

STEREOVISION MONOCULAIRE PAR UNE SEULE IMAGE COULEUR

Thierry Simon¹, Christophe Simon², Majdi Khoudeir³

¹Equipe Perceval, IUT de Figeac, Avenue de Nayrac, 46100 Figeac, France et CROMEP/Ecole des Mines d'Albi, Campus Jarlard, 81013 Albi CT Cedex 09. courriel: thierry.simon@univ-tlse2.fr

²ESSTIN, Universités de Nancy, 2 rue Jean Lamour, 54519 Vandoeuvre les Nancy, France. courriel: christophe.simon@esstin.uhp-nancy.fr

³SIC, Bât. SP2MI, Téléport 2, Bvd Marie et Pierre Curie, BP 30179, 86962 Futuroscope Chasseneuil Cedex, France courriel : khoudeir@sic.sp2mi.univ-poitiers.fr

Résumé: Nous proposons de décrire un algorithme original d'estimation de la profondeur d'une scène observée par une seule caméra couleur. Notre but est de mettre au point une méthode de reconstruction 3D de montages pour l'usinage, à faible coût, pour la surveillance de la validité des trajectoires d'usinage grande vitesse. L'algorithme suit la méthode que nous avons développée précédemment qui s'appuie sur le traitement d'une image nette et d'une image floue composées en rapport des modules des gradients. Cette méthode, que nous décrivons brièvement, exploite le flou optique mesuré sur les contours comme information de profondeur. Le nouvel algorithme utilise trois images floues acquises d'un même point de vue et avec les mêmes réglages des paramètres du système optique. Seule l'ouverture de l'iris change donnant un flou différent selon les profondeurs et selon la taille de l'ouverture. Nous montrons que nous pouvons substituer à l'acquisition des trois images celle d'une seule image couleur.

Mots clés: Stéréovision, flou optique, image couleur.

1. INTRODUCTION

L'obtention d'une information tridimensionnelle par caméras en vision passive reste un enjeu important pour les chercheurs en traitement de l'image. L'application que nous mettons en œuvre concerne le contrôle de la trajectoire de l'outil, produit par un système de FAO, en correspondance avec le montage effectif d'usinage. Ce contrôle prend toute son importance dans l'usinage grande vitesse pour lequel la moindre collision entraîne des surcoûts pouvant remettre en cause son intérêt. Cette concordance entre la trajectoire, placée dans la commande numérique, et le montage réel doit se vérifier à la fois dans les phases de mise au point du programme d'usinage et de production. Cette double contrainte conduit à rechercher des solutions monoculaires, et des algorithmes simples ne nécessitant pas d'architecture spécialisée pour une exécution rapide.

Les travaux en stéréovision monoculaire, depuis Pentland [1], considèrent l'objectif d'une caméra comme un système assurant une projection sur le plan capteur et tiennent compte de la nature ondulatoire de la lumière incohérente par la réponse impulsionnelle du système optique. Ces deux aspects, géométrique et ondulatoire, modélisent complètement la transformation opérée par l'objectif, et permet de lier la géométrie de la scène à son image. La projection conduit à une perte d'information qui sera compensée par l'acquisition de plusieurs images prises d'un même point de vue mais en faisant varier certains paramètres du système optique. Nous avons choisi de développer les techniques utilisant les variations de l'ouverture, celles-ci conduisent à l'exploitation du flou optique comme information de profondeur. Deux classes de méthodes peuvent être identifiées. La première fait appel à une modélisation des discontinuités de luminance, souvent représentant les contours d'objets [8], [2], et ne peut fournir qu'une carte éparse des profondeurs. Généralement, ces méthodes sont rapides, précises mais

nécessitent une interprétation complémentaire de la scène pour structurer les informations de profondeurs. A cette fin, il est nécessaire de joindre les pixels en une chaîne de contours pour assurer la cohérence géométrique. La seconde ne procède pas à une modélisation explicite de l'image, elle utilise un contenu fréquentiel dense, en pratique, des textures, puisque la défocalisation opère un filtrage passe bas dont la résultante doit être comparée pour obtenir la profondeur [3]. Ces méthodes doivent gérer au mieux le compromis entre l'uniformité locale du flou et la précision des résultats par rapport à la taille de la fenêtre de calcul. Ainsi, nous présentons une méthode de mesure de flou sur les discontinuités.

