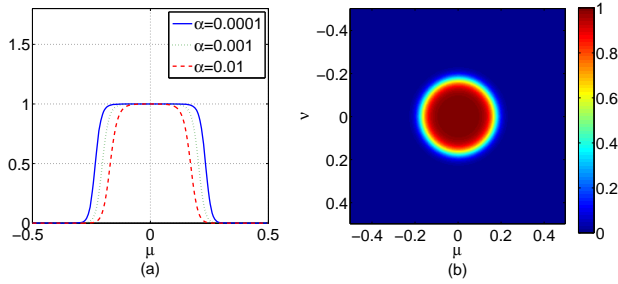


Fig. 11. (a) Slices of g at $\nu = 0$ for three values of the inverse SNR α . (b) Example of g for $\alpha = 0.01$ and $\tau = 2$

As illustrated in Fig. 11, g is close to the indicator function of a disk, whose size depends on the defocus level and the SNR. It is close to 1 for small frequencies, corresponding to $\alpha |\tilde{d}|^2 \ll |\tilde{h}|^2$, and close to 0 for large frequencies when $\alpha |\tilde{d}|^2 \gg |\tilde{h}|^2$. Thus we assume that $g = 1$ for $\mu^2 + \nu^2 < t^2$ and 0 elsewhere. Using polar coordinates with $r^2 = \mu^2 + \nu^2$, Eq. (57) becomes:

$$I_{nt} \simeq 2\pi N \int_0^t r^5 dr = \frac{2\pi}{6} N t^6. \quad (59)$$

We arbitrary fix the parameter t so that it corresponds to the intersection of the curves $|\tilde{h}|^2(\mu, \nu)$ and $\alpha |\tilde{d}|^2(\mu, \nu)$, besides we use the following approximation:

$$4(\sin^2(\pi\mu) + \sin^2(\pi\nu)) \simeq 4\pi^2(\mu^2 + \nu^2), \quad (60)$$

which is valid for small frequencies. This means that t is the solution of the equation:

$$4\alpha\pi^2 t^2 = e^{-4\pi\tau^2(z)t^2}. \quad (61)$$

Eq. (61) can be written as:

$$X e^X = \frac{\tau^2(z)}{\alpha}, \quad (62)$$

with $X = 4\pi\tau^2(z)t^2$. The solution of Eq. (62) corresponds to the Lambert W function value at $\frac{\tau^2(z)}{\alpha}$. Thus one can write:

$$t = \frac{1}{2\pi\tau(z)} \sqrt{W\left(\frac{\tau^2(z)}{\alpha}\right)}. \quad (63)$$

For large x , $W(x) \sim \ln x - \ln \ln x$. To use this equivalence we need to assume that $\tau(z)^2$ is large with respect to the inverse SNR α . This assumption amounts to consider that the defocus blur is high, while α is relatively low. Hence we assume that $|\tau| > 1$ and $\alpha < 1$. The combination of Eq.(59) and Eq.(63) gives:

$$I_{nt} \sim \frac{\left(\ln \frac{\tau^2(z)}{\alpha} - \ln \ln \frac{\tau^2(z)}{\alpha}\right)^3 N}{6(2\pi)^5 \tau^6(z)}. \quad (64)$$

To conclude B is approximated with:

$$B(z) \sim \frac{\sqrt{N} \rho \left(\ln \frac{\tau^2(z)}{\alpha} - \ln \ln \frac{\tau^2(z)}{\alpha}\right)^{3/2}}{\tau^2(z) (2\pi)^{5/2} \sqrt{6}}. \quad (65)$$

Combining Eq. (53) and (65),

$$\sigma_{\text{BCR}}(z) \simeq \kappa \frac{z^2 F / \#p_x}{f^2} \frac{\tau^2(z)}{\sqrt{N} \rho \left(\ln \frac{\tau^2(z)}{\alpha} - \ln \ln \frac{\tau^2(z)}{\alpha}\right)^{3/2}}, \quad (66)$$

with $\kappa = \sqrt{6\pi}$.

Fig. 12(a) and (b) show the variations of σ_{CRB} obtained with Eq. (16) with a Gaussian PSF and with Eq. (66), and the absolute error in percentage, respectively, for a camera with parameters given in Table 1 with the in-focus plane at 1.5 m and the inverse SNR $\alpha = 0.001$. The patch size is 31×31 pixels.

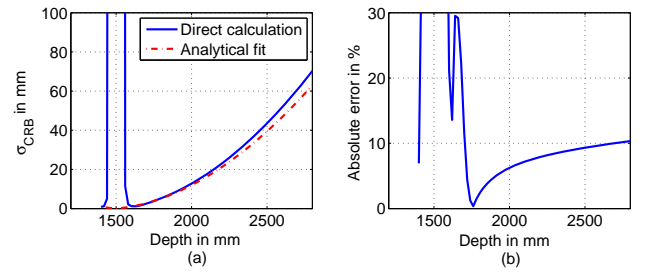


Fig. 12. (a) σ_{CRB} calculated with Eq. (16) and Eq. (66) for the same camera. (b) Absolute error in percentage of the analytical formula value with respect to the direct calculation of (16).

