

Formulation locale du flou optique pour la perception de la profondeur sur des scènes texturées

Abdeljabar Benzougar, Majdi Khoudeir, Michel Leard
Laboratoire IRCOM-SIC, UMR 6615 Université de Poitiers - UFR Sciences - SP2MI
Boulevard 3 Teleport 2, BP 179, 86960 FUTUROSCOPE CEDEX (France)
E-mail: khoudeir@sic.sp2mi.univ-poitiers.fr

Résumé

Dans le cadre des techniques de la perception de la profondeur par exploitation du flou optique, nous proposons une approche locale particulièrement adaptée aux scènes 3D fortement texturées et sans contours facilement identifiables. Nous développons pour cela une nouvelle formulation locale du flou qui exploite les variations du niveau de gris et les variations locales du flou optique. Cette formulation suppose une réponse impulsionnelle gaussienne de l'optique d'acquisition, mais ne requiert aucune modélisation des images. Un algorithme d'estimation du flou utilisant deux images floues acquises avec des paramètres optiques différents est proposé. Les performances de cette méthode sont évaluées d'abord sur des images texturées présentant des flous spatialement variables générés par synthèse, puis sur des scènes réelles contenant des objets 3D fortement texturés.

1 Introduction

La multitude des environnements réels, rencontrés lors d'applications liés à la vision par ordinateur, nécessite de pouvoir disposer de techniques de stéréovision adaptées aux scènes étudiées. Nous développons dans ce cadre une technique de stéréovision monoculaire par exploitation du flou optique, particulièrement adaptée aux environnements fortement texturés, sans contours facilement identifiables incluant des surfaces de profondeur progressive. Ce type de scène, qui exclut les approches par analyse de contours ou par analyse de régions de l'image, conduit à la mise en place d'une analyse locale de la profondeur et donc du flou optique en chaque point de l'image.

L'estimation du flou est généralement basée sur l'utilisation de deux images acquises avec différents paramètres connus de la caméra. Plusieurs approches ont été développées récemment dans ce cadre. Ainsi la méthode STM (S-Transform Method), proposée par Subbarao [10] et basée sur une modélisation par un polynôme cubique de l'image nette dans le domaine spatial, permet d'aboutir à un estimateur local du flou. Cette méthode est adaptée aux

Abstract

Within the context of techniques of depth perception based on the optical blur, we suggest a local approach which is particularly adapted to highly textured 3D scenes which have no easily identifiable contours. For this, we develop a new local formulation of the blur which exploits variations of grey level and local variations of optical blur. This formulation assumes a Gaussian point spread function of the camera, but does not require a preliminary modelling of the image. An algorithm for the estimation of the blur using two blurred images acquired with different optical parameters is suggested. The performance of this method is first assessed on textured images presenting spatially variable blurring generated by synthesis, and then on real scenes containing 3D objects which are highly textured and which include variable depth.

scènes ne contenant qu'une seule profondeur, mais échoue avec des scènes présentant des objets de profondeurs variables et donc des flous spatialement variables.

Plus récemment Rajagopalan et Shaudhauri [6] [7] ont proposé deux méthodes. La première modélise préalablement le système optique de la caméra par un système dit BSV (Block Shift Variant) dans le but de résoudre le problème de repliement de flou entre pixels voisins. Les images sont ensuite subdivisées en sous-images de taille 13×13 où le flou est supposé spatialement invariant; le modèle BSV permet de tenir compte des interactions entre les flous des sous-images voisines. Cette méthode permet d'obtenir des résultats acceptables, mais reste limitée aux cartes de profondeur de faible résolution spatiale tels que des scènes présentant des variations lentes de profondeur.

La seconde exploite pour la détermination du flou deux représentations fréquentielles de l'image, basée pour la première sur le spectrogramme de l'image, et pour la seconde sur sa distribution de Wigner. Cette méthode est certes plus précise que la méthode BSV mais nécessite des temps de calcul importants.

