

# Evaluation du flou optique pour la mesure de profondeur en vision monoculaire

par

**B. HEIT \* - G. SCHNEIDER\* - T. SIMON\*\* - G. MARTIN\* -  
J. BREMONT\***

**\* Université de Nancy I  
Centre de Recherche en  
Automatique de Nancy**

**\*\* Université de Poitiers  
Ecole Supérieure des  
Ingénieurs de Poitiers**

**RESUME :** La méthode présentée dans cet article constitue un outil original de reconstruction de la profondeur des objets dans une scène, par mesure du flou optique en vision monoculaire. La première partie rappelle les principes physiques employés en formation des images et les différentes études effectuées dans ce domaine. La phase délicate de notre approche est la détermination des fonctions de transfert du système optique, dont on cherche à extraire un paramètre assimilé à une mesure de flou. Les relations de l'optique géométrique, complétées par un étalonnage reliant le paramètre extrait à la profondeur du point dans la scène, font l'objet de la deuxième partie. L'attrait de cette technique est sa capacité à apporter une information 3D en s'affranchissant du problème d'appariement nécessaire en stéréovision. Ces travaux intéressent la robotique mobile pour la navigation de véhicules autonomes, et également les systèmes de contrôle ou d'assemblage de pièces.  
**MOTS-CLES :** Vision 3D monoculaire, analyse de scènes pour la robotique, carte des profondeurs, flou optique.

**ABSTRACT :** We present a new monocular computer vision method for depth reconstruction of the objects in a scene by measuring the optical blur. The first part is dedicated to the repetition of the applied optical principles of the imaging process. Then, we summarise the preceding works on the field. The delicate step of our technique is the identification of the optical transfer function, wherefore we extract a parameter assimilated with the amount of blur. In the second part, we give the geometrical optical relations completed by the calibration of the system linking the extracted parameter with the depth of the object points. The convenience of our technique is its ability to perform the depth information avoiding the matching problem as it appears in stereo vision. Our works find an interest in robotics, autonomous vehicle navigation, and also in control or assembly tasks.

**KEYWORDS :** Monocular 3D vision, image analysis for robotics, depth map, optical blur.

## **Table des matières**

### **1. Introduction**

### **2. Le flou optique porteur de l'information de profondeur**

#### **2.1. Image focalisée - image défocalisée**

#### **2.2. Formation des images**

#### **2.3. Modèles de la fonction de transfert optique**

### **3. Extraction de l'information de profondeur**

#### **3.1. Introduction**

#### **3.2. Les approches "Depth From Defocus" existantes**

### **4. Approche par identification des contours**

#### **4.1. Introduction**

#### **4.2. Identification du paramètre de la réponse impulsionnelle**

#### **4.3. Etablissement de la correspondance flou-éloignement**

#### **4.4. Etalonnage**

#### **4.5. Schématisation du traitement**

#### **4.6. Résultats obtenus**

### **5. Conclusion et perspectives**

### **Références bibliographiques**

## **1. Introduction**

L'utilisation de la vision, en tant que capteur extéroceptif, paraît naturelle et l'évolution des capteurs de vision dans leur ensemble, a donné naissance à tout un ensemble de techniques de traitement d'images pour extraire les éléments pertinents. Cependant, pour beaucoup d'applications, une vue bidimensionnelle est insuffisante et l'information tridimensionnelle doit être extraite depuis l'analyse d'images de luminance.

En vision 3D, l'apport d'informations supplémentaires pour déterminer la troisième dimension procède selon diverses démarches aboutissant à des solutions variées spécifiques du domaine d'application. Nous nous intéressons à la vision 3D passive, où l'information acquise provient de plusieurs images de la même scène pour des vues différentes par la position, l'angle, l'ouverture ou la focale. Dans ce contexte, les visions binoculaire et trinoculaire connaissent un fort développement. Les opérations liées à ces techniques demandent des moyens élaborés de représentation symbolique de l'image initiale, puis de mise en correspondance de chaque élément caractéristique par des techniques dites d'appariement.

L'utilisation d'un capteur monoculaire fixe offre une alternative à ces dispositifs complexes. Nous trouvons principalement deux approches : l'une utilise un objectif à focale variable permettant par triangulation d'obtenir des données tridimensionnelles [Lavest-91] et l'autre, décrite dans ce document, pratique le flou optique.

L'évaluation de la profondeur d'objets dans une scène par l'utilisation du flou optique demande une bonne connaissance du processus de formation des images et, plus particulièrement, de la fonction de transfert optique de l'objectif utilisé. Ainsi, pour déterminer avec exactitude la relation entre le flou et l'éloignement, une modélisation de la réponse impulsionnelle optique est nécessaire.

## **2. Le flou optique porteur de l'information de profondeur**

Le comportement physique d'un système optique dans la formation des images peut être modélisé par la fonction de transfert entre la scène observée et sa représentation sur le plan image.

En analyse d'image, le système optique fournit deux modèles principaux. Le premier se base sur l'optique géométrique. Le deuxième modèle tient compte du phénomène de diffraction mais, bien que plus complet, reste complexe.

### **2.1. Image focalisée - image défocalisée**

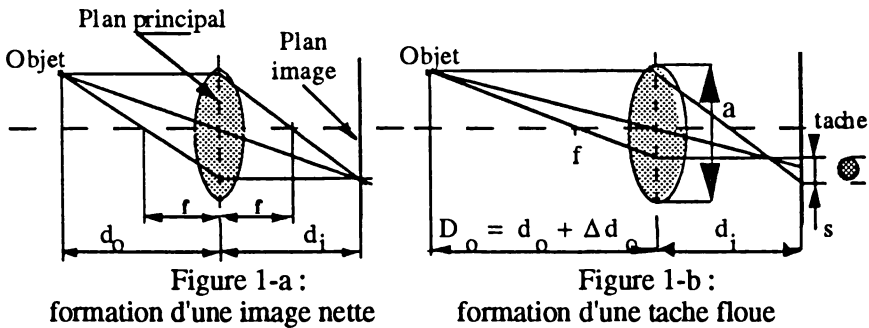
Pour assurer correctement la formation d'une image sur un plan, plusieurs lentilles forment généralement les objectifs photographiques. En ce qui nous concerne, nous assimilerons un objectif (quel que soit le nombre de lentilles le composant) à une lentille convergente, mince et de longueur focale

équivalente. Si l'on étudie le comportement d'une lentille en approximation paraxiale, la loi de conjugaison prend la forme simple donnée par la formule de Gauss :

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i}, \quad (1)$$

où  $f$ ,  $d_o$ ,  $d_i$  sont respectivement la distance focale, la distance plan principal - objet, la distance plan principal - plan image.

