

Amélioration de l'évaluation de la profondeur à partir d'un couple d'images nette et floue

Christophe SIMON¹, Frédérique BICKING¹, Thierry SIMON², Ahmed GOUDALI²

¹ Université Henri Poincaré Nancy 1, Ecole Supérieure des Sciences et Technologies de l'Ingénieur de Nancy

2 Rue Jean Lamour, 54519 Vandœuvre-lès-Nancy Cedex

² Equipe Perceval, IUT de Figeac, Université Toulouse 2

Avenue de Nayrac, 46100 Figeac

Christophe.Simon@esstin.uhp-nancy.fr

Frederique.Bicking@esstin.uhp-nancy.fr, Simon.IUT.Figeac@wanadoo.fr

Résumé - Nous présentons dans cet article la généralisation d'une méthode d'estimation locale de la profondeur sur les contours utilisant un couple d'images nette et floue. Après avoir positionné cette méthode de *depth from defocus*, initialement développée par T. Simon ([9]), nous définissons les relations théoriques mettant en œuvre cette généralisation. Nous apportons une solution à la sensibilité aux variations locales de luminance et montrons les contraintes théoriques de la méthode. Enfin, l'application de la méthode à des images de synthèse et à une scène réelle est présentée.

Abstract – This article deals with the generalization of a local depth estimation method using sharp edges and blurred edges. This Depth from Defocus method, developed by T. Simon ([9]), is explained and the theoretical relations are defined. Improvements concerning the generalization, the noise sensitivity on the depth estimation are developed and application conditions are exposed. Some results on synthetic and real images are presented.

1. Introduction

La profondeur, distance entre la surface visible des objets dans une scène et le capteur de la caméra, est une donnée nécessaire pour le calcul des coordonnées des points appartenant à cette surface dans un espace tridimensionnel de référence. De nombreuses méthodes ont été développées pour obtenir ces coordonnées. Certaines techniques, nommées *Depth From Focus* ou DFF ([2], [4], [5], [12]), recherchent des zones de netteté dans l'image et les lient à la profondeur par les relations classiques de l'optique géométrique. Elles présentent l'inconvénient majeur d'un temps prohibitif d'obtention d'une carte des profondeurs. D'autres méthodes appelées *Depth From Defocus* ou DFD ([3], [7], [8], [10], [13]), exploitent le flou optique perceptible sur les zones non homogènes de l'image correspondant à des contours ou à des textures ([1]) en utilisant au moins deux images d'une même scène acquises avec des réglages différents des paramètres du système optique. Notre technique utilise uniquement un couple d'images nette et floue d'une même scène, acquises selon le même point de vue. L'importance du flou optique, caractéristique de la profondeur, est observée sur les discontinuités de luminance dans l'image floue.

Notre article rappelle dans un premier temps le lien entre la profondeur et le flou optique et précise le modèle des discontinuités de luminance utilisé. Les relations issues de l'algorithme initialement posé par T. Simon ([9]) sont présentées. Enfin, nous proposons des améliorations reposant sur une généralisation du calcul du rapport des modules des gradients des images nette et floue au cas des contours épais

et également à la totalité des points de la discontinuité dans la direction de son gradient. Nous montrons aussi qu'il existe des contraintes d'application de la méthode initiale en terme d'orientation. Ces améliorations conduisent à une meilleure précision et à une meilleure robustesse de l'évaluation de la profondeur.

2. Le flou optique comme information de profondeur

Dans les méthodes relevant du *depth from defocus*, on cherche à établir une relation entre la profondeur, les paramètres de la caméra et le flou optique dans les images. L'objet de ce paragraphe est de montrer comment s'obtient cette relation.

Pour assurer correctement la formation d'une image sur un plan, plusieurs lentilles forment généralement les objectifs photographiques. Pour des raisons de simplification, nous assimilons l'objectif de la caméra à une lentille convergente mince unique et restons dans les hypothèses de Gauss.

A un plan image à une distance fixée et pour une scène contenant plusieurs plans en profondeur, un seul plan donne une image nette, dite focalisée. Les points objets appartenant aux autres plans formeront des taches plus ou moins floues en fonction de leur distance par rapport au plan image. Pour une distance donnée, les relations de l'optique géométrique permettent d'obtenir la distance à un objet sur le plan objet correspondant à une profondeur dans la scène.