Dans le cadre d'analyse de scènes fortement texturées, nous avons également développé une technique basée sur la sensibilité des coefficients de corrélation au flou optique [4]. Un algorithme itératif permet d'estimer le flou à partir de la comparaison des coefficients de corrélation évalués sur deux images floues de la même scène dont la moins floue est filtrée à l'aide d'un filtre gaussien. Cette technique suppose que la scène étudiée est composée uniquement d'objets monoprofondeurs.

L'objectif de la méthode que nous présentons ici est de permettre une évaluation ponctuelle du flou et par conséquent de la profondeur en tout point de la scène texturée étudiée. Dans ce cadre, nous proposons dans un premier temps une formulation locale du flou optique qui nous permettra ensuite d'appréhender le problème de mesure d'un flou spatialement variable. Nous décrivons ensuite un algorithme de perception de la profondeur, mettant en oeuvre des traitements effectués dans le domaine spatial et utilisant deux images floues acquises avec des paramètres d'acquisition différents. Les résultats obtenus seront ensuite présentés dans le cas d'images texturées rendues floues artificiellement, puis dans le cas de scènes réelles totalement texturées.

2 Le flou optique, information de profondeur

Dans cette partie, nous rappelons brièvement le processus de formation des images et décrivons la relation entre flou optique et profondeur. Dans ce but, nous adoptons une modélisation classique du système optique de la caméra [2]. Nous représentons ainsi ce système par une lentille mince associée à un diaphragme de diamètre r (Fig. 1).

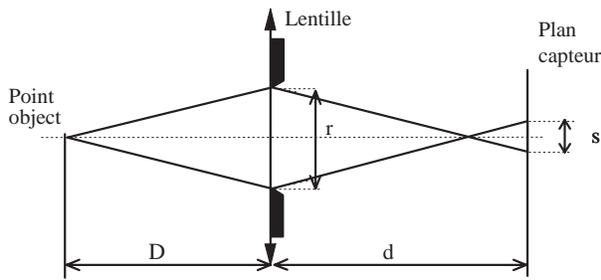


Figure 1: Formation d'une image floue.

Pour un point objet situé à une distance D de la caméra, l'image recueillie sur le plan du capteur, dans le cas d'un réglage défocalisé, est une tache image et non un point image. L'étendue de cette tache varie selon le degré de défocalisation, et sa géométrie est liée à la géométrie du diaphragme utilisé.

Dans le cas d'un diaphragme circulaire, la tache image

sera un disque de diamètre s . Les relations de l'optique géométrique permettent alors d'exprimer s en fonction de la distance focale F , de la distance lentille-capteur d et de la distance D de l'objet, soit $s = rd(\frac{1}{F} - \frac{1}{d} - \frac{1}{D})$. En raison du phénomène de diffraction, le profil de luminance de la tache image n'est pas uniforme. Il est alors nécessaire de modéliser ce profil à l'aide de la réponse impulsionnelle du système optique en vue de la mesure de flou. Le modèle gaussien, habituellement utilisé, est une bonne approximation de la réponse impulsionnelle réelle $h(x,y)$ [3]

$$h(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (1)$$

où l'écart-type σ est proportionnel au diamètre de la tache floue, soit $\sigma = ks$ avec $k > 0$. Le paramètre σ représente alors la mesure du flou et peut s'exprimer en fonction de la distance objet-caméra par

$$\sigma = krd\left(\frac{1}{F} - \frac{1}{d} - \frac{1}{D}\right) \quad (2)$$

Ainsi, l'image défocalisée d'une scène 3D comportant des objets incluant des variations de profondeur (D variable) présente un flou spatialement variable. L'estimation locale du flou permettra alors de déterminer la profondeur de tous les points visibles de la scène.

3 Formulation locale du flou

Considérons l'image nette f d'une scène incluant des objets de profondeurs variables; f sera supposée continue et dérivable. L'image floue g correspondante présente un flou σ variable en tout point (x, y) de l'image, et peut s'exprimer en fonction de l'image nette f et de la réponse impulsionnelle h de la caméra supposée de type gaussien

$$g_\sigma(x, y) = (f * h_\sigma)(x, y) \quad (3)$$

où $*$ est l'opérateur de convolution 2D, et l'indice σ signifie que g présente un flou spatialement variable.

