

# Estimation de profondeur à partir du flou de défocalisation d'un imageur chromatique

Pauline TROUVÉ<sup>1</sup>, Frédéric CHAMPAGNAT<sup>1</sup>, Guy LE BESNERAIS<sup>1</sup>, Jérôme IDIER<sup>2</sup>

<sup>1</sup>Onera-The French Aerospace Lab, Chemin de la Hunière, FR-91123 Palaiseau, France

<sup>2</sup>LUNAM Université, IRCCyN (UMR CNRS 6597) BP 92101, 1 rue de la Noé 44321 Nantes Cedex 3, France  
pauline.trouve@onera.fr, frederic.champagnat@onera.fr, guy.le\_besnerais@onera.fr,  
jerome.idier@irccyn.ec-nantes.fr

**Résumé** – Nous présentons une nouvelle méthode d'estimation de profondeur monovoie passive reposant sur l'utilisation d'une optique possédant du chromatisme longitudinal, combinée à un détecteur couleur. Cette caméra produit, en une seule acquisition, trois images RVB possédant un flou de défocalisation variable spectralement, ce qui permet de faciliter l'estimation de la profondeur par une approche de type *depth from defocus* (DFD). Nous avons développé un algorithme de DFD original dédié à l'estimation de profondeur à partir d'images produites par une caméra chromatique. Nous présentons ici les résultats expérimentaux obtenus avec cet algorithme sur les images produites par un prototype de caméra chromatique.

**Abstract** – In this paper we present a new passive depth estimation method based on the joint utilisation of a lens with chromatic aberration and a color sensor. Such a camera produces in a single snapshot three RGB images having spectrally varying defocus blur, which eases depth estimation using a *Depth from defocus* approach. We have developed an original DFD algorithm dedicated to the processing of images recorded by a chromatic camera. Here we present experimental depth estimation results obtained with a prototype of chromatic camera and the proposed DFD algorithm.

## 1 Introduction

L'estimation de profondeur par les techniques de *Depth From Defocus* (DFD) repose sur le lien entre le niveau de flou de l'image d'un objet et sa position dans l'espace. En effet, lorsqu'un point source est placé dans le plan de mise au point d'un imageur, son image, appelée fonction d'étalement du point (FEP), est une tache d'Airy de largeur souvent négligeable devant la taille du pixel du détecteur. Mais lorsque cet objet ponctuel est placé en dehors du plan de mise au point, la FEP est une tache de défocalisation de largeur  $\epsilon$ , de taille supérieure au pixel du détecteur et responsable du flou observé dans l'image. La largeur théorique  $\epsilon$  s'exprime de la manière suivante :

$$\epsilon = Ds \left| \frac{1}{f} - \frac{1}{p} - \frac{1}{d_{det}} \right|, \quad (1)$$

où  $f$  est la distance focale de l'optique,  $D$  son diamètre, et avec  $d_{det}$  et  $p$  les distances entre l'optique et respectivement le détecteur et l'objet, comme illustré à la figure 1. D'après (1), connaissant la distance focale et la position du détecteur, l'estimation de la largeur de la FEP permet d'estimer la profondeur.

Les premières méthodes de DFD utilisent différentes images de la même scène avec plusieurs réglages du même imageur [1]. Cependant la scène doit alors être statique pendant les acquisitions. D'autres techniques n'utilisent

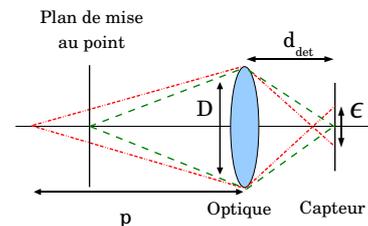


FIGURE 1 – Principe de DFD. En vert l'objet ponctuel est dans le plan de mise au point de la caméra, en rouge il est défocalisé.