Cet article se structure en trois parties. Dans la première, nous rappelons les modèles utilisés. Dans la deuxième, nous précisons notre approche exploitant un couple d'images nettes et floues, puis celle utilisant un triplet d'images floues, pour lequel une substitution par une image couleur est proposée. Enfin nous donnons un exemple de résultat sur des images synthétiques avant de conclure.

2. MODELISATION EN STEREOVISION

2.1 Modèle de réponse impulsionnelle de l'optique

Depuis les travaux de Hopkins [5] la défocalisation peut être représentée par la convolution entre l'image située en position focalisée et la réponse impulsionnelle du système optique. La forme de cette réponse est établie en utilisant la différence de chemins optiques due au changement de position du plan image distant de ξ , montré figure 2. Elle dépend aussi de la position de l'objet par rapport à la lentille et du diamètre de l'ouverture pupillaire. Il s'agit d'exprimer la répartition de l'intensité lumineuse projetée sur le plan capteur, à partir d'un point objet situé à s_o , qui prend la forme d'un disque si le système possède une symétrie circulaire. L'expression proposée par Hopkins donne la fonction de transfert optique et conduit à la réponse impulsionnelle par l'application de la

transformée de Fourier inverse. Cette forme est difficilement utilisable, et Stokseth en a proposé une forme simplifiée (1) que l'on peut retrouver dans [6].

$$H(\rho) = 2 \left(1 - 1.38 \frac{\rho}{\rho_c} + 0.0304 \left(\frac{\rho}{\rho_c} \right)^2 + 0.344 \left(\frac{\rho}{\rho_c} \right)^3 \right) \times \frac{J_1 \left(4k\xi \left(1 - \frac{\rho}{\rho_c} \right) \frac{\rho}{\rho_c} \right)}{4k\xi \left(1 - \frac{\rho}{\rho_c} \right) \frac{\rho}{\rho_c}} \quad (1)$$

où λ est la longueur d'onde, $k = \frac{2\pi}{\lambda}$, $J_1(\cdot)$ la fonctions de Bessel de première espèce et d'ordre 1, $\rho = \sqrt{u^2 + v^2}$ la

fréquence spatiale et $f_c = \frac{L}{s_o \lambda}$ la fréquence de coupure.

En outre, si l'on considère la sensibilité du capteur $S(\lambda)$, comprise entre λ_1 et λ_2 , pour une gamme de couleurs, la forme de la réponse impulsionnelle se déduit de l'expression suivante :

$$H_c(\rho) = \int_{\lambda_1}^{\lambda_2} S(\lambda) H(\rho) d\lambda \quad (2)$$

Ce dernier modèle peut être lui-même approché par le modèle gaussien, le plus utilisé actuellement [3], [2], [8], qui représente un bon modèle expérimental et calculatoire dont la forme simple et directe de la réponse impulsionnelle est donnée par l'équation (3) dans le domaine spatial. Ainsi pour les images couleurs il existe une réponse impulsionnelle par gamme de couleurs, typiquement R, V, B caractérisées chacune par une sensibilité identifiée à partir du comportement du filtre et du capteur et un écart-type ($\sigma_R, \sigma_V, \sigma_B$).

$$h(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (3)$$

La formation de l'image floue if_{s_o} est alors traduite par le produit de convolution entre l'image nette du point source et la réponse impulsionnelle du système optique pour la profondeur s_o , $if_{s_o}(x, y) = in(x, y) \otimes_{2D} h(x, y, s_o)$ où $in(x, y)$ est l'image nette.

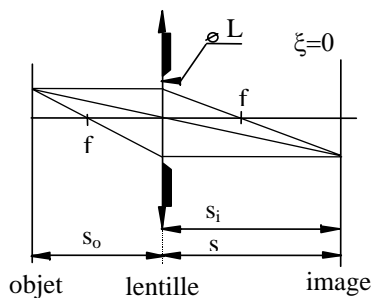


Figure 1 : Image nette focalisée

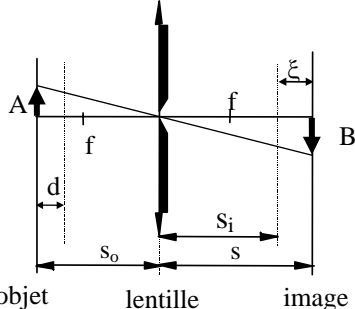


Figure 2 : Image nette, ouverture réduite de l'iris.