Les variations de g par rapport à σ et son Laplacien 2D s'écrivent alors

$$\begin{aligned} \frac{\partial g_\sigma}{\partial \sigma}(x, y) &= (f * \frac{\partial h_\sigma}{\partial \sigma})(x, y) \\ \nabla^2 g_\sigma(x, y) &= (f * \nabla^2 h_\sigma)(x, y) \end{aligned} \quad (4)$$

avec $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$.

En utilisant l'équation 1, on obtient

$$\begin{aligned} \frac{\partial h_\sigma}{\partial \sigma}(x, y) &= \left(\frac{x^2 + y^2}{2\pi\sigma^5} - \frac{1}{\pi\sigma^3}\right) \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \\ \nabla^2 h_\sigma(x, y) &= \frac{1}{\sigma} \left(\frac{x^2 + y^2}{2\pi\sigma^3} - \frac{1}{\pi\sigma}\right) \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \end{aligned}$$

d'où

$$\frac{\partial h_\sigma}{\partial \sigma}(x, y) = \sigma \nabla^2 h_\sigma(x, y) \quad (5)$$

Soit en combinant les relations 4 et 5

$$\frac{\partial g_\sigma}{\partial \sigma}(x, y) = \sigma \nabla^2 g_\sigma(x, y) \quad (6)$$

En utilisant la notation différentielle, l'équation 6 peut aussi s'écrire

$$\sigma d\sigma = \frac{dg_\sigma(x, y)}{\nabla^2 g_\sigma(x, y)} \quad (7)$$

ou finalement

$$d(\sigma^2) = 2 \frac{dg_\sigma(x, y)}{\nabla^2 g_\sigma(x, y)} \quad (8)$$

Cette relation exprime la variation de σ^2 en fonction des variations de g , nous noterons donc désormais g par $g_{\sigma^2}(x, y)$.

La détermination du flou σ en tout point peut alors être appréhendée à partir de deux images floues $g_1 = g_{\sigma^2}$ et $g_2 = g_{\sigma^2 + \Delta(\sigma^2)}$ obtenues pour deux réglages optiques différents, mais suffisamment proches de façon à assimiler la variation $\Delta(\sigma^2)$ à $d(\sigma^2)$. La relation 2 exprimée pour ces deux réglages permet de relier les deux flous correspondant par $\sigma^2 + \Delta(\sigma^2) = (\alpha\sigma + \beta)^2$, où $\alpha = \frac{r_2 d_2}{r_1 d_1}$ et $\beta = kr_2 d_2 (\frac{1}{f_2} - \frac{1}{f_1} + \frac{1}{d_1} - \frac{1}{d_2})$. Nous obtenons ainsi $\Delta(\sigma^2) = (\alpha^2 - 1)\sigma^2 + 2\alpha\beta\sigma + \beta^2$. Les constantes α et β sont déterminées à l'issue d'une phase de calibration [1]. Un choix judicieux des paramètres à modifier lors de l'acquisition des deux images permet de simplifier cette équation. En effet, si l'on modifie uniquement la distance focale, on est dans le cas $\alpha = 1$. Si l'on choisit de modifier uniquement le diamètre r de l'ouverture, on obtient $\beta = 0$. Une faible variation de σ entre les deux images, soit un petit $\Delta(\sigma^2)$, permet d'explicitier l'expression du flou à partir de la relation 8, soit

$$\sigma = \begin{cases} \frac{\Delta(\sigma^2) - \beta^2}{2\beta} & \text{si } \alpha = 1 \\ \left(\frac{\Delta(\sigma^2)}{\alpha^2 - 1} \right)^{1/2} & \text{si } \beta = 0 \end{cases} \quad \text{avec } \Delta(\sigma^2) = 2 \frac{\Delta g_{\sigma^2}(x, y)}{\nabla^2 g_{\sigma^2}(x, y)} \quad (9)$$

où $\Delta g = g_2 - g_1$ et $\nabla^2 g = \frac{\nabla^2 g_1 + \nabla^2 g_2}{2}$.