qu'une seule image [2, 3, 4, 5], l'acquisition est alors moins contraignante, mais il faut estimer localement le niveau de flou, autrement dit la FEP, à partir d'une seule acquisition alors que la scène est inconnue. Plusieurs références utilisent un apprentissage sur une base de scènes [3, 4] ou exploitent une modélisation contraignante sur la scène et sur la forme de la FEP [2]. Dans [5], nous avons développé un algorithme d'estimation de profondeur non supervisé et indépendant de la forme de la FEP, qui sélectionne une FEP dans une famille de FEP potentielles à l'aide d'une modélisation simple des gradients de la scène. Nous avons montré la possibilité de faire de la DFD avec cet algorithme en utilisant par exemple la pupille codée proposée dans [3]. Cependant, si les techniques de DFD mono-image sont plus adaptées au cas de scènes dynamiques que les

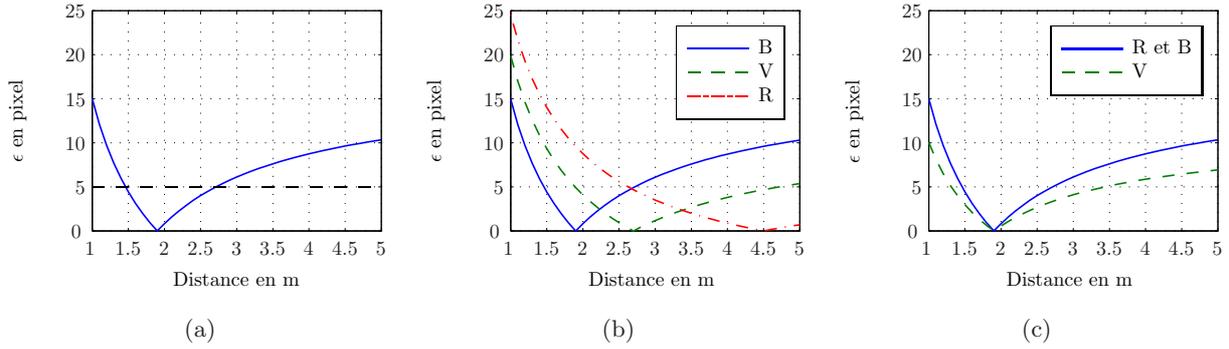


FIGURE 2 – (a) Variation de la largeur de la FEP en fonction de la profondeur d’un objet, pour un imageur de focale 25 mm, de nombre d’ouverture 4 et un plan de mise au point à 2 m. (b) et (c) : effets sur le flou de défocalisation des canaux RVB de l’utilisation respectivement d’une lentille chromatique et de la pupille chromatique proposée dans [9].

techniques multi-images, elles ont cependant deux inconvénients. Tout d’abord, comme illustré à la figure 2(a), il existe une ambiguïté entre le niveau de flou et la profondeur de part et d’autre du plan de mise au point. Ensuite, il n’est pas possible d’estimer précisément la profondeur d’un objet lorsqu’il est placé dans l’intervalle de distances défini par la profondeur de champ de l’imageur, dans laquelle la FEP est inférieure au pixel. En effet, dans cet intervalle de distances aucune variation de flou n’est observable. Pour résoudre ces deux difficultés nous proposons d’utiliser l’aberration chromatique longitudinale.

## 1.1 Principe

Le chromatisme d’une optique se traduit par la variation de la position du plan de mise au point avec la longueur d’onde. La combinaison d’une optique dont le chromatisme longitudinal n’est pas corrigé et d’un détecteur couleur, produisant trois images RVB, permet donc d’avoir trois images avec des degrés de flou différents. Ainsi, comme l’illustre la figure 2(b), pour chaque profondeur il existe un triplet unique de FEP associées aux trois canaux RVB et l’ambiguïté du DFD mono-image disparaît. De plus, lorsqu’un objet est dans la région de profondeur de champ d’un des canaux, sa position peut être estimée grâce à la variation de la FEP dans les deux autres canaux, car chaque canal a sa propre région de profondeur de champ. Ceci permet d’éliminer la zone de forte incertitude autour du plan de mise au point qui existe en DFD mono-image. Enfin, l’acquisition des trois canaux RVB est instantanée, l’estimation de profondeur n’impose donc pas de contrainte de scène statique.