La figure 1-a donne un exemple de formation de l'image nette d'un point objet.



A un plan image à distance  $d_i$  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 sur le plan image (figure 1-b). L'équation 2, dans laquelle "a" représente l'ouverture de l'objectif utilisé, donne le diamètre  $s$  de la tache, appelé cercle de confusion, en fonction de la distance  $D_o$  et du plan focalisé  $d_o$  [Bove-89],

$$s(d_o, D_o, f, a) = a \left| \frac{D_o - f}{D_o} \right| \left| \frac{D_o}{D_o - f} - \frac{d_o}{d_o - f} \right|. \quad (2)$$

Cette expression est à la base de toutes les méthodes qui utilisent le flou optique comme information de profondeur. Cependant, comme le montre la figure 2, des limitations en précision sont dues à des considérations purement physiques, concernant la profondeur de champ et le tassement du flou. Il faudra donc choisir judicieusement l'objectif, en fonction de la scène à analyser.

De plus, les aberrations géométriques et les effets de diffraction limitent la résolution dans le plan image.

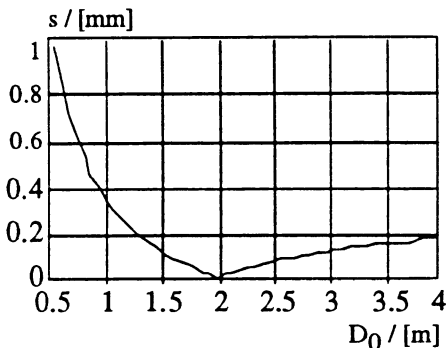


Figure 2 :  
évolution du diamètre de la tache

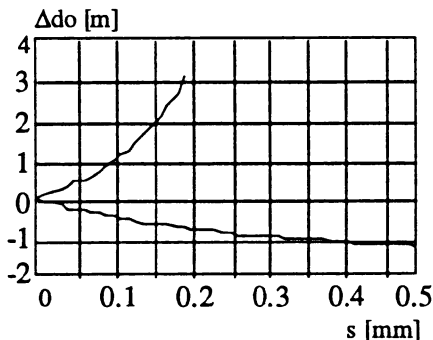


Figure 3 :  
variation de  $\Delta d_0$  fonction de  $s$

Pour calculer la profondeur d'un point objet dans la scène pour un diamètre de tache donné, on inverse l'équation 2 et on obtient pour  $\Delta d_0 = D_0 - d_0$  :

$$\Delta d_0 = \begin{cases} \frac{(f - d_0) d_0 s}{d_0 s - s f + a f}, & 0 \leq s \\ \frac{(d_0 - f) d_0 s}{-d_0 s + s f + a f}, & 0 \leq s < \frac{a f}{d_0 - f} \end{cases} \quad (3)$$

Cette relation, représentée figure 3, montre que pour "s" petit, il y a deux solutions de  $\Delta d_0$  représentant un plan objet en avant ou en arrière du plan net. Afin d'éviter cette ambiguïté, il est nécessaire de choisir  $d_0$  minimal ou maximal pour la scène à observer, c'est-à-dire de focaliser respectivement en avant ou en arrière de la scène observée [Schneider-93].

A partir de ces considérations uniquement valables pour des points objets isolés, nous étendons le principe aux images plus complexes, en modélisant leur formation en une décomposition de points lumineux.

## 2.2. Formation des images

Les relations entre l'objet et l'image deviennent plus facilement exploitables, si l'instrument optique peut être modélisé par un système linéaire. Ainsi, on décompose la fonction d'entrée  $g(x, y)$  représentant les points de l'objet en une infinité de fonctions de Dirac, pondérées par la luminosité :

$$g(x, y) = \iint_E g(x_0, y_0) \delta(x - x_0, y - y_0) dx_0 dy_0, \quad (4)$$

où  $(x_0, y_0)$  sont les coordonnées objets.

Le système optique transforme du signal d'entrée  $g(x, y)$  en un signal  $g'(x, y)$  sur le plan image, ce que décrit par l'opérateur  $T$  :

$$g'(x, y) = T \left\{ \iint_E g(x, y) \delta(x - x_0, y - y_0) dx_0 dy_0 \right\}. \quad (5)$$

Sous l'hypothèse d'illumination incohérente, le système optique se comporte linéairement vis-à-vis des intensités lumineuses  $g(x, y)$ . Nous obtenons donc :

$$\begin{aligned} g'(x, y) &= \iint_E g(x, y) T \left\{ \delta(x - x_0, y - y_0) \right\} dx_0 dy_0 \\ &= g(x, y) \otimes \text{PSF}(x, y), \end{aligned} \quad (6)$$

avec  $\text{PSF}(x, y, x_0, y_0) \equiv T \left\{ \delta(x - x_0, y - y_0) \right\}$ , où  $\text{PSF}(x, y, x_0, y_0)$  est la transformation d'une impulsion en  $(x_0, y_0)$ , traversant l'optique d'un système spatialement invariant, appelée Fonction d'étalement ou Point Spread Function (PSF).