Ainsi, pour l'une ou l'autre des modifications de réglage envisagées, la formulation proposée permet de déterminer le flou σ de manière locale.

En pratique, en raison de la sensibilité au bruit de l'opérateur Laplacien, les estimations du flou ainsi obtenues peuvent être entachées d'erreur; une estimation plus robuste pourra alors être obtenue en moyennant Δg et $\nabla^2 g$ sur un petit domaine C , soit

$$\Delta(\sigma^2)^2 = 4 \frac{\iint_C (\Delta g_{\sigma^2}(x, y))^2 dx dy}{\iint_C (\nabla^2 g_{\sigma^2}(x, y))^2 dx dy} \quad (10)$$

Le domaine C considéré sera pris suffisamment réduit afin de limiter l'effet de lissage introduit par ce moyennage sur les transitions de flou et donc de profondeur.

4 Résultats expérimentaux

Pour présenter les résultats expérimentaux obtenus par notre approche, nous considérons deux types d'images texturées: les images de la banque du groupe de recherche GDR-ISIS, que nous avons rendues floues par convolution, et ensuite des images de scènes réelles incluant des objets texturés de profondeur variable.

4.1 Validation sur des images rendues floues par construction

Les images sont construites à partir des images "sable" et "herbe" de la banque d'image du GDR-ISIS. Afin de simuler un flou spatialement variable, la première image floue du couple d'images nécessaire à la détermination du flou est obtenue par convolution de l'image nette avec un filtre gaussien non stationnaire. L'écart-type de la réponse impulsionnelle de ce filtre varie spatialement selon une gaussienne de la forme $\sigma(i, j) = a_1 + a_2 \exp(-((i - M/2)^2 + (j - N/2)^2)/2b^2)$, où $M \times N$ est la taille de l'image, a_1 , a_2 et b des constantes positives. La seconde image floue est obtenue de la même façon avec un flou différent $(\sigma^2 + \Delta(\sigma^2))^{1/2} = \alpha\sigma$. Pour l'ensemble des résultats donnés, ces constantes sont $M = N = 128$, $a_1 = 1.0$, $a_2 = 4.0$ et $b = 50.0$.

La figure 2 présente les images nettes "herbe" et "sable", ainsi que les deux couples d'images floues obtenues avec $\alpha = 1.1$. Les valeurs théoriques de $\sigma(i, j)$ sont présentées en figure 3.a. Les équations 9 et 10 permettent de déterminer les estimations de $\sigma(i, j)$ avec un domaine carré C de taille réduite. Les figures 3.b-c montrent les estimations de σ pour les deux textures, obtenues en considérant un domaine C de 13×13 pixels. Les erreurs obtenues sont de 1.3% et 1.5% respectivement pour les textures "herbe" et "sable". Bien que les deux images "herbe" et "sable" présentent des textures d'aspect différent, l'erreur sur la mesure du flou σ reste du même ordre de grandeur.

Afin d'évaluer l'évolution de l'erreur d'estimation sur σ en fonction du rapport α , deux autres images floues sont construites à partir de l'image nette "herbe" avec $\alpha = 1.3$ et 1.43. Les erreurs obtenues sont alors respectivement 3% et 5.4%. Comme on peut le prévoir, les erreurs obtenues dépendent de l'écart de α à l'unité.

Pour tester la sensibilité au bruit de notre méthode, un bruit gaussien a été ajouté aux images floues "herbe" pour obtenir un rapport signal/bruit allant jusqu'à 25dB. La figure 4 montre la carte de flou obtenue. L'erreur obtenue est alors de l'ordre de 4%, mais reste dans une limite acceptable.