## 1.2 État de l’art

Le principe de DFD à l’aide du chromatisme a été proposée dans [6], cependant la mise en œuvre algorithmique permet de traiter uniquement des bords de plage noirs et blancs. Par ailleurs, le chromatisme est utilisé pour étendre la profondeur de champ d’une image dans [7, 8] et

dans [7] la possibilité d’estimer la profondeur n’est que mentionnée. À notre connaissance, il n’existe donc pas de réalisation concrète de DFD à partir d’un imageur chromatique pour des images réelles. Notons qu’une autre approche consiste à introduire des filtres dans le diaphragme de l’objectif afin d’avoir une ouverture différente, donc un flou différent, pour les trois canaux RVB [9]. Cependant, puisque ce système n’a qu’un seul plan de mise au point, comme le montre la figure 2(c), ce dernier possède les mêmes inconvénients que la DFD mono-image, liés à l’ambiguïté de part et d’autre du plan de mise au point et à la profondeur de champ.

## 2 Algorithme de DFD chromatique

Une des difficultés du DFD à l’aide du chromatisme est que les trois images traitées sont issues de scènes qui ne sont que partiellement corrélées. Ainsi, les techniques de DFD multi-images ne peuvent pas s’appliquer car elles supposent que la scène est identique dans chaque image. C’est pourquoi, nous avons développé un nouvel algorithme appelé C-DFD (pour *Chromatic Depth from Defocus*) qui étend l’approche que nous avons proposée pour le DFD mono-image dans [5] au traitement d’images couleurs produites par un imageur chromatique. Dans [5], nous utilisons un formalisme bayésien pour sélectionner localement une FEP parmi un jeu de FEP potentielles, obtenues par calibrage. La sélection d’une FEP se fait d’après un critère dérivé d’un maximum de vraisemblance marginale dont l’expression analytique est obtenue avec un modèle de scènes à gradients gaussiens et isotropes et un modèle de bruit blanc gaussien. Nous rejetons par pré-traitement les régions homogènes où la FEP, donc la profondeur, est inobservable. Dans le cas d’acquisitions d’images couleurs par une caméra chromatique, le modèle de formation d’images peut s’écrire :

$$\mathbf{Y} = H_C \mathbf{X} + \mathbf{B}, \quad (2)$$

$$\text{avec } H_C = \begin{bmatrix} H_R(p) & 0 & 0 \\ 0 & H_V(p) & 0 \\ 0 & 0 & H_B(p) \end{bmatrix}, \quad (3)$$

avec  $\mathbf{Y} = [\mathbf{y}_R^t \ \mathbf{y}_V^t \ \mathbf{y}_B^t]^t$  et  $\mathbf{X} = [\mathbf{x}_R^t \ \mathbf{x}_V^t \ \mathbf{x}_B^t]^t$ , où  $\mathbf{y}_R, \mathbf{y}_G$  et  $\mathbf{y}_B$  (resp.  $\mathbf{x}_R, \mathbf{x}_G$  et  $\mathbf{x}_B$ ) représentent la concaténation des  $M$  (resp.  $N$ ) pixels des trois fenêtres des images (resp. des scènes) RVB.  $\mathbf{B}$  correspond au bruit d'acquisition modélisé par un processus aléatoire gaussien centré de variance  $\sigma_b^2$ . Chaque matrice  $H_c(p)$  est une matrice de convolution de taille  $N \times M$  relative au canal  $c$  qui dépend de la profondeur de la scène  $p$ . Nous proposons de sélectionner un triplet de FEP parmi un jeu de triplets potentiels à l'aide d'un critère de maximum de vraisemblance marginale, mais par rapport à [5], nous modélisons la corrélation entre les canaux RVB à l'aide d'une décomposition en luminance et chrominance ( $L, C_1, C_2$ ) :