The absolute error away from the in-focus plane is below 10%, which is acceptable for an asymptotic formula. As σ_{CRB} calculated with Eq. (66) is derived from the approximation of circulant matrices, one can observe that the error decreases when the patch size increases. In contrast, increasing the patch size induces a

burdensome calculation of Eq. (16) due to inversion of P_z and of $H^t H + \alpha D^t D$ for large matrices. Therefore, the asymptotic analytical formula should be favored for large patches.

References

- [1] A. Pentland, "A new sense for depth of field". IEEE Trans. Pattern Anal. Mach. Intell. **4**, 523–531 (1987).
- [2] M. Subbarao, "Parallel depth recovery by changing camera parameters," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 1988), pp. 149–155.
- [3] P. Green, W. Sun, W. Matusik and F. Durand, "Multi-aperture photography," ACM Trans. on Graphics **26**, 1–7 (2007).
- [4] H. Nagahara, C. Zhou, C., T. Watanabe, H. Ishiguro and S. Nayar, "Programmable aperture camera using LCoS," in *Proceedings of IEEE European Conference on Computer Vision* (IEEE, 2010), pp. 337–350.
- [5] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," ACM Trans. on Graphics **26**, 1–12 (2007).
- [6] A. Levin, R. Fergus, F. Durand and W.T. Freeman, "Image and depth from a conventional camera with a coded aperture," ACM Trans. on Graphics **26**, 1–9 (2007).
- [7] S. Zhuo and T. Sim, "On the recovery of depth from a single defocused image," in *International Conference on Computer Analysis of Images and Patterns*, (Springer 2009), pp. 889–897.
- [8] M. Martinello and P. Favaro: "Single image blind deconvolution with higher-order texture statistics," Video Processing and Computational Video, 124–151 (2011).
- [9] P. Trounev, F. Champagnat, G. Le Besnerais and J. Idier, "Single image local blur identification," in *Proceedings of IEEE Conference on Image Processing* (IEEE, 2011), pp. 613–616.
- [10] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Comp. Science Tech. Report, 2005.
- [11] A. Chakrabarti and T. Zickler, "Depth and deblurring from a spectrally varying depth of field," in *Proceedings of IEEE European Conference on Computer Vision* (IEEE, 2012), pp. 648–661.
- [12] P. Trounev, F. Champagnat, J. Sabater, T. Avignon, G. Le Besnerais and J. Idier, "Passive depth estimation using chromatic aberration and a depth from defocus approach," Applied Optics **52**, 7152–7164 (2013).
- [13] Y. Schechner and N. Kiryati. "Depth from defocus vs. stereo: How different really are they?," Int. J. of Computer Vision, **39**, 141–162 (2000).
- [14] 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).
- [15] R. Blendowske, "Role of optics in the accuracy of depth-from-defocus systems: comment", J. Opt. Soc. Am. A **24**, 3242–3244 (2007).
- [16] 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. of Comp. Vision **30**, 175–190 (1998).
- [17] S.M. Kay, Fundamentals of statistical signal processing: estimation theory, Prentice-Hall, 1993.
- [18] F. Aguet, D. Van De Ville and M. Unser, "A maximum-likelihood formalism for sub-resolution axial localization of fluorescent nanoparticles", Opt. Express **13**, 10503–10522 (2005).
- [19] C. Fournier, L. Denis and T. Fournel, "On the single point resolution of on-axis digital holography", J. Opt. Soc. Am. A **27**, 1856–1862 (2010).
- [20] G. Johnson, E. Dowski and W. Cathey, "Passive ranging through wave-front coding: information and application," Applied Optics **39**, 1700–1710 (2000).
- [21] A. Greengard, Y. Schechner and R. Piestun, "Depth from diffracted rotation," Opt. Letters **31**, 181–183 (2006).
- [22] A. Levin, Y. Weiss, F. Durand and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms", in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2009), pp. 88–101.
- [23] F. Orieux, J.-F. Giovannelli, T. Rodet, "Bayesian estimation of regularization and point spread function parameters for Wiener-Hunt deconvolution", J. Opt. Soc. Am. A **27**, 1593–1607 (2010).
- [24] J. Idier, *Bayesian approach to inverse problems*, ISTE Ltd and John Wiley & Sons Inc, 2008.
- [25] J.W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, 1996.
- [26] M. Delbracio, P. Musé, A. Almansa and J. Morel, "The non-parametric sub-pixel local point spread function estimation is a well posed problem," Int. J. Comp. Vision **96**, 175–194 (2012).
- [27] J. Besag and C. Kooperberg, "On conditional and intrinsic autoregressions," Biometrika **82**, pp. 733 (1995).
- [28] A. Neumaier, "Solving Ill-Conditioned and Singular Linear Systems: A Tutorial on Regularization", SIAM Rev. **40**, 636–666 (1998).
- [29] G. Wahba, "A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem", The Annals of Statistics, JSTOR **13**, 1378–1402 (1985).
- [30] P.R. Halmos, *Measure theory*, Int. Student Editions, D. Van Nostrand Company, New York, 1969.
- [31] B. Porat and B. Friedlander, "Computation of the exact information matrix of Gaussian time series with stationary random components", IEEE Trans. on Acoustic, Speech and Signal Proc. **34**, 118–130 (1986).