Design of a chromatic 3D camera with an end-to-end performance model approach

P. Trouvé[1] F. Champagnat[1] G. Le Besnerais[1] G. Druart[1] J. Idier[2]

[1]ONERA-The French Aerospace Lab
F-91761 Palaiseau, France
{pauline.trouve,frederic.champagnat}@onera.fr
{guy.le_besnerais,guillaume.druart}@onera.fr

[2]LUNAM Université, IRCCyN
UMR CNRS 6597 BP 92101
1 rue de la Noë, 44321 Nantes Cedex 3, France
jerome.idier@irccyn.ec-nantes.fr

Abstract

In this paper we present a new method for the design of a 3D single-lens single-frame passive camera. This camera has a chromatic lens and estimates depth based on a depth from defocus technique (DFD). First we develop an original calculation of the Cramér Rao Bound to predict the theoretical camera accuracy. This model takes into account the optical parameters through the camera Point Spread Function (PSF) and the algorithms parameters applied to the raw image for depth estimation and image restoration. This model is then used for the end-to-end design of a chromatic camera, dedicated to a small UAV, that is realized and experimentally validated.

1. Introduction

The increasing interest for 3D camera has led to the development of several depth measure techniques. For instance, cheap active cameras such as Kinect [7] are now available. They estimate depth using a projected pattern but are sensitive to perturbation of this pattern, for instance due to sunlight. Passive 3D cameras using parallax effects are either cumbersome, when using two separated cameras as in stereoscopy, or reduce image resolution, in the case of the plenoptic camera [16], or reduce signal to noise ratio (SNR) in the case of the color filtered aperture of [1]. Other passive solutions use depth from focus (DFF), which relies on estimation of the sharpest image among a set of images acquired with varying focus [15], or depth from defocus (DFD), i.e. local estimation of the defocus blur by comparing two or more images [17]. Such multiple-frames DFD or DFF cameras require the scene to be static during the acquisitions which restricts the field of applications. Thus, although more computationally demanding, single frame DFD methods address a larger field [11, 14, 13, 22]. Yet these techniques have limitations: they suffer from a dead zone in the depth of field region where blur is quasi-uniform so that depth can not be estimated and there is an ambiguity between depths ahead and behind the in-focus plane, which yield similar defocus blurs.

In the literature various modifications of the camera optics are proposed to overcome these defaults and to improve depth estimation accuracy with a single frame DFD method. Coded apertures are thoroughly studied in [11, 14, 13]. Another approach is to use a lens with spectrally varying blur using either a chromatic aperture [2], or a chromatic lens with some amount of longitudinal chromatic aberration [8, 21]. The latter approach can avoid depth ambiguity or dead zone, in contrast with all other solutions [11, 14, 13, 2]. Besides, the camera light intensity is not reduced as in [1, 11, 14, 13] which leads to a higher SNR. Finally as mentioned in [9, 5, 21], the use of chromatic aberration tends to reduce the lens dimension.

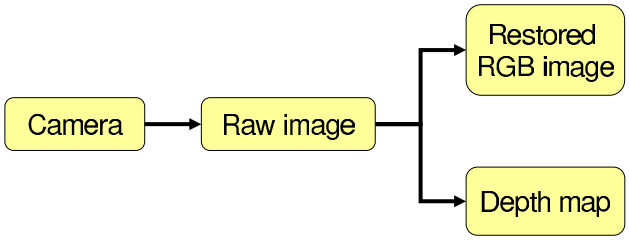


Figure 1. Principle of a computational 3D camera.

Note that these various optical choices increase depth estimation accuracy but also usually lead to a degradation of the image quality that has to be corrected by a dedicated processing. Thus, as illustrated in figure 1, a computational 3D camera will have two main processing blocks, one for depth estimation and one for image restoration. Besides, such solutions also require to rethink the lens design.

Here we base the lens design on the optimization of a theoretical criterion which accounts for both optic and processing parameters, an approach referred to as codesign, that is briefly reviewed in the next section.

978-0-7695-4990-3/13 $26.00 © 2013 IEEE
DOI 10.1109/CVPRW.2013.140

1.1. Codesign for computational 3D cameras

The design of computational camera requires to add criteria related to the processing parameters to the classical design criteria only based on optical image quality. Here, as illustrated in figure 1 for a computational 3D camera, we need two criteria: a first one for depth estimation accuracy and another one for post-processing image quality.

Lens optimization for depth estimation relies on increasing PSF variation with depth. In [11] the coded aperture is optimized with a maximisation of the Kullback distance between potential codes. However this optimization concerns only the aperture shape and not other sensor or processing parameters, which precludes the codesign of a complete imaging system. In [13] a depth estimation performance evaluation based on the imaging system PSFs simulation and a Bayesian formalism is proposed. However the proposed evaluation criterion is calculated using a learned database, which reduces the generality of the performance evaluation. Besides, the calculation is too computationally intensive to be applied for a complete imaging system optimization. Note also that both references [11, 13] only provide a global score to an imaging system without physical interpretation of the performance. In contrast, we propose here, in the case of a single frame DFD 3D camera, a complete theoretical performance model that predicts for each depth its estimation accuracy. This model relies on an original calculation of the Cramér-Rao bound (CRB) which gives the best expectable standard deviation for depth estimation. Based on a generic scene model, the CRB can be computed for any given imaging system. Moreover, it can be efficiently evaluated using Fast Fourier Transform (FFT), allowing joint optimization of all relevant camera parameters at an affordable cost.