$$\begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_V \\ \mathbf{x}_B \end{bmatrix} = T \otimes \mathbf{I}_{M,M} \mathbf{X}^{LC} = T \otimes \mathbf{I}_{M,M} \begin{bmatrix} \mathbf{x}_l \\ \mathbf{x}_{c_1} \\ \mathbf{x}_{c_2} \end{bmatrix}, \quad (4)$$

$$\text{où } T = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{bmatrix}, \quad (5)$$

où  $\otimes$  correspond au produit de Kronecker et  $\mathbf{I}_{M,M}$  est la matrice identité de taille  $M \times M$ . Ainsi, (2) devient :

$$\mathbf{Y} = H_C^c(p) \mathbf{X}^{LC} + \mathbf{B} \quad (6)$$

$$H_C^c(d) = H_C(p) T \otimes \mathbf{I}_{M,M}. \quad (7)$$

Nous modélisons ensuite la densité de probabilité de chaque composante à l'aide d'un a priori gaussien isotrope, décorrélié des deux autres composantes ce qui donne :

$$p(\mathbf{X}^{LC}, \sigma_x^2, \mu) \propto \exp\left(-\frac{\|D_C(\mu) \mathbf{X}^{LC}\|^2}{2\sigma_x^2}\right) \quad (8)$$

$$\text{avec } D_C(\mu) = \begin{bmatrix} \sqrt{\mu}D & 0 & 0 \\ 0 & D & 0 \\ 0 & 0 & D \end{bmatrix}. \quad (9)$$

$D$  correspond à la concaténation verticale des matrices de convolution associées aux vecteurs  $[-1 \ 1]$  et  $[-1 \ 1]^t$  et  $\mu$  modélise la différence de variance entre les composantes ( $L, C_1, C_2$ ). Ce type d'a priori est dit impropre car  $D\mathbf{1} = \mathbf{0}$ . Cependant [10] montre qu'il est possible d'exprimer une vraisemblance marginale des données sous la forme :

$$p(\mathbf{Y} | H_C^c(p), \sigma_b^2, \alpha) \propto |\sigma_b^2|^{-\frac{1}{2}} P(\alpha, p) e^{-\frac{\mathbf{Y}^t P(\alpha, p) \mathbf{Y}}{2\sigma_b^2}}, \quad (10)$$

avec  $P(\alpha, p) = I_{N,N} -$

$$H_C^c(p)(H_C^c(p)^t H_C^c(p) + \alpha D_C^t D_C)^{-1} H_C^c(p)^t.$$

$\alpha = \sigma_b^2 / \sigma_x^2$  est un terme de régularisation permettant de tenir compte de la variation locale du rapport signal à

bruit. Comme dans [5], on peut montrer que maximiser la vraisemblance marginale revient à minimiser le terme :

$$GL_C(p, \alpha) = \frac{\mathbf{Y}^t P(\alpha, p) \mathbf{Y}}{|P(\alpha, p)|_+^{1/(3N-3)}}, \quad (11)$$

$P(\alpha, p)$  a trois valeurs propres nulles, une pour chaque composante luminance et chrominances. Comme dans [11], nous fixons  $\mu$  à 0.04. L'algorithme C-DFD consiste alors à définir une fenêtre de l'image et à sélectionner le triplet qui minimise le critère  $GL_C$ , à partir d'une famille de triplet de flous potentiels.

## 3 Validation expérimentale

### 3.1 Objectif chromatique

Un objectif chromatique de distance focale 25 mm et de nombre d'ouverture 4 a été conçu dans le cadre d'une collaboration IOGS-ONERA. Cet objectif a un chromatisme longitudinal de 200  $\mu\text{m}$  et il est associé à un détecteur couleur possédant des pixels de 3.45  $\mu\text{m}$ , afin d'obtenir une caméra chromatique. Les plans de mise au point des canaux RVB sont placés respectivement à 4.5, 2.7 et 1.9 m, comme illustré à la figure 2(b).