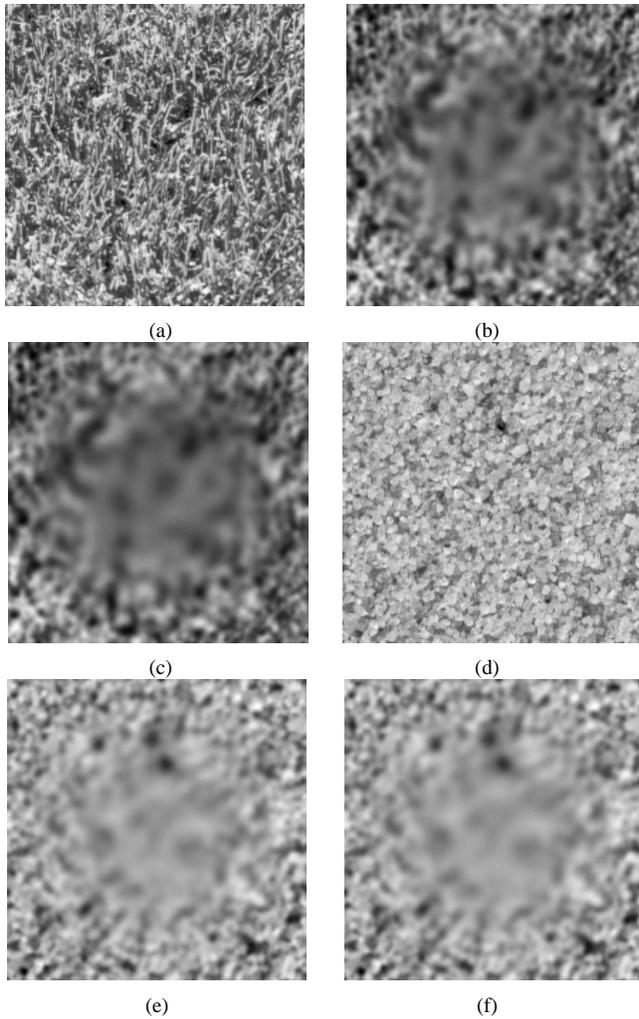


Figure 2: Les images texturées rendues floues : (a) image nette "herbe", (b) et (c) images floues "herbe" avec $\alpha = 1.1$, (d) image nette "sable", (e) et (f) images floues "sable" avec $\alpha = 1.1$.

4.2 Application aux images réelles

Les images réelles ont été acquises avec une caméra de type courant dont l'objectif est à focale variable. Les résultats que nous allons présenter ont été établis avec une focale de $30mm$ et deux nombres d'ouverture fixés à 6.8 et 8 soit $\alpha = 1.18$. Le plan net est placé entre la caméra et la scène. Les autres paramètres de la caméra, intervenant dans la relation entre le flou et la profondeur, ont été préalablement déterminés à l'issue d'une phase de calibration. Afin de réduire le bruit sur les images réelles, ces images sont préalablement filtrées avec un filtre gaussien.

Deux scènes réelles sont considérées. La première scène est composée d'un plan incliné constitué avec une texture "herbe" dont la profondeur varie de $80cm$ à $110cm$. Les deux images floues de cette scène sont présentées en figure

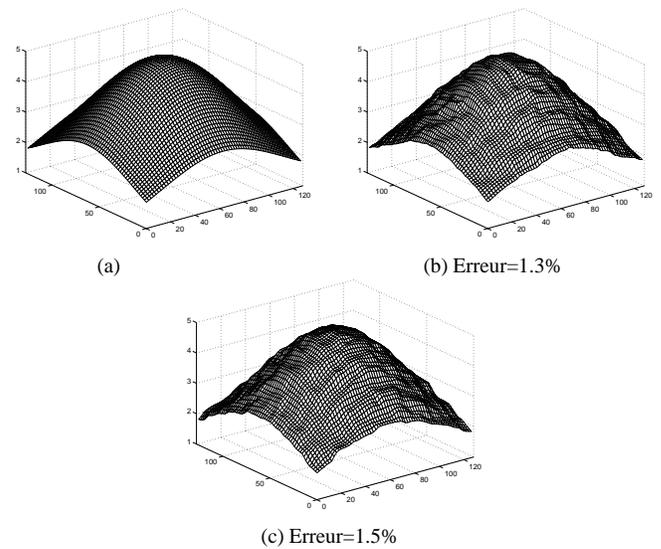


Figure 3: Résultats sur les images construites : (a) Valeurs exactes de σ , (b) Estimation de σ pour la texture "herbe", (c) Estimation de σ pour la texture "sable"