Our work can be related to references [19, 21] which also use CRB. However, the CRB in [19] is used in multiple images DFD to optimize the blur ratio between the two images, while we address here single frame chromatic DFD and study the accuracy variation with respect to depth. In [21] the CRB is used to compare theoretical accuracy and experimental depth estimation performance of an existing chromatic lens. In contrast, we use the CRB for the design of a chromatic DFD 3D camera.

Concerning the image restoration part of the system, its optimization can be related to the problem of depth of field extension (EDOF). In most works concerning EDOF, imaging systems are optimized in order to have a PSF invariant to depth for a large range allowing a global image deconvolution algorithm. For instance in [5] an optical design software is used to optimize the polychromatic PSF of the chromatic lens to make it approximately invariant to depth over a large depth range. In several works, the optimization is essentially based on deconvolution errors or post-processing SNR [6, 23, 20]. In the case of a 3D camera with chromatic aberration, another approach for image restoration is to transfer high frequency content of sharp channels to blurred ones [9]. For such approaches we propose a simple evaluation of the image restoration performance based on the union of each channel depth of field.

1.2. Paper organization

In Section 2, we present our performance model for depth estimation accuracy of a 3D DFD camera. We use it to evaluate the two chromatic solutions described in [2, 21], and predict that the chromatic lens leads to better performances than a chromatic aperture. Section 3 is devoted to algorithmic issues, both for depth estimation and image restoration. This section ends with an empirical comparison of the systems of [2, 21] on simulated images, which confirms the better potential of the chromatic lens of [21]. Therefore, we focus on the chromatic lens approach and propose the codesign of a chromatic camera in Section 4, using theoretical performance criteria both for depth estimation and image quality. Here we focus on the design of a 3D camera for a small UAV. The resulting camera is evaluated in Section 5: it is shown that the experimental accuracy is in good agreement with the expectations. Finally, we conclude and give directions of current research in Section 6.

2. Performance model

2.1. Cramér Rao Bound

We propose to use the Cramér Rao Bound (CRB) in order to predict the theoretical depth estimation accuracy of a 3D DFD camera. For the estimation of a parameter θ from a data vector $\boldsymbol{y}$ the CRB writes:

$$\mathrm{var}(\hat{\theta}) \geq \sigma^2_{\mathrm{CRB}} = \mathrm{FI}(\theta)^{-1} = -E\left[\frac{\partial^2 \ln p(\boldsymbol{y}|\theta)}{\partial^2 \theta}\right]^{-1}, \quad (1)$$

where FI is the Fisher Information, $p(\boldsymbol{y}|\theta)$ the likelihood of the data $\boldsymbol{y}$ and E the expectation function. For a Gaussian vector $\boldsymbol{y}$ of density proportional to $\exp -\frac{1}{2}\boldsymbol{y}^t Q_\theta \boldsymbol{y}$, one can show that:

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

where Q_θ is the precision matrix (i.e. the inverse of the covariance matrix) of the data $\boldsymbol{y}$. The proof of (2) can be found in [18], the main difference is that we parameterize the data density using the precision matrix instead of the covariance matrix. According to equation (2) the FI, and thus the CRB, depends only on the precision matrix that we analytically express in the following using simple scene and data priors.

2.2. Image model

Defocus blur is a spatially varying blur, so an image patch is usually modeled with the local convolution of a scene patch with the PSF and addition of random acquisition noise. Using the vector representation on image and scene patches we have:

$$\boldsymbol{Y} = H\boldsymbol{X} + \boldsymbol{N}, \tag{3}$$

with $\boldsymbol{Y} = [\boldsymbol{y}_R^t \boldsymbol{y}_G^t \boldsymbol{y}_B^t]^t$, $\boldsymbol{X} = [\boldsymbol{x}_R^t \boldsymbol{x}_G^t \boldsymbol{x}_B^t]^t$, where for each channel c, $\boldsymbol{y}_c$ (respectively $\boldsymbol{x}_c$) collects pixels of the image (resp. scene) patch in the lexicographical order. $\boldsymbol{N}$ stands for the noise process which is modeled as a zero mean white Gaussian noise (WGN) with variance σ_n^2. The observation matrix writes:

$$H(d) = \begin{bmatrix} H_R(d) & 0 & 0 \\ 0 & H_G(d) & 0 \\ 0 & 0 & H_B(d) \end{bmatrix}. \tag{4}$$

Each $H_c(d)$ is a convolution matrix which depends on the defocus PSF of the channel c. 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 $\boldsymbol{x}_c$ is enlarged with respect to the one of $\boldsymbol{y}_c$ according to the PSF support [10, Section 4.3.2]. N is the length of each vectors $\boldsymbol{y}_c$, M the length of each vector $\boldsymbol{x}_c$ thus each H_c is a convolution matrix of size $N \times M$. Note that the proposed formalism allows to model both 3CCD and color filter array (CFA) sensors. Modeling a CFA sensor just amounts to remove adequate lines from full convolution matrices H_c.