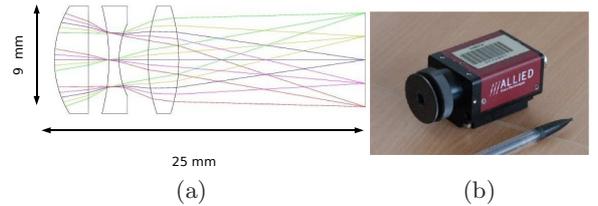


FIGURE 3 – (a) Schéma des lentilles de l'objectif chromatique. (b) Caméra chromatique.

### 3.2 Mesure de profondeur sur l'axe

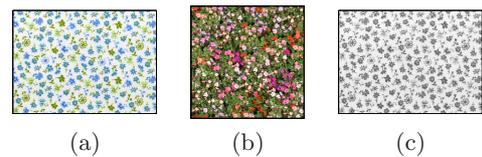


FIGURE 4 – Cibles

Les FEP de la caméra chromatique sont calibrées entre 1 et 5 m avec un pas de 5 cm. La figure 5 présente les résultats d'estimation de profondeur obtenus sur l'axe pour trois scènes présentées à la figure 4 déplacées successivement entre 1 et 4.5 m. Les fenêtres de chaque canal RVB traitées sont de taille  $21 \times 21$  pixels. La profondeur est estimée à l'aide de l'algorithme C-DFD à l'intérieur d'une région centrée de l'image et les résultats présentés correspondent à la moyenne et à l'écart-type calculés sur 100 résultats d'estimation de profondeur. Les résultats obtenus sont comparés à une vérité terrain donnée par la Kinect,

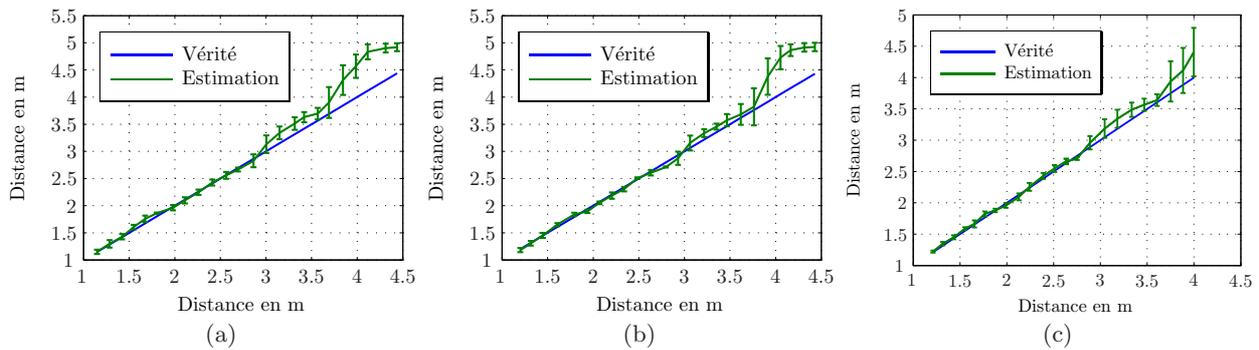


FIGURE 5 – Résultats de l’algorithme C-DFD sur l’axe pour les trois cibles de la figure 4.

qui est un capteur 3D actif de précision inférieure au cm. Pour les trois cibles, l’estimation de profondeur a un biais négligeable et un écart-type inférieur à 5 cm entre 1 et 3 m, ce qui illustre la robustesse de l’estimation de profondeur vis-à-vis de la couleur de la scène. Mais pour des distances supérieures, l’estimation de profondeur se dégrade.