Le flou optique est lié à la fonction de transfert du système optique ([14]) dans le domaine spatial représentée par sa

réponse impulsionnelle dite *Point Spread Function* (PSF). Ainsi, une relation de convolution (1) s'établit entre l'image nette $I_n(i, j)$, l'image floue $I_f(i, j)$ et la $PSF(i, j)$ où i et j représentent les coordonnées d'un pixel et \otimes_{2D} l'opérateur de convolution bidimensionnelle :

$$I_f(i, j) = I_n(i, j) \otimes_{2D} PSF(i, j) \quad (1)$$

La PSF dépend autant des propriétés des matériaux de l'optique que de la forme géométrique de la lentille et des paramètres de prise de vue. Plusieurs modèles ont été proposés et nous avons retenu le modèle de distribution de luminance donné par une gaussienne bidimensionnelle avec un paramètre d'étalement σ_{s_o} correspondant à une profondeur s_o . Ce modèle est donné par la relation (2) :

$$PSF_{s_o}(i, j) = \frac{1}{2\pi\sigma_{s_o}^2} \exp\left(-\frac{i^2 + j^2}{2\sigma_{s_o}^2}\right) \quad (2)$$

La relation entre le paramètre d'étalement σ_{s_o} et la profondeur peut être obtenue par la relation (3) où c et m sont des constantes représentant un réglage spécifique de la caméra obtenu au cours d'une phase de calibrage [11] :

$$\frac{1}{s_o} = \frac{\sigma_{s_o}}{m} - \frac{c}{m} \quad (3)$$

3. Méthode initiale pour l'estimation du paramètre de flou optique

Les approches classiques d'estimation de la profondeur exploitent soit le contenu spatial de l'image par les caractéristiques géométriques ([6], [7]) ou la forme des objets dans la scène ([2]), soit le contenu fréquentiel par une analyse de Fourier ([10]).

On effectue l'acquisition d'une image nette en tout point avec un iris fermé qui fournit facilement la position des contours grâce à un opérateur de gradient lisse comme celui de Prewitt. En faisant l'acquisition d'une seconde image, floue (iris ouvert), le contenu informationnel de la scène est isolé. En établissant le rapport des modules des gradients des images nette et floue de la même scène qui permet de donner la même importance à chaque contour, il est possible d'identifier le paramètre d'étalement σ_{s_o} et ainsi d'estimer la profondeur par la relation (3). Ce rapport s'écrit selon la relation (4) où $|\nabla_p I_n(i, j)|$ et $|\nabla_p I_f(i, j)|$ sont les modules des gradients de Prewitt des images nette et floue.

$$R(i, j) = \frac{|\nabla_p I_n(i, j)|}{|\nabla_p I_f(i, j)|} \quad (4)$$

Considérons que l'image nette $I_n(i, j)$ est formée de discontinuités en rampe d'amplitude $(b-a)$, de longueur ε selon une orientation unique θ et que dans l'orientation perpendiculaire à θ , l'image nette ne présente pas de variations. La fonction de niveaux de gris, $c_n(x)$, de cette discontinuité s'écrit selon la relation (5) où x , gradué en pixels, représente une ligne ou une colonne de l'image :

$$c_n(x) = \begin{cases} a & x < x_0 \\ a + \left(\frac{b-a}{\varepsilon}\right)(x-x_0) & x_0 \leq x \leq x_0 + \varepsilon \\ b & x > x_0 + \varepsilon \end{cases} \quad (5)$$

Si $\theta = 0 + n\pi$, $I_n(i-1, j) = I_n(i, j) = I_n(i+1, j)$, alors le module du gradient de Prewitt de l'image nette peut s'écrire selon la relation (6) :

$$|\nabla_p I_n(i, j)| = \left[\left[\begin{array}{ccc} -1 & 0 & 1 \\ \frac{1}{3} & -1 & 0 & 1 \\ -1 & 0 & 1 \end{array} \right] \otimes_{2D} I_n(i, j) \right]^2 + \left[\begin{array}{ccc} -1 & -1 & -1 \\ \frac{1}{3} & 0 & 0 & 0 \\ 1 & 1 & 1 \end{array} \right] \otimes_{2D} I_n(i, j) \right]^2 \quad (6)$$

$$= \left| \begin{array}{ccc} -1 & 0 & 1 \end{array} \right| \otimes_{1D} I_n(i, j) \quad \forall i$$

Si l'orientation de la discontinuité est $\theta = \pi/2 + n\pi$, on aura $I_n(i, j-1) = I_n(i, j) = I_n(i, j+1)$ et un raisonnement similaire est mené sur ce contour pour tout j .