2.3. Scene model

In the context of local PSF estimation, a Gaussian prior on the scene is often very effective as shown for instance in [12, 3, 22]. However as mentioned in [21], when dealing with chromatic data the components in the RGB decomposition are partially correlated. Following [4, 21] 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:

$$\begin{bmatrix} \boldsymbol{x}_R \\ \boldsymbol{x}_G \\ \boldsymbol{x}_B \end{bmatrix} = T \otimes \mathbf{I}_{M,M}, \begin{bmatrix} \boldsymbol{x}_L \\ \boldsymbol{x}_{C_1} \\ \boldsymbol{x}_{C_2} \end{bmatrix} \tag{5}$$

where $\otimes$ stands for the Kronecker product, $\mathbf{I}_{M,M}$ is the identity matrix of size $M \times M$ and:

$$T = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{bmatrix}. \tag{6}$$

According to [4], the three components of the luminance/chrominance (LC) decomposition can be assumed to be uncorrelated. We then use the Gaussian prior:

$$p(\boldsymbol{X}^{LC}, \sigma_{x_C}^2) \propto \exp\left(-\frac{\|D_C \boldsymbol{X}^{LC}\|^2}{2\sigma_{x_C}^2}\right) \tag{7}$$

where $\boldsymbol{X}^{LC} = [\boldsymbol{x}_L^t \boldsymbol{x}_{C_1}^t \boldsymbol{x}_{C_2}^t]^t$ and:

$$D_C = \begin{bmatrix} \sqrt{\mu_c} D & 0 & 0 \\ 0 & D & 0 \\ 0 & 0 & D \end{bmatrix}. \tag{8}$$

D is the vertical concatenation of the convolution matrices relative to the horizontal and vertical first order derivation operator, and μ_C is the ratio of the luminance and the chrominance variances. As in [4] μ_c is fixed at 0.05. Thus the image model becomes:

$$\boldsymbol{Y} = H_C(d)\boldsymbol{X}^{LC} + \boldsymbol{N} \tag{9}$$

$$H_C(d) = H(d)T \otimes \mathbf{I}_{M,M}. \tag{10}$$

2.4. Likelihood marginalisation

The data likelihood is then derived through a marginalisation of the scene [3, 12, 22],

$$p(\boldsymbol{Y}|d, \sigma_n^2, \sigma_{x_C}^2) = \tag{11}$$

$$\int p(\boldsymbol{Y}|\boldsymbol{X}^{LC}, d, \sigma_n^2) p(\boldsymbol{X}^{LC}, \sigma_{x_C}^2) d\boldsymbol{X}^{LC}, \tag{12}$$

which is tractable only for a Gaussian prior on the scene. Replacing (7) into (12) and using a Gaussian density for the noise process we obtain:

$$p(\boldsymbol{Y}|\boldsymbol{\theta}) = \left|\frac{Q_{\boldsymbol{\theta}}}{2\pi}\right|_+^{\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{Y}^t Q_{\boldsymbol{\theta}} \boldsymbol{Y}\right), \tag{13}$$

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

$$Q_{\boldsymbol{\theta}} = \frac{1}{\sigma_n^2}\left[I - H_C(d)(H_C^t H_C(d) + \alpha D_C^t D_C)^{-1} H_C(d)^t)\right]. \tag{14}$$

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

$$\text{FI}(\psi) = \frac{1}{2}\text{tr}\left(P_\psi^{-1}\frac{dP_\psi}{d\psi}P_\psi^{-1}\frac{dP_\psi}{d\psi}\right). \tag{15}$$

2.5. Computation of the CRB

To simplify the calculation of the CRB, we assume that the signal to noise ratio (i.e. α) is known and focus only on depth estimation. This amounts to assume that $\psi = \{d\}$.

For each depth d we compute the convolution matrices H_R, H_G and H_B. These PSFs can be simulated either using a simple Gaussian or pill-box model, or based on Fourier optics principles or using an optical design software such as Zemax. Then for a given value of α, we compute the matrice P_ψ using (10), (14) and $P_\psi = \sigma_n^2 Q_\theta$. To compute FI($d$) with (15), we use the numerical differentiation:

$$\frac{dP_\psi}{d\psi} \simeq \frac{P_{\psi+\delta} - P_{\psi-\delta}}{2\delta}, \tag{16}$$

where δ is a small depth variation with respect to d. Taking the inverse square root of the result gives the theoretical minimum standard deviation σ_{CRB} of Eq. (1).