### 3.3 Carte de profondeurs

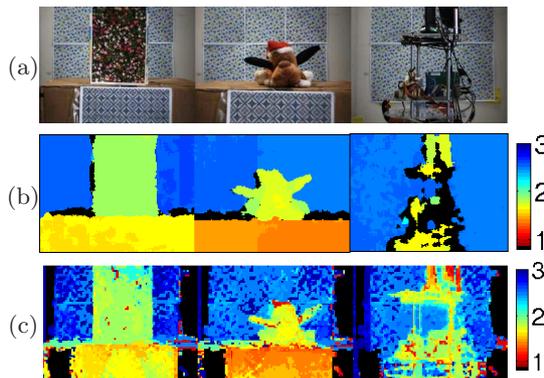


FIGURE 6 – (a) Scènes acquises avec l’imageur chromatique. (b) à (d) Cartes de profondeurs en m obtenues resp. avec la Kinect, avec l’algorithme C-DFD. La couleur noire correspond aux régions homogènes non informatives.

À partir des FEP calibrées, l’estimation de profondeur peut également être réalisée sur des images complètes. Trois exemples de carte de profondeur obtenues avec l’algorithme C-DFD sont présentées à la figure 6 en comparaison avec celles données par une Kinect. Les valeurs de profondeur estimées avec l’imageur chromatique sont cohérentes avec celles de la Kinect, bien que plus bruitées. Notons cependant que le résultat de la Kinect est fortement régularisé et une régularisation de nos résultats permettrait d’éliminer les valeurs aberrantes et de propager l’information dans les régions homogènes non renseignées. Enfin, dans le troisième exemple, la Kinect donne des résultats aberrants, car la mire projetée se réfléchit mal sur cet objet très linéique, alors que nous observons correctement la structure complexe de l’objet sur notre résultat.

## 4 Conclusion et perspectives

Nous avons développé une méthode d’estimation de profondeur à partir d’images produites par un imageur chromatique et validé cette méthode expérimentalement. Une perspective de ce travail est de développer un imageur produisant simultanément une carte de profondeur et une image couleur de bonne qualité. Pour cela, il est nécessaire de restaurer l’image couleur car le chromatisme dégrade légèrement la qualité optique de l’imageur, par exemple à l’aide d’un traitement du type de transfert de hautes fréquences [7]. Une autre perspective de ces travaux est l’optimisation conjointe des paramètres de la caméra chromatique et des traitements.

### Remerciements

Nous remercions T. Avignon, L. Jacobowicz et J. Sabater de l’IOGS pour la réalisation de l’imageur chromatique.

### Références

- [1] A. Pentland, *A new sense for depth of field*, *Patt. An. M Int.*, 1987.
- [2] S. Zhuo et T. Sim : *Defocus map estimation from a single image*, *Pattern Recognition*, **44**, 2011.
- [3] A. Levin et al. *Image and depth from a conventional camera with a coded aperture*, *ACM Trans. Graph.*, **26**, 2007.
- [4] M. Martinello et P. Favaro, *Single image blind deconvolution with higher-order texture statistics*, *VPCV*, 2011.
- [5] P. Trouvé et al. *Single image local blur identification*, *ICIP*, 2011.
- [6] J. Garcia et al. *Chromatic aberration and depth extraction*, *ICPR*, **1**, 2000.
- [7] F. Guichard et al., *Extended depth-of-field using sharpness transport across color channels*, *SPIE*. **7250**, 2009.
- [8] O. Cossairt et S. Nayar, *Spectral Focal Sweep : Extended Depth of Field from Chromatic Aberrations*, *ICCP*, 2010.
- [9] A. Chakrabarti et T. Zickler, *Depth and deblurring from a spectrally varying depth of field*, *ECCV*, 2012.
- [10] F. Champagnat *Inference with Gaussian improper distributions*, *Rapport Interne Onera n° RT 5/14983 DTIM*, 2012.
- [11] L. Condat, *A generic variational approach for demosaicking from an arbitrary color filter array*, *ICIP*, 2009.