Le modèle se complète sur les scènes tridimensionnelles pour des points objets appartenant à "n" plans différents et caractérisés par leur réponse impulsionnelle  $\text{PSF}_i$ . En ne superposant que les contributions qui sont visibles par l'objectif, il en résulte :

$$\begin{aligned} g'(x, y) &= \sum_{i=1}^n \iint_{E_i} g_i(x, y) \text{PSF}_i(x - x_0, y - y_0) dx_0 dy_0 \\ &= \sum_{i=1}^n g_i(x, y) \otimes \text{PSF}_i(x, y). \end{aligned} \quad (7)$$

Dans le domaine fréquentiel, le comportement d'un système linéaire et spatialement invariant est décrit de façon équivalente par :

$$G'(u, v) = \text{OTF}(u, v) \cdot G(u, v), \quad (8)$$

où OTF (Optical Transfer Function) représente la transformation Fourier 2D (F) de la réponse impulsionnelle.

$$\begin{aligned} \text{OTF}(u, v) &= F \{ \text{PSF}(x, y) \} \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \text{PSF}(x, y) \exp[-j(ux + vy)] dx dy. \end{aligned} \quad (9)$$

## 2.3. Modèles de la fonction de transfert optique

Un point lumineux de la scène est interprété en tant qu'impulsion de Dirac. On modélise la formation du flou dans le domaine spatial, par la fonction d'étalement (PSF). Cependant, la PSF dépend autant des propriétés des matériaux de l'optique (indice de réfraction), que de la forme géométrique de la lentille (longueur focale) et des paramètres de prise de vue (distance de l'objet, diaphragme, éclairage) [Born-90].

Une modélisation réelle, tenant compte aussi bien des aberrations géométriques que des effets de diffraction, ne s'effectue que difficilement. Une forme connue pour la PSF est la forme gaussienne 2D, utilisée dans notre démonstration, et décrite par :

$$\text{PSF}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \quad (10)$$

Son écart type  $\sigma$  représente le diamètre de la tache floue sur le plan image. La relation de proportionnalité entre  $\sigma$  et le diamètre  $s$  de la tache, formée par un point lumineux, peut être modélisée par :

$$s = k \sigma. \quad (11)$$

Pour caractériser un système, on utilise également les réponses du système optique à une impulsion de Dirac translatée suivant une direction appelée Line Spread Function (LSF). La réponse à un front lumineux (fonction échelon ou fonction d'Heavyside) est appelée Edge Spread Function (ESF). De plus,

$$\begin{aligned} \text{LSF}(x) &= \int_{y=-\infty}^{+\infty} \text{PSF}(x, y) dy, \\ \text{ESF}(x) &= \int_{-\infty}^x \text{LSF}(x_0) dx_0. \end{aligned} \quad (12)$$

## 3. Extraction de l'information de profondeur

### 3.1. Introduction

Dans le domaine spatial, les méthodes optiques de vision 3D monoculaire passive, exploitant ce flou optique, se regroupent en deux types. Le premier, que l'on qualifie de "Depth From Focus" (D.F.F.), mesure la profondeur à partir de la focalisation optimale pour chaque point ou région d'intérêt [Krotkov-89]. Coûteux en temps de calcul, il s'applique surtout à l'évaluation

de scènes provenant de la microscopie et ne sera pas développé. Le deuxième type, qualifié "Depth From Defocus" (D.F.D.), mesure la profondeur des objets à partir de deux ou plusieurs images de la même scène. On identifie alors les PSF, pour chaque point d'intérêt dans une image défocalisée, par comparaison avec les autres images.

### 3.2. Les approches "Depth From Defocus" existantes

Les deux approches principales sont les techniques exploitant le contenu spatial de l'image par les caractéristiques géométriques ou la forme des objets dans la scène, et les techniques exploitant le contenu fréquentiel par une analyse de Fourier. Les méthodes présentées utilisent principalement, dans le domaine spatial, les caractéristiques géométriques de l'image. Cependant, dans la plupart des méthodes, on ne mesure pas directement la taille de la tache, mais la valeur du paramètre d'étalement de la fonction PSF ou LSF, qui modélise le processus de formation de la tache.

Les solutions actuelles sont proposées par A.P. Pentland, P. Grossmann, M. Subbarao et R.A. Jarvis.

#### 3.2.1. Approche de PENTLAND

Les images sont des scènes contenant des objets réels et l'auteur propose deux approches, pour évaluer le paramètre d'étalement  $\sigma$ , que sont l'utilisation de fronts nets et la comparaison d'images à des ouvertures différentes [Pentland-82-87-89].

Dans la première approche à partir de fronts nets, on utilise la mesure du passage par zéro du laplacien pour déterminer les contours. Pour un point  $f(x, y)$  du plan image placé sur un front de discontinuité en  $x_0$ , nous avons :

$$f(x, y) = \begin{cases} a & \text{si } x < x_0, \\ a + b & \text{si } x \geq x_0. \end{cases} \quad (13)$$

Nous définissons le contour par :

$$C(x, y) = \nabla^2 (G(r, \theta) \otimes f(x, y)) \quad (14)$$

où  $\otimes$  désigne l'opérateur de convolution, et  $G(r, \sigma)$ , la PSF 2D non normalisée de forme gaussienne en coordonnées polaires.

Pour  $x_0 = 0$  il vient alors :

$$C(x, y) = b \frac{dG(r, \theta)}{dx} = -\frac{b x}{\sqrt{2\pi} \sigma^3} \exp\left(\frac{-x^2}{2\pi\sigma^2}\right). \quad (15)$$



Sous forme logarithmique, cette équation s'écrit :  $A x_i + B = C_i$ .  
Tous les points situés sur une frontière  $b$  suivent cette équation, qui peut être interprétée comme celle d'une droite de régression, et nous obtenons :

$$A = \frac{\sum_{i=1}^N (x_i^2 - x^2) C_i}{\sum_{i=1}^N (x_i^2 - x^2)}. \quad (16)$$

Pour un  $b$  donné nous avons :  $\sigma = (2 A)^{-\frac{1}{2}}$  (17)

Pour la deuxième approche, dont la figure 4 nous donne le schéma, la profondeur de champ est modifiée par une ouverture différente. L'amplitude de cette différence est une fonction simple de la distance entre le capteur et le point image. Pour obtenir une estimation de la profondeur, il suffit de comparer localement deux images suivant les contours.