Note that to reduce calculation time, we decompose P_ψ with a Fourier transform and reorganize the frequencies to gather together the Fourier components of the three channels at the same frequency. We then deal with a 3×3 block diagonal matrix. Given the imaging system RGB PSFs at depth (d, $d-\delta$, $d+\delta$) and a patch size of 21×21 pixels, a value of σ_{CRB} is obtained in 80ms with a 3GHz processor.

2.6. Comparison of two 3D cameras performances

To illustrate the genericity of the proposed performance model, we compare the theoretical accuracy in depth estimation of two imaging systems. The first one has a chromatic lens as in [21] and the other one a chromatic aperture as proposed in [2]. The two imaging systems have the same main focal length, main f-number and detector pixel size, in order to impose the same optical constraints on both camera. We simulate for both cases the PSFs of the three RGB channels with Gaussian functions whose standard deviations, normalized with the pixel size, are given by:

$$\sigma_c(d) = \rho \frac{f_c d_{det}}{p_x F_{\#c}} \left(\frac{1}{f_c} - \frac{1}{d} - \frac{1}{d_{det}} \right), \tag{17}$$

where f_c is the focal length of each channel c, $F_{\#c}$ is the f-number of the channel c, d_{det} is the distance between the optic and the detector, p_x the detector pixel size, and ρ is a corrective parameter set to 0.25, so as to fit a Fourier optics model for defocusing.

According to [21], the green channel focal length of the chromatic lens is 25 mm and the f-number of the three channels is 4. The parameter d_{det} is calculated with the lens law in order to put the green channel in-focus plane at 2.7 m, as mentioned in [21]. The red and blue focal lengths are then calculated using the lens law so that the red and blue in-focus planes are respectively at 5 m and 1.9 m. For the chromatic aperture case, the focal length is set to 25 mm for the three channels. The f-number for the red and blue channels is set to 4 and is set to 6.8 for the green channel in order to have an aperture radius ratio of 0.59, as in [2]. The in-focus plane is put at 1.9 m. Figure 2 presents the values of σ_{CRB} obtained for the chromatic aperture and chromatic lens cases, with $\alpha = 0.001$, δ=1 mm and a patch size of 21×21 pixels.

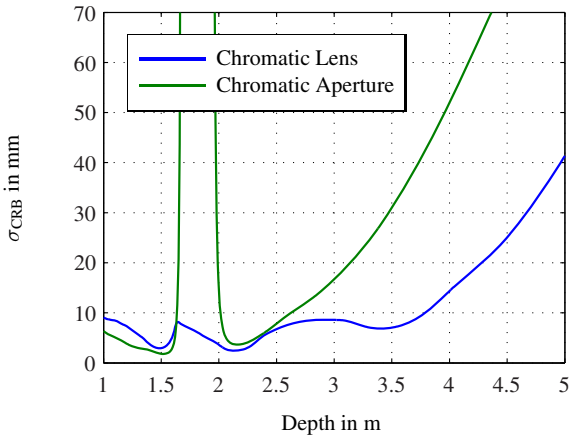


Figure 2. Comparison of theoretical depth estimation accuracy using the proposed performance model for an imaging system having a chromatic lens or a chromatic aperture.

For the chromatic aperture there is a divergence of σ_{CRB} around 1.9 m. This divergence corresponds to the dead zone for depth estimation around the in-focus plane, where no PSF variation can be observed. There the derivative of P_ψ is equal to zero, which gives an infinite σ_{CRB}. With the chromatic lens there is no divergence, thanks to the combination of the three in-focus planes. Figure 2 shows that before 1.6 m the accuracy of the chromatic aperture slightly outperforms the one of the chromatic lens but after 2.5 m the accuracy of the chromatic lens is much better. This study highlights the advantage of chromatic lens in terms of long range depth estimation accuracy. Note that the previous analysis is enabled by a distinctive feature of our performance model: it gives a performance index for each considered depth whereas previous works on DFD performance proposed global scores for a given imaging system [11, 13].

3. Algorithms for a chromatic 3D camera

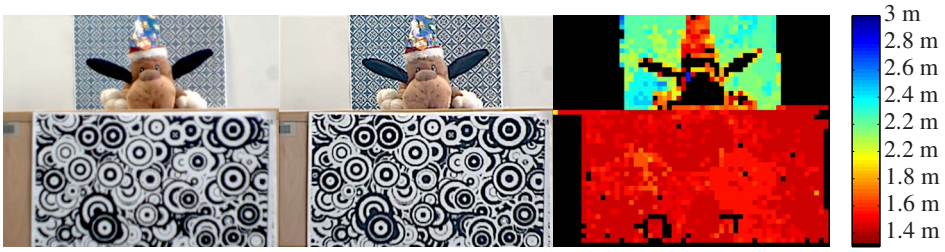
The present section is dedicated to the two algorithms required by a computational 3D illustrated in figure 1.