2.2 Modèle de la relation flou profondeur

La plupart des travaux en imagerie numérique modélisent le système optique dans son comportement projectif par une lentille simple comme le montre la figure 1, même dans le cas d'utilisation d'un système composé de plusieurs lentilles (zoom). Dans cette hypothèse, nous pouvons retrouver les lois simples de l'optique géométrique (4) et (5) :

$$\frac{1}{s_o} + \frac{1}{s_i} = \frac{1}{f} \quad (4)$$

$$2R = \frac{Ls}{\frac{1}{f} - \frac{1}{s_o} - \frac{1}{s_i}} \quad (5)$$

Cette équation, issue des relations des triangles semblables, facilement identifiable sur la figure 4, permet de trouver la relation entre le diamètre du disque représentant l'image du point et la distance entre l'objet et la lentille, exprimée par la relation (4). Dans le cas d'utilisation du modèle gaussien une relation de proportionnalité est choisie entre le paramètre σ et le diamètre du disque flou. σ devient ainsi une mesure de flou et une mesure de profondeur. Les techniques mises en œuvre dans de nombreux travaux visent à l'évaluation de σ . L'équation (6) résume la relation entre le flou et la profondeur, distance entre le plan lentille et le plan objet, pour une partie de la courbe (I) ou (II) présentée figure 3 :

$$\begin{aligned} s_o &= \frac{fs}{s_i - f - 2R \frac{f}{L}} \quad \text{pour } \xi < 0 \\ s_o &= \frac{fs}{s_i - f + 2R \frac{f}{L}} \quad \text{pour } \xi \geq 0 \end{aligned} \quad (6)$$

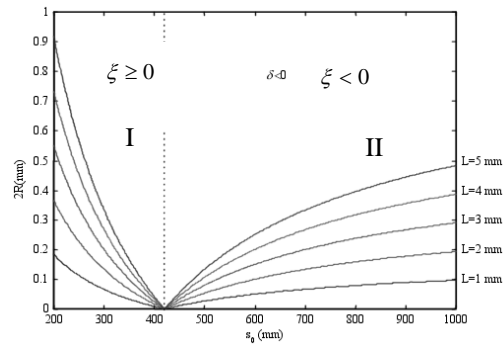


Figure 3 : Relation entre le flou et la profondeur.

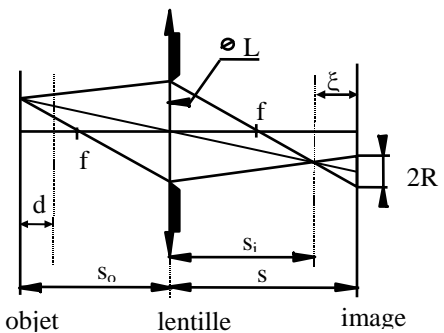


Figure 4 : Image défocalisée, floue.

Les paramètres m et c (7) sont déterminés dans une phase de calibration, ils englobent les paramètres intrinsèques du système optique. Une fois la profondeur connue, les coordonnées parallèles au plan capteur sont calculées, dans

le repère de référence, en application inverse de la projection en utilisant le facteur de grossissement B/A mis en évidence sur la figure 2.

$$s_o^{-1} = m\sigma + c \quad (7)$$

2.3 Modèle de discontinuités de luminance

Le modèle le plus simple s'appuie sur la forme d'une fonction de Heaviside, dans la direction du gradient, $\delta = b - a$ étant le contraste dans l'expression (8) de $c_1(x)$:

$$c_1(x, y) = \begin{cases} a & x < x_o, \forall y \\ a + \delta & x \geq x_o, \forall y \end{cases} \quad (8)$$

$$c_2(x) = \begin{cases} a & x < x_o, \forall y \\ a + \frac{\delta}{\varepsilon} (x - x_o) & x_0 \leq x \leq x_o + \varepsilon, \forall y \\ a + \delta & x > x_o + \varepsilon, \forall y \end{cases} \quad (9)$$