Si  $f_k(r, \theta)$  exprime, en coordonnées polaires, une partie de l'image  $I_k(x, y)$  centrée en  $(x_0, y_0)$ , avec une ouverture  $a_k$ , son expression prend alors les formes suivantes :

pour l'ouverture  $a_1$ ,

$$f_1(r, \theta) = I_1(x_0 + r \cos \theta, y_0 + r \sin \theta), \quad (18)$$

pour l'ouverture  $a_2$ ,

$$f_2(r, \theta) = I_2(x_0 + r \cos \theta, y_0 + r \sin \theta). \quad (19)$$

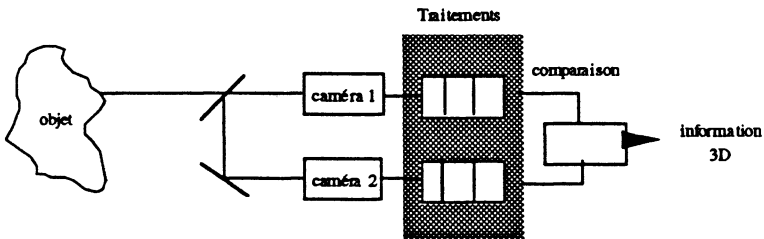


Figure 4 : schéma de principe [Pentland-89]

Après avoir traversé le système optique, l'expression des images prend la forme:

$$f_k(r, \theta) = f_0(r, \theta) \otimes G_k(r, \theta), \quad (20)$$

avec

$$G_k(r, \theta) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(\frac{-r^2}{2\pi\sigma_k^2}\right), \quad (21)$$

où  $G(r, \theta)$  représente la réponse impulsionnelle du système optique et  $f_0$  l'expression de l'image nette.

Puis, nous calculons localement la transformée de Fourier des deux images :

$$F_k(\lambda, \theta) = F_0(\lambda, \theta) \cdot G_k\left(\lambda, \frac{1}{\sigma_k\sqrt{2\pi}}\right) \quad (22)$$

et élaborons le rapport :

$$\frac{F_1(\lambda, \theta)}{F_2(\lambda, \theta)} = \frac{G_1\left(\lambda, \frac{1}{\sqrt{2\pi}}\right)}{G_2\left(\lambda, \frac{1}{\sqrt{2\pi}}\right)} = \exp\left(\lambda^2 2\pi(\sigma_2^2 - \sigma_1^2)\right). \quad (23)$$

En passant au logarithme, nous pouvons, pour les mêmes raisons que dans l'approche à partir des fronts nets, considérer l'équation comme celle d'une droite de régression. De plus, si la caméra 1 a une petite ouverture, nous savons que  $\sigma_1 = 0$ . Nous déterminons  $\sigma_2$  qui est identique (équation 17) à celui trouvé dans la méthode utilisant les fronts de discontinuité.

### 3.2.2. Approche de GROSSMANN

Pour l'image utilisée, constituée de lignes horizontales et verticales noires sur un fond blanc, et ne possédant qu'un seul plan net, l'acquisition se fait à grande ouverture. La méthode proposée s'articule en trois étapes [Grossmann-87].

En premier lieu, on extrait la primitive contour par un opérateur de type gradient. On évalue alors la largeur de la distribution "W" sur les contours et on compte le nombre de pixels contenus dans le pic de la distribution, ayant des valeurs de gris supérieures à la valeur à mi-hauteur. Une correction est faite pour les directions autres qu'horizontales et verticales. Dans le cas où la largeur de W obtenue est créée par deux contours proches, on modifie alors l'échelle, pour permettre la dissociation.

La distribution observée est vue comme le résultat de la convolution entre la ligne nette (paramètre d'étalement  $W_i$ ) et une fonction de paramètre d'étalement "B" de type gaussien; il vient alors :

$$W^2 = W_i^2 + B^2. \quad (24)$$

$W_i$  étant en général inconnu, l'auteur propose que  $W_i$  soit le même pour toute l'image. Enfin, La conversion de "B" en mesure de distance objet-captteur s'obtient par l'équation  $D_o = k B$ , où  $k$  est fonction des paramètres de la caméra (distance focale,...).

### 3.2.3. Approche de SUBBARAO

La méthode proposée par M. SUBBARAO [Subbarao-88] procède par l'acquisition d'une image nette et consécutivement de plusieurs images floues pour des distances positionnées de part et d'autre du plan net, prises dans les mêmes conditions d'éclairément. Les images ayant servi de références sont de taille 64x64 pixels et ne présentent qu'un front lumineux. Le calcul de la profondeur se fait à l'aide de la LSF calculée d'après une discontinuité floue d'ordre zéro.

Ainsi, pour un échelon lumineux équation (13), l'image observée est définie par :  $g(x, y) = \text{PSF}(x, y) \otimes f(x, y)$ . Pour caractériser la PSF, on utilise le modèle gaussien, pour lequel  $\sigma = k.s$  avec  $k > 0$ ; la valeur de  $k$  s'obtient par une phase de calibration.

Si l'on dérive la fonction  $g(x, y)$  suivant l'axe du gradient, il vient l'équation 25 :

$$\begin{aligned} \frac{\partial g(x, y)}{\partial x} &= \text{PSF}(x, y) \otimes \frac{\partial f(x, y)}{\partial x} \\ &= \text{PSF}(x, y) \otimes b \delta(x) \\ &= \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} \text{PSF}(\xi, \eta) b \delta(x - \xi) d\xi \right) d\eta \\ &= \int_{-\infty}^{+\infty} \text{PSF}(x, \eta) d\eta. \end{aligned} \tag{25}$$

Soit  $b \times \text{LSF}(x) = \frac{\partial g(x, y)}{\partial x}$ , on peut alors déterminer la LSF par la relation :

$$\text{LSF}(x) = \frac{\frac{\partial g(x, y)}{\partial x}}{\int_{-\infty}^{+\infty} \frac{\partial g(x, y)}{\partial x} dx} \tag{26}$$

Pour simplifier les calculs, nous prenons le paramètre d'étalement de  $\text{LSF}(x)$  comme la racine carrée du moment centré d'ordre deux de la distribution de  $\text{LSF}(x)$ , interprété comme la déviation standard et défini par l'équation 27 :

$$\sigma_1^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 \text{LSF}(x) dx, \quad (27)$$

où  $\bar{x} = \int_{-\infty}^{+\infty} x \text{LSF}(x) dx$  avec  $\int_{-\infty}^{+\infty} \text{LSF}(x) dx = 1$ .