3.1. Chromatic DFD algorithm

3.1.1 Algorithm description

Many DFD algorithms estimate depth within a set of potential depths. This amounts to select a PSF among a set of potential PSF obtained by calibration [11, 13, 22]. More specifically here we have to select a PSF triplet among a set of potential triplets. This is done with a criterion derived from a maximum likelihood estimator. Using the same formalism as in section 2 the marginalized likelihood writes:

$$p(\boldsymbol{Y}|\psi, \sigma_n) = \left| \frac{P_\psi}{2\pi\sigma_n^2} \right|_+^{\frac{1}{2}} \exp\left(-\frac{1}{2} \frac{\boldsymbol{Y}^t P_\psi \boldsymbol{Y}}{\sigma_n^2} \right). \tag{18}$$



Raw image Restored image Raw depth map

Figure 3. Example of input and outputs of the codesign 3D camera. Black label is for homogeneous regions insensitive to defocus blur.

This likelihood depends on depth, noise variance, and the scene variance. In order to reduce the number of parameters we maximise this likelihood with respect to the noise variance. This leads to $\widehat{\sigma}_n^2 = \boldsymbol{Y}^t P_\psi \boldsymbol{Y}/(N-3)$. Replacing $\widehat{\sigma}_n$ into equation (18) gives:

$$p(\boldsymbol{Y}|\psi, \widehat{\sigma}_n^2) \propto |P_\psi|_+^{\frac{1}{2}} (\boldsymbol{Y}^t P_\psi \boldsymbol{Y})^{-(\frac{3N-3}{2})}.$$

Maximisation of this *generalized likelihood* amounts to minimise the criterion:

$$\text{CGL}(\psi) = \boldsymbol{Y}^t P_\psi \boldsymbol{Y}\, |P_\psi|_+^{\frac{-1}{(3N-3)}}. \qquad (19)$$

where CGL stands for chromatic generalized likelihood. If one writes $\hat{\alpha}_k = \arg\min_\alpha \text{CGL}(d_k, \alpha)$, depth can be estimated using the following criterion:

$$\hat{d} = \arg\min_k \text{CGL}(d_k, \hat{\alpha}_k). \qquad (20)$$

Note that this criterion can be seen as a generalization of the criterion proposed in [22] to the case of a chromatic lens. In this paper we use the same implementation than the one proposed in [22] based on generalized singular value decomposition of the matrices H_C and D_C.

3.1.2 Empirical chromatic DFD performance

Figure 4. Natural scenes used to simulate image patches.

In this section, we use the previous chromatic DFD algorithm on simulated images for the two chromatic imaging systems: a chromatic aperture and chromatic lens, related to [2, 21] described in the section 2.6. For each imaging system, we generate 120 image patches of size 21×21 pixels using scenes patches extracted from natural scenes presented in figure 4. For each depth and each imaging system the images are obtained with a convolution of scene patches with the corresponding Gaussian PSF. White Gaussian noise is added to the result with a standard deviation of 0.01, given that the scenes have a normalized intensity. Table 1 shows the bias and the standard deviation of the depth estimation results. For depths below 3 m the performances of both imaging systems are quite close, but after 3 m the chromatic lens system shows a much better performance than the chromatic aperture. This is consistent with the theoretical accuracy comparison made in section 2.6.

Depth	2 m		3 m		4 m		5 m	
Ref.	B	S	B	S	B	S	B	S
[21]	1	1.2	1.3	2.3	0.7	4.1	1.6	12
[2]	1.7	2.2	0.3	9.3	0.3	22	4.1	36

Table 1. Bias (B) and standard deviation (S) in cm on depth estimation results for two chromatic optical solutions using simulated images.

3.2. Image restoration

Chromatic aberration induces a inhomogeneous resolution among the RGB channels. Thus, as illustrated in Figure 3, the raw RGB image is blurred and requires restoration processing. As proposed in [9] a high frequencies transfer can be used to improve image resolution. Formally the restored channel is the sum of the original image with a weighted sum of the high frequencies of each channels. In [9], the weights depend on a relative sharpness measure and are set with a calibration step. In our case we simply use the depth map estimated to determine these weights. Thus we propose to restore each channel image using:

$$y_{i,\text{out}} = y_{i,\text{in}} + a_{d,R}\text{HP}_R + a_{d,G}\text{HP}_G + a_{d,B}\text{HP}_B. \qquad (21)$$

HP_c are the high frequencies of the channels c, obtained with a high pass filter. The values of $a_{d,c}$ are decreasing functions of $|d - d_{0,c}|$ and vary from 0 to 1 where d_0 is the in-focus plane position of the channel c. Figure 3 illustrates the resolution gain obtained with such restoration algorithm and the corresponding depth map obtained with the chromatic DFD algorithm on a real image acquired with the camera described in section 4.

4. Codesign of a chromatic 3D camera

The performance study conducted in section 2 illustrates theoretically and experimentally the high potential of a chromatic lens for depth estimation. Besides, the chromatic lens has no ambiguity around the in-focus plane which enlarges the depth estimation range. Thus we now proceed with the optimization of a DFD 3D camera with a chromatic lens using our performance model.