Mais la réalité expérimentale montre que la forme la plus rencontrée est celle dont le modèle est $c_2(x)$ (9). Un autre modèle de contour permet de s'approcher encore plus finement des contours rencontrés en pratique. L'expression (10) modélise ce type de discontinuité de luminance. Plusieurs paramètres des éléments du système d'acquisition, de l'objectif, du capteur, les aberrations chromatiques et le comportement électronique de la surface sensible, le temps d'intégration [7] peuvent contribuer à l'obtention de tels contours. L'ensemble des ces contributions peut être modélisé par une réponse impulsionnelle de forme gaussienne du dispositif optoélectronique :

$$c_3(x, y) = c_1(x, y) \otimes_D p(x, y) \quad (10)$$

$$\text{avec } p(x, y) = \frac{1}{2\pi\sigma_0^2} \exp\left(-\frac{x^2 + y^2}{2\sigma_0^2}\right)$$

3. METHODE AVEC UNE IMAGE NETTE ET UNE IMAGE FLOUE

Notre méthode suit le synoptique de la figure 5, le flou est évalué sur les contours, détectés dans l'image nette. La valeur du flou est estimée à partir de la composition du rapport des modules des gradients qui donne la relation

$$\text{suivante : } R(x, y) = \frac{|\nabla \text{in}(x, y)|}{|\nabla \text{if}(x, y)|} = \frac{|\nabla \text{in}(x, y)|}{|\nabla (\text{in}(x, y) \otimes h(x, y))|} \quad (11)$$

Cette relation se réduit, compte tenu du modèle du contour net, à :

$$R(x) = \frac{|\nabla c_2(x)|}{|\nabla c_f(x)|} = \frac{|\nabla c_2(x)|}{|\nabla c_2(x) \otimes h(x)|} \quad (12)$$

avec $h(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{x^2}{2\sigma^2})$, qui finalement donne le système (13):

$$R(x) = \begin{cases} \frac{1}{h(x-x_o) + h(x-(x_o+\varepsilon)) + 2\sum_{u=1}^{\varepsilon-1} h(x-(x_o+u))} & \text{pour } x = x_o \text{ et } x = x_o + \varepsilon \\ \frac{h(x-x_o) + h(x-(x_o+\varepsilon)) + 2\sum_{u=1}^{\varepsilon-1} h(x-(x_o+u))}{2} & \text{pour } x \in]x_o; x_o + \varepsilon[\end{cases} \quad (13)$$

Avec cette expression du rapport des modules des gradients, nous sommes capables de calculer $(\varepsilon + 1)$, valeurs de σ pour tout $x \in [x_o; x_o + \varepsilon]$. On obtient une valeur finale par la moyenne des valeurs estimées. La carte des profondeurs produite est éparse, elle dépend des contours détectés.

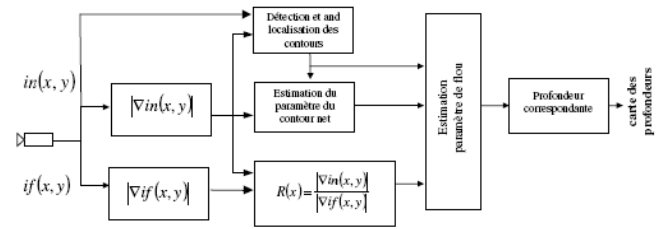


Figure 5 : Schéma de principe

4. METHODES AVEC TROIS IMAGES FLOUES

L'inconvénient majeur de la méthode précédente est la nécessité de disposer d'une image nette dont l'acquisition est réalisée avec une très petite ouverture de l'iris. La quantité de lumière arrivant sur le capteur devient alors faible ce qui conduit à un rapport signal sur bruit défavorable à la détection de contours. Nous proposons d'utiliser trois images floues, acquises avec trois ouvertures différentes, présentant sur chaque contour σ différents. En s'appuyant sur le modèle $c_3(x)$ nous pouvons écrire :

$$R^{jk}(x) = \frac{|\tilde{\nabla} \text{if}_j(x)|}{|\tilde{\nabla} \text{if}_k(x)|} = \frac{\sqrt{\sigma_0^2 + \sigma_k^2}}{\sqrt{\sigma_0^2 + \sigma_j^2}} \exp\left(-\frac{x^2}{2} \left(\frac{\sigma_k^2 - \sigma_j^2}{(\sigma_0^2 + \sigma_j^2)(\sigma_0^2 + \sigma_k^2)} \right)\right) \quad (14)$$