La relation entre la LSF(x) et la profondeur est alors donnée à partir d'une relation entre s et  $\sigma_1$ , qui s'écrit :  $\sigma_1 = k_1 s$  d'où :

$$\sigma_1 = -\frac{k_1 a d}{u} + \frac{k_1 a (d-f)}{f} = \frac{m}{u} + c. \quad (28)$$

Les constantes m et c peuvent être déterminées par des techniques de calibration.

La relation entre  $\sigma_1$  de la LSF (x) et  $\sigma$  de la PSF est donnée par :

$$\sigma = \sqrt{2} \sigma_1. \quad (29)$$

On peut apporter une extension du calcul de la profondeur d'une scène contenant des objets simples, pour la phase de calibration, à partir d'images de format 512 \* 512 pixels. Cependant, la taille du masque utilisé doit être adaptée à la défocalisation.

### 3.2.4. Approche de JARVIS

L'approche présentée exploite comme mesure de profondeur le rapport des modules du gradient du laplacien des images nette et floue. L'étude est assortie d'essais sur des images réelles [Jarvis-90].

Pour déterminer la forme du contour, l'auteur utilise l'amplitude du gradient au passage par zéro du laplacien. Ainsi, si  $I_n$  représente l'image nette et  $I_f$  l'image floue, l'indication de défocalisation, après avoir appliqué un masque de type Laplacien sur une image adoucie par un filtre gaussien (G), est définie par :

$$R = \frac{|\nabla(\nabla^2 G \otimes I_n)|}{|\nabla(\nabla^2 G \otimes I_f)|}. \quad (30)$$

## 4. Approche par identification des contours

### 4.1. Introduction

Puisque le flou est perceptible, principalement sur les contours d'objets, notre méthode ne s'applique qu'aux points de contour de position connue. Nous retenons, comme Pentland, l'acquisition d'une image supplémentaire, nette en tout point avec un iris fermé, qui fournit plus facilement la position des contours. En faisant l'acquisition simultanément d'une image nette et d'une image contenant des parties floues en fonction de la profondeur, nous séparons le contenu informationnel de la scène. Une estimation, communément admise du flou est obtenue à partir du module des gradients de luminance. Pour donner la même importance à chaque contour, nous avons choisi de construire le rapport des modules des images nette et floue, afin d'identifier le paramètre d'étalement  $\sigma$  [ Simon-91].

La relation entre l'image nette  $in(x, y)$  et l'image floue  $if(x, y)$  est modélisée par :

$$if(x, y) = in(x, y) \otimes PSF(x, y). \quad (31)$$

En nous plaçant localement sur l'axe du gradient ( $\bar{x}$ ), nous obtenons une expression monodimensionnelle de son module :

$$|\bar{\nabla} in(x, y)| = |\bar{\nabla} in(\bar{x})|. \quad (32)$$

Le calcul du gradient net et flou est alors approximé aux différences finies :

$$G_{net}(\bar{x}) = \frac{|in(\bar{x} + 1) - in(\bar{x} - 1)|}{2} \quad (33)$$

$$G_{flou}(\bar{x}) = \frac{|if(\bar{x} + 1) - if(\bar{x} - 1)|}{2} \quad (34)$$
$$= \frac{\left| \sum_{u=-P}^P in(\bar{x} + 1 - u) LSF(u) - \sum_{u=-P}^P in(\bar{x} - 1 - u) LSF(u) \right|}{2}$$

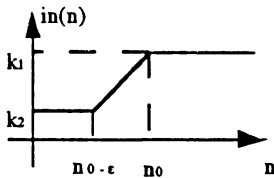
où nous considérons que le support de la réponse impulsionnelle est borné à  $[-P, +P]$ .

## 4.2. Identification du paramètre de la réponse impulsionnelle

Dans le but d'extraire localement le paramètre d'étalement de la LSF à partir de lignes de contours, nous établissons le rapport net sur flou des modules des gradients de deux images de la même scène, différentes dans leur netteté. Nous tenons compte, dans cette construction, des différences de quantité de lumière reçue sur le capteur, pour une zone d'analyse.

$$R(\bar{x}) = \frac{|\bar{\nabla} \text{in}(\bar{x})|}{|\bar{\nabla} (\text{in}(\bar{x}) \otimes \text{LSF}(\bar{x}))|}. \quad (35)$$

Dans la direction du gradient, nous trouvons une définition monodimensionnelle. Monga [Monga-90] a classifié ces contours en 5 groupes et pour la forme de contour la plus rencontrée (C4) dans une image, nous avons :



$$C(4) = \begin{cases} k_2 & -\infty < n \leq n_0 - \varepsilon \\ k_1 + \frac{k_1 - k_2}{\varepsilon} (n - n_0) & n_0 - \varepsilon < n < n_0 \\ k_1 & n_0 \leq n < +\infty \end{cases}$$

Figure 4 : contour en rampe

Equation (36).