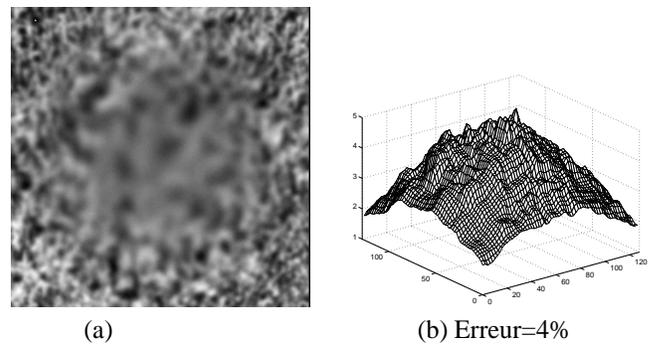


Figure 4: Estimation de σ pour la texture "herbe" en présence de bruit : (a) image floue bruitée, (b) Estimation de σ

5.a-b. L'estimation locale du flou à partir des deux images de cette scène, donnée en figure 5.c, met bien en évidence une variation du flou selon un plan incliné. La carte de profondeur correspondante est présente en figure 5.d. L'erreur d'estimation de cette carte de profondeur est inférieure à 8% . Les pics observés sur les estimations de distances, pour l'extrémité haute du plan incliné, correspondent à une zone de début de saturation du point de vue du flou, du système optique.

La seconde scène est composée de trois bandes, de faible épaisseur, dont la texture est de type "sable". Le fond de l'image, placée à $183cm$ de la caméra, est une texture du même type. La figure 6.a-b montre les images floues de cette scène. Les bandes, de gauche à droite sur cette figure, sont positionnées verticalement à $124cm$, $160cm$ et $100cm$ de la caméra.

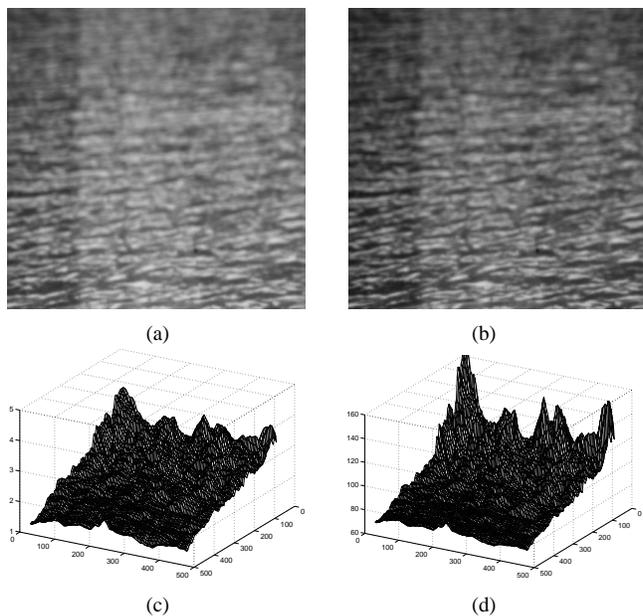


Figure 5: Estimations du flou sur une scène réelle composée d'un plan incliné type "herbe" : (a) et (b) images floues, (c) Estimation de σ , (d) Estimation de la carte de profondeur (en cm).

Les résultats d'estimation du flou et de la profondeur sont donnés en figure 6.c-d. L'allure de ces cartes fait apparaître clairement la différence de flou et de profondeur sur les différentes bandes ainsi que sur le fond de la scène. Les transitions de profondeur sont clairement identifiables sur la carte de profondeur, comme on peut le voir notamment entre la bande de droite et le fond texturé. La sensibilité au bruit de l'opérateur Laplacien peut engendrer une erreur ponctuelle d'estimation de l'ordre de 20%. L'erreur moyenne de la carte de profondeur reste inférieure à 9%.