Le logarithme de ce rapport fait apparaître deux grandeurs

A^{jk} et B^{jk} :

$$B^{jk} = \frac{1}{2} \ln \left(\frac{\sigma_0^2 + \sigma_k^2}{\sigma_0^2 + \sigma_j^2} \right) \text{ et } A^{jk}(x) = -\frac{\sigma_k^2 - \sigma_j^2}{2(\sigma_0^2 + \sigma_j^2)(\sigma_0^2 + \sigma_k^2)}$$

L'estimation des A^{jk} et B^{jk} peut se faire sur le contour dans

la direction du gradient en utilisant une régression. La contribution de l'image nette apparaît dans l'expression des grandeurs A^{jk} et B^{jk} . Une troisième image composée avec la première fournit deux équations supplémentaires afin de connaître la contribution de la défocalisation.

Pour A^{jk} , nous pouvons écrire une équation linéaire du type

$$Y = MX \text{ avec } Y = (\ln R^{jk}(x_1) \dots \ln R^{jk}(x_n)), M = (A^{jk} \ B^{jk}),$$

$$X = \begin{pmatrix} x_1^2 & \dots & x_n^2 \\ 1 & \dots & 1 \end{pmatrix} \text{ et } x_j \text{ les positions des points du}$$

contour. Une fois deux valeurs de A^k connues, par exemple $A = A^{12}$ et $A^{13} = E$ constituées à partir de trois images floues, nous pouvons trouver σ_1 par l'équation (15) et σ_0 par (16).

$$\sigma_1 = \pm \sqrt{\frac{(A(1 - \alpha_{13}^2) + E(\alpha_{12}^2 - 1))^2}{2EA(\alpha_{13}^2 - 1)(\alpha_{12}^2 - 1)(\alpha_{12}^2 - \alpha_{13}^2)(A - E)}} \quad (15)$$

et

$$\sigma_0 = \pm \sqrt{\frac{(A(1 - \alpha_{13}^2) + E(\alpha_{12}^2 - 1))(A\alpha_{12}^2(\alpha_{13}^2 - 1) + E\alpha_{13}^2(1 - \alpha_{12}^2))}{2EA(\alpha_{13}^2 - 1)(\alpha_{12}^2 - 1)(\alpha_{12}^2 - \alpha_{13}^2)(A - E)}} \quad (16)$$

Seules les racines positives et réelles sont à considérer. Les coefficients α_{jk} représentent le rapport des flous σ_i qui dépendent des réglages de l'ouverture. En effet, selon la relation entre le flou et l'éloignement qui dépend du diamètre de l'ouverture, une modification de celle-ci entraîne une modification proportionnelle du flou [9]. Ces rapports sont constants quelque soit la profondeur et peuvent être déterminés dans une phase de calibration. L'algorithme de la méthode est résumé sur le schéma fonctionnel de la figure 5.

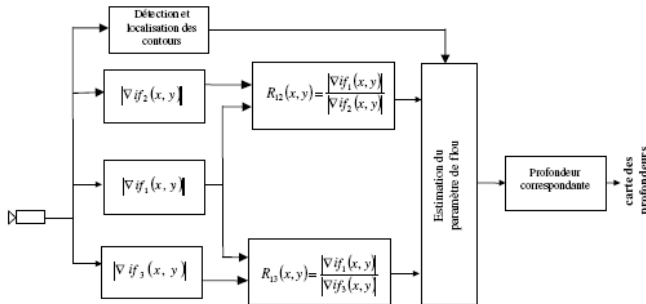


Figure 6 : Schéma fonctionnel.