Sur ces discontinuités et pour le modèle gaussien, le rapport prend des valeurs exploitables dépendant de leur écart  $\varepsilon$ . Ainsi, aux point  $n_0$  et  $n_0 - \varepsilon$ , nous obtenons :

$$R(n_0) = \frac{\sqrt{2\pi} \sigma}{\sum_{u=-\varepsilon+1}^{-1} u(A) + \varepsilon(B)}, \quad (37)$$

où

$$A = \exp\left(-\frac{(1-u)^2}{2\sigma^2}\right) - \exp\left(-\frac{(1+u)^2}{2\sigma^2}\right)$$

$$B = \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right) + \exp\left(-\frac{(1+\varepsilon)^2}{2\sigma^2}\right),$$

et :

$$R(n_0 - \varepsilon) = \frac{\sqrt{2\pi} \sigma}{\sum_{u=-\varepsilon+1}^{-1} u(C) + \varepsilon(D)}, \quad (38)$$

où

$$C = \exp\left(-\frac{(1-\varepsilon-u)^2}{2\sigma^2}\right) - \exp\left(-\frac{(1+\varepsilon+u)^2}{2\sigma^2}\right)$$

$$D = 1 + \exp\left(-\frac{1}{2\sigma^2}\right).$$

Dans la majorité des cas rencontrés,  $\varepsilon$  appartient à l'intervalle [1,3]. Ces valeurs entières conduisent à une forme analytique simple du rapport contenant le paramètre de la LSF, ce qui nous donne :

$$R(n_0 - 1) = R(n_0) = \frac{\sqrt{2\pi} \sigma}{1 + \exp\left(-\frac{1}{2\sigma^2}\right)} \quad (39)$$

$$R(n_0 - 2) = R(n_0) = \frac{\sqrt{2\pi} \sigma}{1 + \exp\left(-\frac{1}{2\sigma^2}\right) + \exp\left(-\frac{4}{2\sigma^2}\right)} \quad (40)$$

$$R(n_0 - 3) = R(n_0) = \frac{\sqrt{2\pi} \sigma}{1 + 2 \exp\left(-\frac{1}{2\sigma^2}\right) + 2 \exp\left(-\frac{4}{2\sigma^2}\right) + \exp\left(-\frac{9}{2\sigma^2}\right)} \quad (41)$$

En pratique, nous explorons, dans un voisinage de 4 points autour du contour, la valeur des gradients dans la direction de la discontinuité pour déterminer les valeurs de  $\varepsilon$ . Une fois  $\varepsilon$  connu, nous choisissons l'équation correspondante, pour trouver  $\sigma$ , qui donne la profondeur par la loi de conjugaison des lentilles minces (équation 3).

### 4.3. Etablissement de la correspondance flou-éloignement

Pour établir la relation flou-éloignement, il est important de maîtriser parfaitement le calcul du paramètre d'étalement. Pour cette raison, nous avons élaboré des images réelles et de synthèse, au format 512\*512 pixels avec 256 niveaux de gris, dans lesquelles le paramètre  $\sigma$  de la réponse

impulsionnelle varie simplement. Les images floues sont fabriquées à partir d'images nettes par convolution avec la PSF de forme gaussienne. On produit une variation linéaire de la profondeur, dans la direction verticale, ce qui entraîne une variation linéaire du paramètre d'étalement suivant la figure 5. Les résultats obtenus par notre algorithme montrent, sur la figure 6, que la mesure suit bien l'hyperbole donnée.

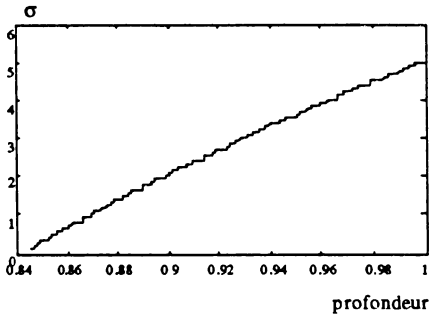


Figure 5 :  
variation linéaire de la profondeur

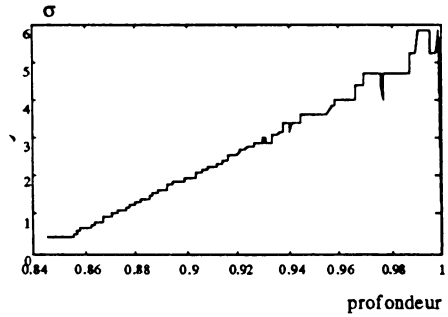


Figure 6 :  
correspondance flou-éloignement

Pour les images réelles nette et floue (figure 7a et 7b), la correspondance flou-éloignement est donnée par la figure 8 :