4.1. Specifications

Our aim is to design a 3D camera that could be embedded on a UAV that is moving in outdoor and indoor conditions. For this application we set the depth estimation range from 1 to 5 m with a required depth accuracy of 10 cm. This imaging system intends to allow the UAV to reach a point in front of it, so it does not require to have a large field of view. Thus, we restrict the field of view at 25°. Since there is a finite family of existing color sensors, we can hardly continuously optimize the sensor parameters. Thus we choose a color sensor and optimize a chromatic lens for it. The chosen color sensor has a pixel size of 3.45μm with a resolution of 2046×2452 pixels.

These specifications give us information about the camera. Indeed the values of the sensor size and the field of view lead to a focal length 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. Besides, we want the UAV to identify obstacles such as electric wires, posts or scaffolds, thus the depth map spatial X-Y resolution is fixed to approximately 2 cm at 3 m. Thus the depth map spatial resolution is of 160 μm in the image plane. This resolution limits the patch size to 46×46 pixels on the sensor. Since we use a Bayer color sensor it amounts to process patches of size 23×23 pixels. Yet we need to define the amount of longitudinal chromatic aberration of the lens, characterized by the RGB in-focus planes positions.

4.2. Design criteria

4.2.1 Depth estimation accuracy

In order to optimize depth estimation for some depth range we propose a design criterion named C_1 based on the mean value of the σ_{CRB} in the range L:

$$C_1(L) = < \sigma_{\text{CRB}}(d) >_{d \in L} . \qquad (22)$$

4.2.2 Image quality

As chromatic aberration reduces image quality and, as mentioned in section 3.2, we use a high frequencies transfer method to improve image resolution. To manage this transfer, we need to have at least a sharp channel at each depth. Thus we define an image quality criterion that measures the union of the depth of field (DOF) of each channels inside of the sought camera depth range L. This criterion named C_2 can be interpreted as a generalized depth of field (GDOF) of the camera after processing, a quantity that is defined as:

$$C_2 = \text{GDOF} = L \bigcap \left(\bigcup_{c=R,G,B} \text{DOF}_c \right) . \qquad (23)$$

Figure 5 illustrates this quantity.

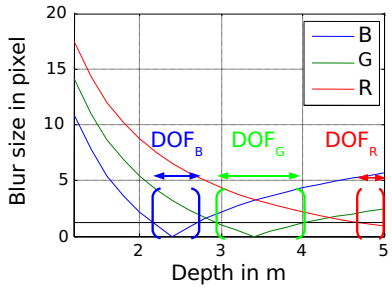


Figure 5. Generalized depth of field.

4.3. First order camera optimization

In this section our aim is to use our performance criteria to find a first approximation of the optimal RGB in-focus planes positions. To do so we simulate the PSFs associated to various chromatic imaging systems, having a focal length of 25 mm at the green channel, a f-number of 3, and a pixel size of 3.45μm. Each system has a different triplet of RGB in-focus planes. As in section 2.6, the PSFs are Gaussian with a standard deviation defined in equation (17). We calculate for each potential system the criteria C_1 and C_2 and thus obtain figure 6. As shown in this figure, maximisation of C_2 or minimisation of C_1 does not lead to the same in-focus planes. Hence, a trade-off has to be found: 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 triplets. The obtained triplet is plotted with the red cross on figure 6. Table 6 presents the optimal RGB in-focus planes obtained respectively after minimisation of C_1, or maximisation of C_2, or with the proposed trade-off approach.

Criterion	Min(C_1)	Max(C_2)	Trade-off
${d_0}^B$(m)	2.2	2.8	2.2
${d_0}^V$(m)	3.6	3.4	3.4
${d_0}^R$(m)	4.2	4.4	4.2

Table 2. Optimal RGB in-focus planes triplets ($[{d_0}^R {d_0}^V {d_0}^B]$) regarding resp. minimisation of criterion C_1, maximisation of criterion C_2 and the trade-off.

According to the proposed trade-off, we select the triplet of in-focus planes for the RGB channels respectively at

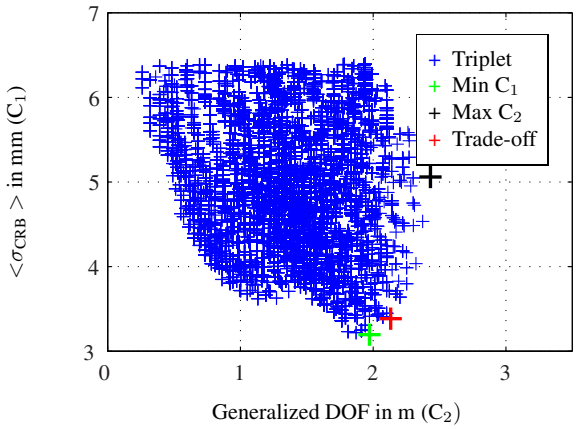


Figure 6. Simulated systems scores with respect to C_1 and C_2.