L'utilisation d'une seule image couleur donnant trois flous $(\sigma_R, \sigma_V, \sigma_B)$ selon la gamme de couleur permet de substituer ce dernier triplet au triplet $(\sigma_1, \sigma_2, \sigma_3)$.

5. QUELQUES RESULTATS

Les images synthétiques présentent trois formes rectangulaires sombres de niveau de gris valant 60 pour l'objet 1, 225 pour l'objet 2 et 130 pour l'objet 3, sur un fond clair uniforme simulant trois parallélépipèdes vus de dessus. Le paramètre de contour est considéré constant et égal à 1 pour les trois objets. Un flou différent par objet (1, 2, 3) est simulé, respectivement $\sigma_1 = 2$, $\sigma_2 = 2.1$ et $\sigma_3 = 2.2$.

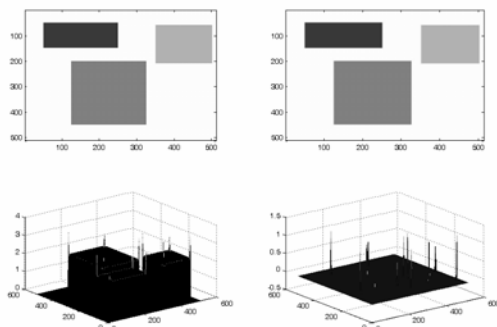


Figure 7 : Synthèse des résultats.

Les résultats, présentés de façon synthétique par la figure 7, montrent que la méthode donne une estimation exacte des valeurs de flous sur les points du contour. En revanche, les coins présentent une erreur non négligeable due au non respect de l'hypothèse de forme du contour.

6. CONCLUSION

La méthode que nous proposons montre que nous pouvons utiliser une seule image couleur pour retrouver la profondeur sur les discontinuités de luminance dans une scène observée par une seule caméra. Nous obtenons une carte éparsée des profondeurs mais la simplicité des traitements permet d'envisager une implantation matérielle peu onéreuse assurant une exécution à la cadence vidéo. La phase d'étalonnage doit être particulièrement soignée notamment dans l'évaluation des rapports de flou a. Il faut s'assurer de l'homogénéité dans le comportement optoélectronique des capteurs pour chaque couleur. Une évaluation de sensibilité aux bruits de la méthode doit être réalisée. Nous étudions également l'influence d'une erreur sur la localisation des contours ou d'une modification trop importante de la valeur de contraste sur les performances.

REFERENCES

- [1] A. Pentland, *A new sense for depth of field*, IEEE Trans. PAMI, vol. 9, n°4, pp. 523-531, 1987
- [2] M. Subbarao and J. K. Tyan, *Selecting the Optimal Focus Measure for Autofocusing and Depth-from-Focus*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 8, pp. 864-870, August 1998.
- [3] F. Deschênes, D. Ziou, P. Fuchs, *An unified approach for simultaneous and cooperative estimation of defocus and spatial shifts*, Image and Vision Computing, n° 22, pp. 35-57, 2004.
- [4] M. Léard, M. Khoudair, J. Brochard, J. Bernard, *Depth measurement of textured surface in a 3D scene*, IEEE International Conference on Electronics Circuits and Systems, Cairo, pp 435-438, December 1997
- [5] H. H. Hopkins, *The frequency response of a defocused optical system*, Proc. Of the Royal Society of London series A, pp. 91-103, February 1955.
- [6] N. Dey, A. Boucher, M Thonnat, *Image formation model of a 3D translucent object observed in light microscopy*, International Conference on Image Processing, Rochester (New York), pp. II-469- II-472 vol.2, 22-25 September 2002.
- [7] Orly Yadid-Pecht, Ralph Etienne-Cummings, *CMOS Imagers: From phototransduction to image processing*, Kluwer Academic Publishers, 2004.
- [8] C. Simon, F. Bicking, T. Simon, *Depth Estimation Based on Thick Edges in Images*, IEEE ISIE'04, Ajaccio, France, 3-8 Mai, 2004.
- [9] A. BENZOUGAR, *Passive Monocular Stereovision, Extension of the Subbarao Method*, IEEE Proceeding MECATRONICS, The 2nd Japan-France Congress on Mecatronics, Takamatsu, Kagawa, Japan, 1994, pp 693-696.