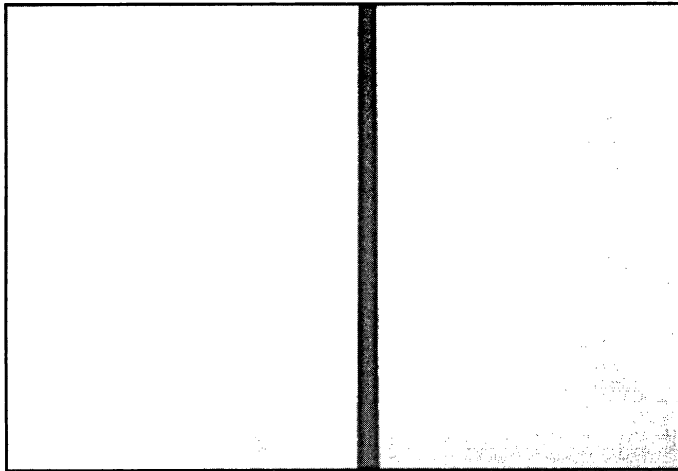


Figure 7-a : image réelle nette



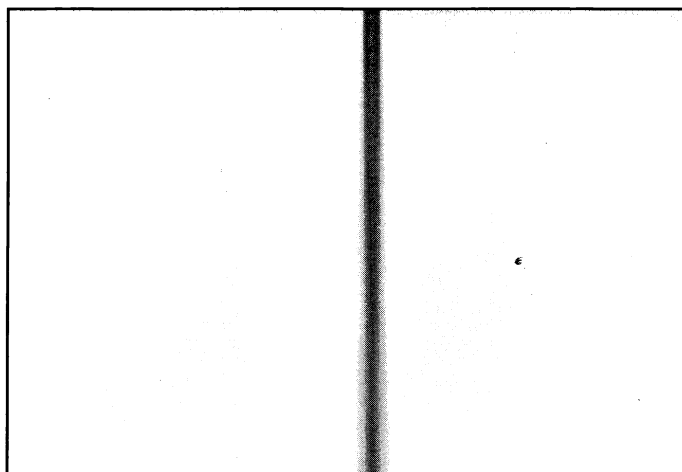


Figure 7-b : image réelle floue

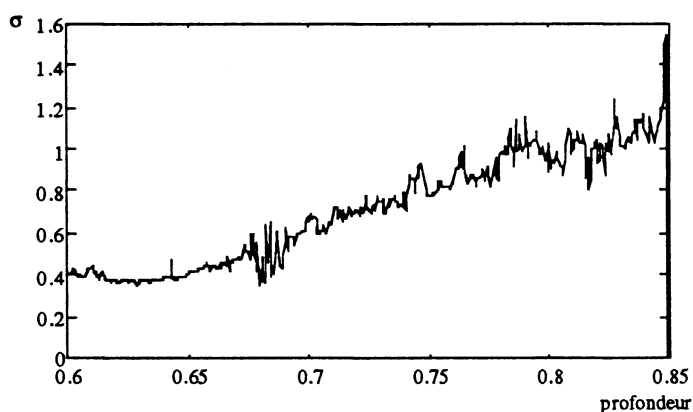


Figure 8 : image réelle, relation flou-éloignement

#### 4.4. Etalonnage

Le calcul de la correspondance flou-éloignement donne une courbe bruitée dans le cas d'images réelles (figure 8). Pour son utilisation, il est nécessaire de filtrer avant de rechercher la partie constante représentant la profondeur de champ et la partie hyperbolique modélisant le reste de la courbe.

Le bruit est provoqué par les conditions de prises de vues (non-uniformité de l'éclairage dans l'espace et le temps, imperfection optique, bruit électronique, etc ...) et par la numérisation (quantification, approximation des opérateurs, position des contours). Le bruit, extrait du signal de la vue réelle

(figure 8) peut être pris comme "gaussien" (figure 9). Son spectre est reproduit sur la figure 10.

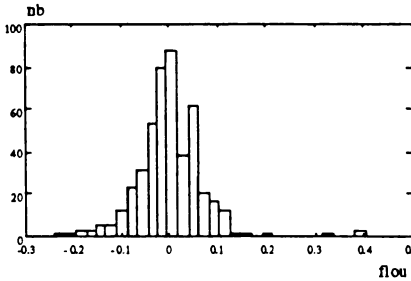


Figure 9 : histogramme du bruit

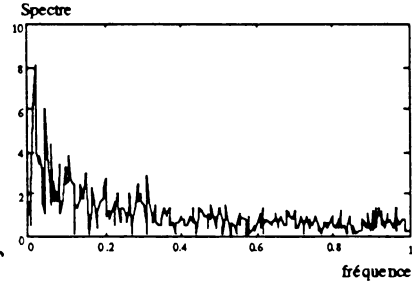


Figure 10 : spectre du bruit

Nous avons retenu le filtre médian lisseur, du fait qu'il élimine bien le bruit de type "gaussien", et respecte l'allure générale du signal, notamment les sauts de moyenne.

L'étalonnage est réalisé sur la scène décrite, qui varie continûment en profondeur, entre deux bornes inférieure et supérieure qui ne doivent pas constituer une limite de détection. Nous étendons la relation flou-éloignement au-delà de la borne supérieure, en modélisant cette relation par une fonction, dont l'expression est de type hyperbolique :

$$s = \frac{x}{D_0} + b \quad (37)$$

$$\sigma = k \cdot s \quad (38)$$

où  $s$  est la mesure de flou et  $D_0$  la profondeur.

La phase d'étalonnage se termine par des essais sur des scènes diverses contenant des profondeurs connues. La relation hyperbolique est modifiée en conséquence par le paramètre  $k$ .

#### 4.5. Schématisation du traitement

L'ensemble des opérations utilisées dans notre méthode de calcul de la profondeur peut être réalisé sur des calculateurs spécialisés en traitements d'images, pour une exécution rapide. Ces calculateurs spécifiques sont aujourd'hui très courants.

Nous calculons le gradient de façon identique sur les deux images. La procédure passe par deux phases, dont la première constitue l'apprentissage par initialisation de la table de transfert représentant la relation flou-éloignement, et la deuxième, l'estimation effective de la profondeur.

La figure 11 donne le synoptique du système de vision 3D où chaque bloc fonctionnel peut être réalisé par une carte électronique pour une conception modulaire de traitements d'images.

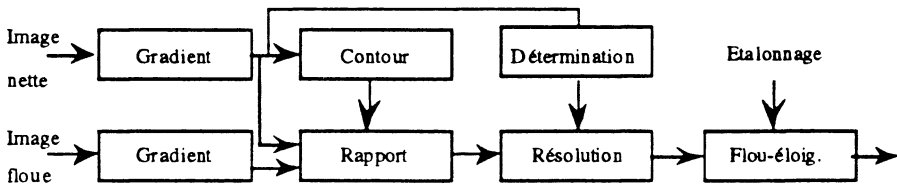


Figure 11 : synoptique de fonctionnement

#### 4.6. Résultats obtenus

Pour utiliser la relation flou-éloignement, nous devons effectuer un étalonnage, mais au préalable, nous devons choisir l'objectif du système optique en fonction de la scène. Nous procédons en deux étapes:

- choix de la focale pour l'observation du champ, vérification de la présence d'un plan net, des plans flous sans saturation du capteur, et de la possibilité d'avoir une image nette en tout point,
- apprentissage sur une scène simple, une ligne, et détermination de l'ouverture maximale et minimale.

Enfin, nous faisons l'acquisition des images nette et floue pour l'évaluation de la profondeur. Le plan net est placé le plus proche de la caméra.

Nous donnons, figure 12-a, une scène d'intérieur, dont la profondeur ne dépasse pas 4m.