A partir du modèle de discontinuité en rampe dans la direction θ , de la constance de l'image nette dans la direction perpendiculaire à θ et de la relation (1), l'expression de l'image floue devient :

$$I_f(i, j) = I_n(i, j) \otimes_{1D} \frac{1}{\sqrt{2\pi\sigma_{s_o}}} \exp\left(-\frac{i^2}{2\sigma_{s_o}^2}\right) = I_n(i, j) \otimes_{1D} PSF(i) \quad \forall j$$

$$I_f(i, j) = I_n(i, j) \otimes_{1D} \frac{1}{\sqrt{2\pi\sigma_{s_o}}} \exp\left(-\frac{j^2}{2\sigma_{s_o}^2}\right) = I_n(i, j) \otimes_{1D} PSF(j) \quad \forall i$$

Le module du gradient de l'image floue s'écrit selon (8) :

$$|\nabla_p I_f(i, j)| = \left[\left[\begin{array}{ccc} -1 & 0 & 1 \\ \frac{1}{3} & -1 & 0 & 1 \\ -1 & 0 & 1 \end{array} \right] \otimes_{2D} I_f(i, j) \right]^2 + \left[\begin{array}{ccc} -1 & -1 & -1 \\ \frac{1}{3} & 0 & 0 & 0 \\ 1 & 1 & 1 \end{array} \right] \otimes_{2D} I_f(i, j) \right]^2 \quad (8)$$

$$= \left| \begin{array}{ccc} -1 & 0 & 1 \end{array} \right| \otimes_{1D} I_f(i, j) \quad \forall j$$

Cette correspondance 2D-1D obtenue dans le cas particulier des images nettes que nous avons décrites permet de poursuivre l'étude directement sur le profil des contours net $c_n(x)$ et flou $c_f(x)$ avec la relation (9) :

$$c_f(x) = c_n(x) \otimes_{1D} PSF(x) \quad (9)$$

Les modules des gradients des images nette et floue peuvent alors s'écrire :

$$\begin{aligned} |\nabla c_n(x)| &= \left| c_n(x) \otimes_{1D} [-1 \ 0 \ 1] \right| \\ |\nabla c_f(x)| &= \left| \nabla c_n(x) \otimes_{1D} PSF(x) \right| \end{aligned} \quad (10)$$

Le rapport des modules des gradients des images nette et floue, donné par la relation (4), prend des valeurs exploitables en fonction des valeurs de la longueur de la rampe ε . Le rapport exprimé aux seuls deux points x_0 et $x_0 + \varepsilon$ avec $R(x_0) = R(x_0 + \varepsilon)$ présente l'une des formes analytiques simples suivantes :

$$\begin{aligned} \text{Pour } \varepsilon = 1, R(x_0) &= \frac{\sigma_{s_0} \sqrt{2\pi}}{1 + \exp\left(-\frac{1}{2\sigma_{s_0}^2}\right)} \\ \text{Pour } \varepsilon = 2, R(x_0) &= \frac{\sigma_{s_0} \sqrt{2\pi}}{1 + 2 \exp\left(-\frac{1}{2\sigma_{s_0}^2}\right) + \exp\left(-\frac{4}{2\sigma_{s_0}^2}\right)} \end{aligned} \quad (11)$$

Pour $\varepsilon \geq 3$,

$$R(x_0) = \frac{\sigma_{s_0} \sqrt{2\pi}}{1 + 2 \exp\left(-\frac{1}{2\sigma_{s_0}^2}\right) + 2 \exp\left(-\frac{4}{2\sigma_{s_0}^2}\right) + \exp\left(-\frac{9}{2\sigma_{s_0}^2}\right)}$$

Ces relations, liant le paramètre d'étalement σ_{s_0} à la valeur du rapport des modules des gradients, sont résolues numériquement. Le rapport des modules des gradients au point x_0 permet de trouver une valeur de σ_{s_0} et d'obtenir une profondeur s_0 grâce à la relation (3).