4.2 m, 3.4 m and 2.2 m. This corresponds to a longitudinal chromatic aberration $f_R - f_B$ around 130 μm.

4.4. Refining the lens optimization

The previews first order optimization gives us the approximate optimal position of the RGB in-focus planes and the required amount of longitudinal chromatic aberration. According to this constraints, a first architecture is designed using the optical software `Zemax`. In contrast to section 4.3, we now deal with the real physical lens parameters as for instance lens curvature radius, thickness or glass type. Starting from this first architecture a more accurate optimization of the lens parameters is conducted. In this case we use jointly the two criteria C_1 and C_2, evaluated using the PSF simulated by `Zemax`, and the image quality optimization algorithms of this software. This leads to a chromatic lens of longitudinal chromatic aberration of 100μm, with RGB in-focus planes respectively at 3.7, 2.7 and 2.2 m. The resulting lens architecture is shown in figure 7.

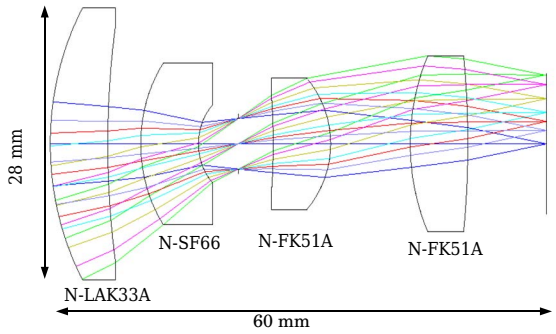


Figure 7. Architecture of the codesigned lens.

5. Experimental results

We have realized the chromatic lens, according to the specifications obtained in section 4.4 and evaluate here its experimental performance.

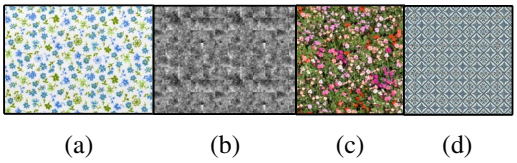


Figure 8. Scenes used as targets in the experiments.

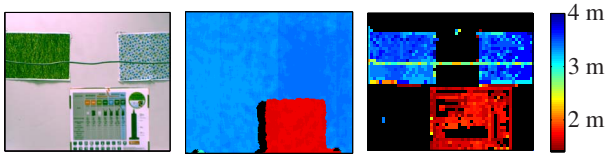


Figure 10. From left to right: RGB image, Kinect and codesigned camera depth maps. Black label is for homogeneous regions.

5.1. On axis depth estimation accuracy

The PSFs of each channel of the codesigned chromatic lens are calibrated from 1 m to 5 m with a step of 5 cm, with a ground truth given by a telemeter. Acquisitions are made of colored textured plane scenes put at different distances from the lens. For each scene and at each distance, depth is estimated with the proposed DFD chromatic algorithm on image 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 patches overlapping of 50%. Figure 8(a) to (d) show four of the scenes used in the experiment and figure 9(a) to (d) show the corresponding mean and the standard deviation of the depth estimation results with respect to the ground truth. Table 3 gives the statistical results for each scene on the full range 1 to 5 m.

Figure 9	(a)	(b)	(c)	(d)
Mean bias (absolute value)(cm)	4	4	10	4
Mean standard deviation (cm)	7	6	7	6

Table 3. Experimental performances of the codesigned camera.

For each scene, bias is comparable to the PSF calibration step (5 cm) and standard deviation is on the order of 7 cm. This results illustrates the good performance of the camera for depth estimation in the specified depth range.

5.2. Depth map

Figure 10 shows an example of depth map obtained with our camera. Because of the PSF variation with field angle, PSF calibration is carried out off axis for 9 image regions where the PSF is assumed to be constant. The depth map obtained with the chromatic DFD algorithm is compared to the depth map given by the Kinect. On textured regions, both 3D camera give the same depth levels. In contrast to the Kinect, which is an active system, we do not estimate depth on homogeneous regions, because they are insensitive to defocus. On the other hand, the wire is visible in our depth map and does not appear with the Kinect.

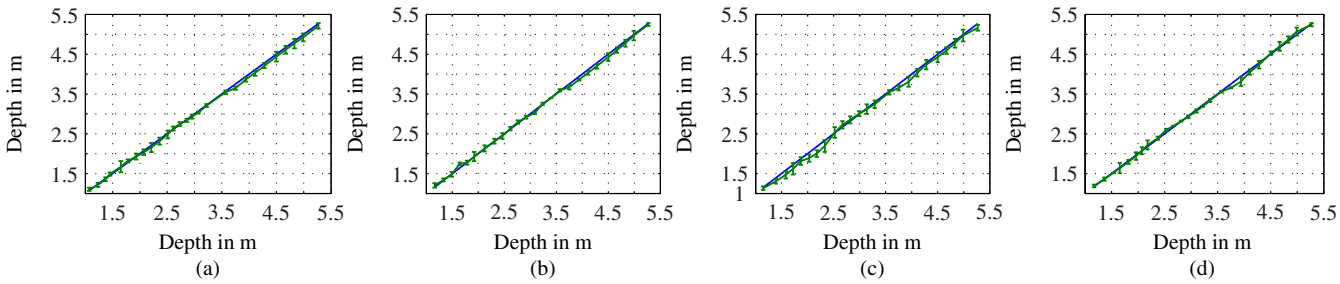


Figure 9. Four on axis depth estimation results resp. for the four targets presented in figure 8. Experimental mean and standard deviation are plotted with error bars (green) with respect to the ground truth given by a telemeter (blue).