Figure 12-a : scène d'intérieur, image nette



Figure 12-b : scène d'intérieur, image floue

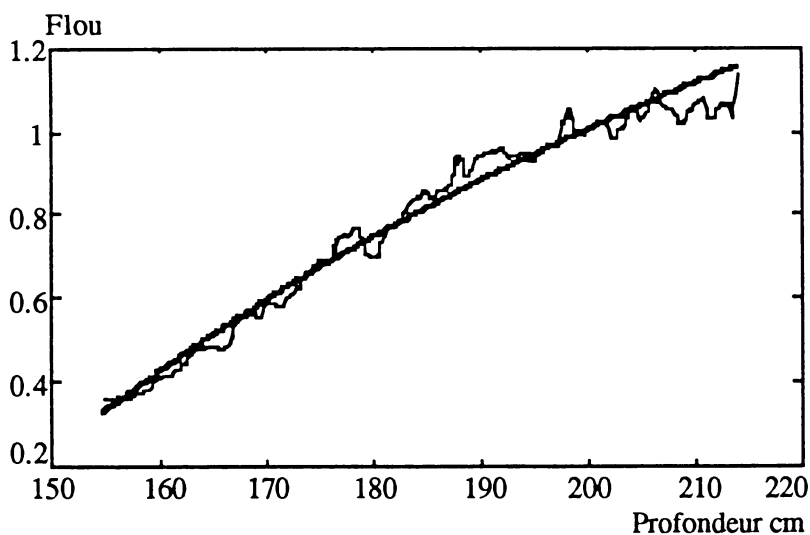


Figure 13 : relation flou-éloignement scène d'intérieur

Les paramètres de l'hyperbole obtenus sont:  $x=-464$  et  $b= 3.48$ .

Les résultats obtenus sont traduits en niveaux de gris et présentés dans la figure 14, qui donne une image de la carte des profondeurs.

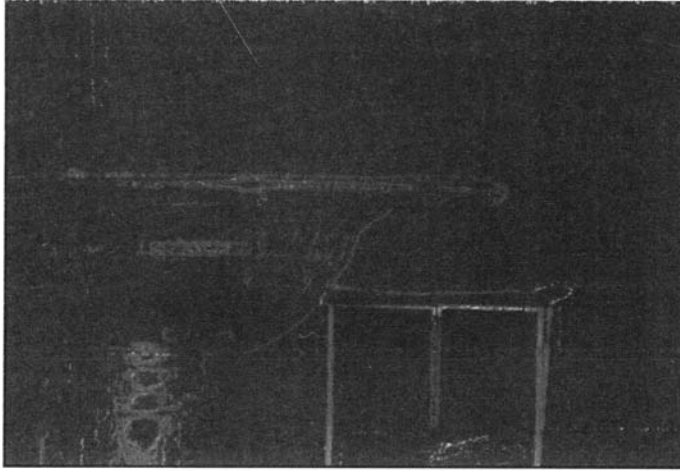


Figure 14 : carte des profondeurs

## 5. Conclusion et perspectives

Nous avons montré, dans cette étude l'intérêt et la réalité des techniques de vision 3D passive. La méthode développée et les résultats obtenus excluent, dans cette première approche, son application à la mesure précise d'objets. Les indications obtenues sont relatives et qualifient les positions des plans objets en fonction d'un nombre de zones de profondeur. Cependant, dans un contexte de vision embarquée ou de robotique mobile, elle trouve son efficacité dans la détection d'événements d'une scène. Ainsi, l'information de flou peut-être complémentaire pour un système classique de vision 3D à base de stéréo-vision en guidant le choix de l'appariement et contribuerait de façon directe au calcul de ressemblance des points potentiellement à appairer. Des conditions expérimentales plus précises, actuellement à l'étude, devraient nous conduire à une relation plus stricte.

### Références bibliographiques

[Grossman-87] P. GROSSMAN, "Depth from focus", Pattern recognition letters, North Holland, 1987, pp 63-68.

[Monga-90] O. MONGA, "Segmentation d'images: où en sommes nous?", rapport de recherche INRIA-Rocquencourt, N°1216, support de cours pour le congrès PIXIM 89, 1990.

[Pentland-82] A. P. PENTLAND, "Depth of scene from depth of field.", Proceeding of DARPA image understanding workshop, Paló Alto, 1982, pp 253-259.

[Pentland-87] A. P. PENTLAND, "A new sense for depth of field.", IEEE Transactions on pattern analysis and machine intelligence, 1987, pp 523-531.

[Pentland-89] A. P. PENTLAND, T. DARVELL, M. TURK, W. HUANG "A simple, real time range camera", IEEE computer Society, Conference on Computer Vision and Pattern Recognition, 1989, pp 256-261.

[Simon-91] T. SIMON, "Appréciation de la profondeur depuis des images nette et floue en vision 3D passive", Thèse de l'Université de Nancy I, juillet 1991.

[Subbarao-88] M. SUBBARAO, N. GURUMOOTHY, "Depth recovery from blurred edges.", IEEE computer society conference on computer vision and pattern recognition, Washington, 1988, pp 498-503.

[Jarvis-88] R.A.JARVIS, "Range from out of focus blur", AI'88, 2nd Australian Joint Artificial Intelligence Conference Proceedings, Adelaide, AUSTRALIA, 15-18 Novembre 1988, pp 347-342.

[Bove-89] M. BOVE, "Discrete fourier transform based depth from focus understanding and machine vision", Technical Digest Series, 1989, V14, pp 118-121.

[Born-70] M.BORN, E.WOLF, Principle of optics, Pergamon Press, Oxford, 1970.

[Schneider-93] G. SCHNEIDER, B. HEIT, T. SIMON, J. BREMONT, "Depth sensing by comparison of sharp and blurred images", Opto 93, 1993, pp 186-188.

[Krotkov-89] E.P. KROTKOV, "Active computer vision by cooperative focus and stereo", Springer-Verlag, New York, Berlin, Heidelberg, 1989.

[Lavest-91] J.M. LAVEST, G. RIVES, M. DHOME, "Utilisation d'un objectif à focale variable en vision monoculaire en vue de la reconstruction 3D", 8ème Congrès Reconnaissance des Formes et Intelligence Artificielle, AFCET, 1991, pp 293-301.

---

Bulletin des Académie et Société Lorraines des Sciences: 1993, 32, n°4  
-Communication présentée à la séance du 13 mai 1993