4. Améliorations de la méthode

Le premier point que nous relevons sur cette méthode est l'invalidité de l'expression du rapport pour des longueurs de profils de discontinuités ε strictement supérieures à 3. En étudiant les gradients des profils des discontinuités avec des valeurs de ε plus importantes, on obtient une expression plus générale du module du gradient de l'image floue et donc du rapport des modules des gradients des images nette et floue. L'étude de la convolution du gradient du contour net en rampe par la PSF monodimensionnelle du contour, nous permet d'exprimer le gradient flou en tout point x entre x_0 et $x_0 + \varepsilon$ du contour par la relation (12) :

$$\forall \varepsilon \in N_*^+ \quad \nabla c_f(x) = \left(\frac{b-a}{\varepsilon} \left[\frac{PSF(x-x_0) + PSF(x-(x_0+\varepsilon))}{1 + 2 \sum_{u=1}^{\varepsilon-1} PSF(x-(x_0+u))} \right] \right) \quad (12)$$

Le rapport des modules des gradients des images nette et floue s'écrit alors selon la relation (13) :

$$R(x) = \frac{1}{\left(\frac{PSF(x-x_0) + PSF(x-(x_0+\varepsilon))}{1 + 2 \sum_{u=1}^{\varepsilon-1} PSF(x-(x_0+u))} \right)} \quad (13)$$

pour $x = x_0$ et $x = x_0 + \varepsilon$

$$R(x) = \frac{2}{\left(\frac{PSF(x-x_0) + PSF(x-(x_0+\varepsilon))}{1 + 2 \sum_{u=1}^{\varepsilon-1} PSF(x-(x_0+u))} \right)}$$

pour $x_0 < x < x_0 + \varepsilon$

Comme dans la méthode initiale, ces relations sont résolues numériquement au point x_0 pour déterminer la valeur du paramètre d'étalement σ_{s_0} . L'estimation de σ_{s_0} est sensible au bruit sur la valeur des images au point x_0 . Afin d'améliorer la méthode en limitant l'influence du bruit, nous utilisons tous les points de la discontinuité $x \in [x_0, x_0 + \varepsilon]$ pour réaliser $\varepsilon + 1$ estimations. Nous déduisons statistiquement une valeur représentative des estimations obtenues que l'on attribue au point x_0 dans la carte des paramètres d'étalement. La méthode se trouve un peu alourdie en temps de calcul mais l'estimation est globalement meilleure face au bruit dans les images.

Enfin, l'hypothèse sur l'orientation des contours de l'image qui nous a permis d'écrire les relations (6) et (8) est en réalité une contrainte stricte pour utiliser la relation (13). Cette contrainte n'était pas précisée strictement dans le développement initial de la méthode. Dès lors que l'image présente une discontinuité d'orientation $\theta \neq 0 + n\pi/2$, la relation (13) n'est plus valable et son exploitation ne donnera que des valeurs approchées du paramètre d'étalement et donc de la profondeur.

5. Relation flou-éloignement : calibrage

Nous avons opté pour une technique de calibrage classique dite « point par point » qui consiste à placer une discontinuité unidirectionnelle à plusieurs profondeurs. Les couples d'images nettes et floues obtenus sont utilisés pour l'estimation de σ_{s_0} pour les différentes profondeurs s_0 . Par régression, nous obtenons les constantes m et c spécifiques aux réglages de la caméra utilisée. Cette approche, lourde en temps de calcul, est réalisée hors ligne. L'imprécision des mesures, liée aux conditions d'acquisition, est prise en compte lors de l'estimation de m et c .

Cette phase nous a permis d'obtenir, sur une plage de valeurs de la profondeur restreinte (6-17 cm), la relation flou-éloignement suivante : $\hat{s}_0^{-1} = -17.5 \cdot 10^{-3} \sigma_{s_0} + 15.6 \cdot 10^{-1}$.

6. Essais et résultats expérimentaux

Lorsque l'on applique notre méthode sur une forme simple comme cela est montré sur les figures 1a et 1b avec $\sigma_s = 2,5$, l'estimation de σ_s , notée $\hat{\sigma}_s$, présente des résultats corrects. Néanmoins, les estimations sont

surévaluées dès lors que les contours sont proches les uns des autres. En effet, au centre de l'image, l'estimation tient compte de la contribution des contours qui lui sont proches perpendiculairement. Ce phénomène de contribution mutuelle des contours est d'autant moins significatif que σ_s est petit et la distance entre les contours relativement grande.

Sur des images réelles, les contours sont beaucoup moins homogènes que ceux d'images de synthèse, ce qui entraîne une disparité des valeurs estimées de σ_s . Le calibrage est déjà caractéristique de ce phénomène. La méthode proposée est sensible au bruit notamment lorsque l'on calcule σ_s au seul point x_0 . L'estimation de σ_s sur une scène réelle, pour des orientations horizontales et verticales, implique une forte réduction du nombre de points de calcul (figure 2) car le bruit induit des orientations différentes de $n\pi/2$. Par ailleurs, la dispersion des résultats est importante. En utilisant les équations généralisées aux points de contours dans l'intervalle $[x_0, x_0 + \varepsilon]$, même s'il n'y a pas plus de points d'estimation, les résultats présentent moins de dispersion. La contrainte d'orientation reste une contrainte forte sur les conditions d'application de la méthode notamment pour obtenir une carte des profondeurs plus dense mais plus juste.