6. Conclusion

In this paper we have presented the end-to-end design of a 3D chromatic camera. The accuracy of the depth estimation of such a camera is modeled using an original calculation of the CRB with generic prior on the scene. The model is used to compare a priori two chromatic concepts, namely chromatic lens and chromatic aperture. The model predicts a better depth accuracy on a larger range for the chromatic lens concept. This prediction is confirmed by empirical evaluation of the depth estimation error statistics for a proposed chromatic DFD estimator on simulated images.

Following the path of the chromatic lens concept, we have designed a chromatic camera using two criteria: one for the image quality and another one for depth estimation accuracy. A prototype of the codesigned camera has been built and its depth estimation accuracy was empirically assessed to around 7 cm from 1 to 5 m range. The prototype was able to locate fine structures (wires). Note that the proposed approach could be straightforwardly extended to different requirements as those used here for a small UAV. Further works involve comparisons of the codesigned chromatic lens with existing chromatic lens cameras as in [21] and co-operation of a chromatic lens with a coded aperture.

References

[1] Y. Bando, B. Chen, and T. Nishita. Extracting depth and matte using a color-filtered aperture. *ACM TOG*, 27(5), 2008.

[2] A. Chakrabarti and T. Zickler. Depth and deblurring from a spectrally varying depth of field. In *ECCV*, 2012.

[3] A. Chakrabarti, T. Zickler, and W. Freeman. Analyzing spatially-varying blur. In *CVPR*, 2010.

[4] L. Condat. A generic variational approach for demosaicking from an arbitrary color filter array. In *ICIP*, 2009.

[5] O. Cossairt and S. Nayar. Spectral focal sweep: Extended depth of field from chromatic aberrations. In *ICCP*, 2010.

[6] F. Diaz, F. Goudail, B. Loiseaux, and J. Huignard. Increase in depth of field taking into account deconvolution by optimization of pupil mask. *Optics Letters*, 34(19), 2009.

[7] B. Freedman, A. Shpunt, M. Meir, and A. Yoel. Depth mapping using projected patterns, US patent 20100118123., 2010.

[8] J. Garcia, J. Sánchez, X. Orriols, and X. Binefa. Chromatic aberration and depth extraction. In *ICPR*, 2000.

[9] F. Guichard, H. Nguyen, R. Tessières, M. Pyanet, I. Tarchouna, and F. Cao. Extended depth-of-field using sharpness transport across color channels. In *Proc. of SPIE*, 2009.

[10] J. Idier, editor. *Bayesian approach to inverse problems*. ISTE Ltd and John Wiley & Sons Inc, apr. 2008.

[11] A. Levin, R. Fergus, F. Durand, and W. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM TOG*, 26(3), 2007.

[12] A. Levin, Y. Weiss, F. Durand, and W. Freeman. Understanding and evaluating blind deconvolution algorithms. In *CVPR*, 2009.

[13] M. Martinello and P. Favaro. Single image blind deconvolution with higher-order texture statistics. *Video Proc. and Comp. Video*, 2011.

[14] H. Nagahara, C. Zhou, T. Watanabe, H. Ishiguro, and S. Nayar. Programmable aperture camera using lcos. In *ECCV*, 2010.

[15] H. Nair and C. Stewart. Robust focus ranging. In *CVPR*, 1992.

[16] 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.

[17] A. Pentland. A new sense for depth of field. *IEEE Trans. on PAMI*, 4, 1987.

[18] B. Porat and B. Friedlander. Computation of the exact information matrix of gaussian time series with stationary random components. *IEEE Trans. on ASSP*, 34(1), 1986.

[19] A. Rajagopalan and S. Chaudhuri. Performance analysis of maximum likelihood estimator for recovery of depth from defocused images and optimal selection of camera parameters. In *IJCV*, volume 30, 1998.

[20] D. Stork and M. Robinson. Theoretical foundations for joint digital-optical analysis of electro-optical imaging systems. *Applied Optics*, 47(10):64–75, 2008.

[21] P. Trouvé, F. Champagnat, G. L. Besnerais, G. Druart, and J. Idier. Chromatic depth from defocus : a theoretical and experimental performance study. In *COSI*, 2012.

[22] P. Trouvé, F. Champagnat, G. Le Besnerais, and J. Idier. Single image local blur identification. In *ICIP*, 2011.

[23] C. Zhou and S. Nayar. What are good apertures for defocus deblurring? In *ICCP*, 2009.