5 Conclusion

L'approche proposée permet d'estimer la carte de profondeur en tout point d'une scène, à partir de deux images floues acquises avec deux ouvertures différentes. Nous avons établi une formulation locale du flou optique spatialement variable en chaque point de l'image. Cette formulation permet d'éviter une analyse isolée de sous-images qui introduit des erreurs dues au repliement de flou entre régions voisines. Par ailleurs, cette approche ne fait aucune hypothèse restrictive sur la forme de l'image nette si ce n'est les conditions de continuité et de dérivabilité généralement posées en traitement d'image.

Les performances de cette méthode ont été évaluées sur des images texturées construites, puis réelles. Les essais sur les images construites ont permis de reconstruire la

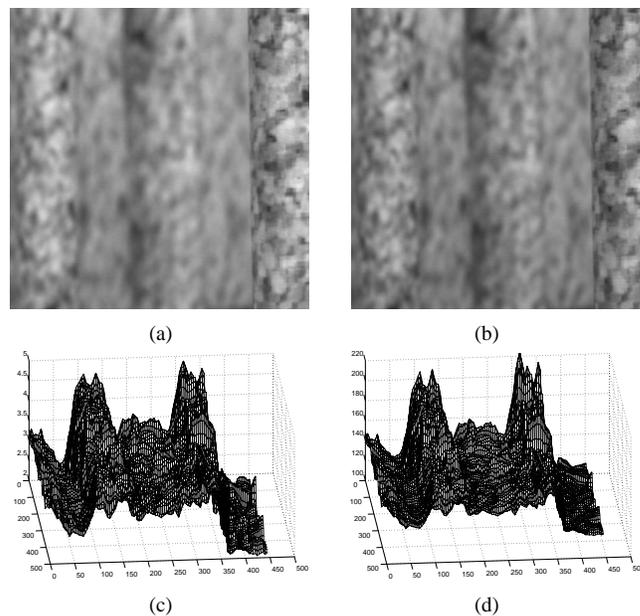


Figure 6: Estimations du flou sur une scène réelle composée de trois bandes texturées : (a) et (b) image floues, (c) Estimation de σ , (d) Estimation de la carte de profondeur (en cm).

carte de flou avec une précision inférieure à 1.5% dans le cas d'images non bruitées et inférieures à 5% dans le cas d'images bruitées avec un rapport signal/bruit allant jusqu'à 25dB. Les résultats obtenus à partir de scènes réelles contenant des objets 3D montrent que cette méthode est particulièrement adaptée aux scènes fortement texturées.

References

- [1] A. Benzougar, "Perception de la Profondeur en Stéréovision Monoculaire Passive par exploitation du Flou Optique: approche basée sur les Moments Spatiaux", Thèse de doctorat de l'Université de Poitiers, 1997.
- [2] M. Born, E. Wolf, "Principles of Optics", Pergamon, London, 1965.
- [3] BKP. Horn, "Robot Vision" - MIT Press, McGraw-Hill Book Company, Cambridge, Mass., New-York, 1986.
- [4] M. Khoudair, J. Brochard, M. Leard, "Mesure de profondeur sur des surfaces texturées par exploitation du flou optique", Seizième colloque GRETSI, Grenoble, pp. 757-760, septembre 1997.
- [5] A.P. Pentland, "A New Sense for Depth of Field", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol .9, no. 4, pp. 523-531, 1987.

- [6] A.N. Rajagopalan, S. Chaudhuri (1997), "A Variational Approach to Recovering Depth from Defocused Images", PAMI, vol. 19, no. 10, pp. 1158-1164, october 1997.
- [7] A.N. Rajagopalan, S. Chaudhuri , "Space-variant approaches for the recovery of depth from defocused images", CVGIP, Computer Vision and Image Understanding, vol. 68, no. 3, pp. 309-329, December 1997.
- [8] T. Simon, "Appréciation de la profondeur depuis des images nettes et floue en vision 3D passive", Thèse de doctorat de l'Université de NANCY I, 1991.
- [9] M. Subbarao, "Parallel depth recovery by changing camera parameters", Proc. IEEE Conference on Computer Vision, pp. 149-155,Florida, december 1988.
- [10] M. Subbarao, G. Surya, "Depth from Defocus: A Spatial Domain Approach", International Journal of Computer Vision, vol. 13, no. 3, p. 271-294, 1994.