7. Conclusion et perspectives des travaux

Dans cet article, nous avons généralisé une méthode de perception locale de la profondeur sur les contours en étendant son application à des contours épais. Une seconde généralisation porte sur l'estimation de la profondeur à partir de tous les points du profil du contour, permettant ainsi de réduire l'influence du bruit sur les données images. En outre, nous avons montré que les relations permettant l'estimation de la profondeur n'étaient valables que pour des orientations $\theta = n\pi/2$. Le respect de cette contrainte réduit fortement le nombre de points d'estimation. Elle devra être levée afin d'obtenir une carte plus dense des profondeurs en généralisant l'expression du rapport des modules des gradients des images nette et floue à d'autres directions.

Références

[1] A. Benzougar, M. Khoudair et M. Leard. Formulation du flou pour la perception de la profondeur sur des scènes texturées. *Vision Interface '99*, Trois-Rivières, Canada, 19-21 mai, pp. 111-116, 1999.

[2] T. Darell et K. Wohn. Depth from focus using a pyramid architecture. *Pattern Recognition Letters*, 11(12), pp. 787-796, 1990.

[3] J. Ens et P. Lawrence. A matrix based method for determining depth from focus. Proc. of the *IEEE Computer Society Conf. On Computer Vision and Pattern Recognition*, 1991.

[4] P. Grossman. Depth from focus. *Pattern Recognition Letters*, North Holland, pp. 63-68, 1987.

[5] E. Krotkov. Focusing. *International Journal of Computer Vision*, 1, pp. 223-237, 1987.

[6] A.P. Pentland. Depth of a scene from depth of field. *Proceedings of DARPA image understanding workshop*, Palo Alto, pp. 253-259, 1982.

[7] A.P. Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4), pp. 523-531, 1987.

[8] D.T.Pham et V. Aslants. Depth from defocus using a neural network. *Pattern Recognition*, 32, pp. 715-727, 1999.

[9] T. Simon. *Appréciation de la profondeur depuis des images nette et floue en vision 3D passive*. Thèse de Doctorat de l'Université de Nancy 1, 1991.

[10] M. Subbarao. Parallel depth recovery by changing camera parameters. Proc. of *IEEE Conference on Computer Vision*, Florida, pp. 149-155, 1988.

[11] M. Subbarao. *Determining distance from defocused images of simple objects*. Technical Report No 89-07-20, Computer vision laboratory, Dpt of Electrical Engineering, State University of New-York, Stony Brook, NY 11794-2350, USA, 1989.

[12] M. Subbarao, T. Choi et A. Nikzad. Focusing techniques. *Journal of Optical Engineering*, 32(11), pp. 2824-2836, 1993.

[13] M. Subbarao et G. Surya. Depth from Defocus: a spatial domain approach. *International Journal of Computer Vision*, 13(3), pp. 271-294, 1994.

[14] C.S. Williams et O.A. Becklund. *Introduction to the OTF*. Wiley Series in Pure and Applied optics, 1989.

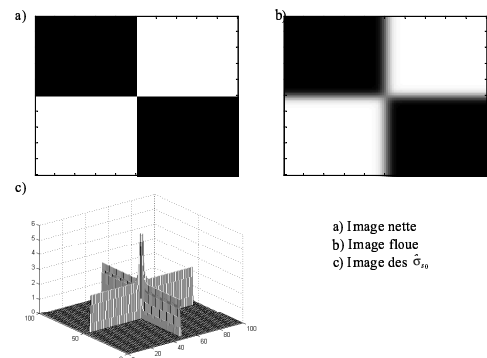


FIG. 1 : illustration de l'incidence des contours voisins sur l'estimation de σ_{s_0} .

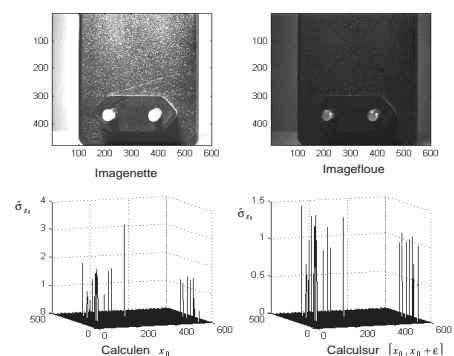


FIG. 2 : application à une scène réelle