

Perception de la profondeur dans les images: Apport d'une approche comportementale

Bernard Heit, Christophe Simon

▶ **To cite this version:**

Bernard Heit, Christophe Simon. Perception de la profondeur dans les images: Apport d'une approche comportementale. 3e Conférence Internationale sur l'Automatisation Industrielle, Jun 1999, Montréal, Canada. pp.CDROM. hal-00331811

HAL Id: hal-00331811
https://hal.science/hal-00331811

Submitted on 17 Oct 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Perception de la Profondeur dans les Images : Apport d'une Approche Comportementale

B. Heit & C. Simon

CENTRE DE RECHERCHE EN AUTOMATIQUE DE NANCY
CRAN CNRS UPRES A 7039
Université Henri Poincaré, Nancy 1
BP. 239 – 54500 Vandoeuvre–lès–Nancy cedex – FRANCE
Tél. : ++33 383 50 33 72 / Fax. : ++33 383 50 16 32
Email : Bernard.Heit@cran.esstin.u-nancy.fr

Résumé

Nous proposons dans cet article un outil de perception de la profondeur dans des images pour des applications industrielles nécessitant l'utilisation de la troisième dimension. La technique est basée sur un modèle de connaissance des lois physiques de formation des images. La méthode est améliorée à l'aide d'une approche comportementale basée dans un premier temps sur un modèle flou d'experts, puis sur un modèle comportemental exprimé par des équations de relation floue spécifiquement optimisées.

Mots clés : Perception, Profondeur, Instrumentation

Abstract

A range sensing method based on monocular computer vision is presented. The technique exploits the physical effect that the imaging properties of an optical system depend upon the acquisition parameters and the object distance. The basic principle is to compare the blur in a pair of defocused images of the same scene taken with different apertures. Fuzzy logic was used to improve the performance of depth from defocus method. The determination of the fuzzy rules set was carried out by human experts. Thereafter, it was improve by a computerize learning method based on optimization algorithm.

Keywords : Perception, Depth, Instrumentation

1. Introduction

Les systèmes automatiques tels que les robots mobiles ou véhicules autonomes, les systèmes de contrôle ou d'assemblage de pièces évoluent dans un univers en trois dimensions qu'ils ont parfois besoin de percevoir. L'enjeu est important et les nombreux travaux de recherche en vision tridimensionnelle se heurtent à une grande difficulté de réalisation.

La mise en évidence de la troisième dimension diffère selon la nature de l'éclairage. On distingue deux grandes familles d'approches regroupées sous la terminologie de vision active et de vision passive. La vision active regroupe les méthodes où les scènes sont observées sous un éclairage particulier et contrôlé, dont les effets servent directement à la mesure. La vision passive, qui nous concerne plus particulièrement, regroupe les méthodes où l'information de profondeur est extraite directement de l'image, sans éclairage particulier. Cette information tridimensionnelle revêt des formes différentes selon que les images proviennent de systèmes monoculaires ou binoculaires. En vision monoculaire, objet de cet article, les données tridimensionnelles sont issues de plusieurs images de la même scène obtenues avec des paramètres de prises de vue différents comme l'ouverture, la longueur focale,...

La méthode adoptée est basée sur les techniques de « Depth From Defocus ». Elle évite la procédure lourde d'un grand nombre d'acquisitions en utilisant qu'un seul couple d'images plus ou moins défocalisées.

Notre technique est basée sur la quantification du flou optique localisé sur les contours. Pour cela, un modèle de connaissance du procédé de formation des images permet, après d'importantes simplifications, d'obtenir cette quantification. Néanmoins, le nombre et le type de ces hypothèses simplificatrices, ainsi que l'imprécision sur ces données ouvre le champ à l'utilisation d'un modèle comportemental et notamment par une approche en logique floue.

Dans la première partie de cet article, nous présentons l'approche par un modèle de connaissance à partir des lois physiques qui lient les variables du procédé, et décrivons le principe du « Depth From Defocus ». Dans la seconde partie, nous précisons une approche d'identification sur la base d'une étude phénoménologique qui recherche les relations entrées - sortie par une analyse sémantique, d'abord sous la forme d'un modèle comportemental expert, puis sous une forme optimisée. Enfin, dans les deux dernières parties, nous évaluons les performances des différents modèles appliqués à différentes images et concluons sur l'évolution de cette approche comportementale.

2. Modèle de connaissance

Le modèle de connaissance s'appuie sur les principes physiques de formation des images et des relations de l'optique géométrique. On modélise le système par une fonction de transfert établie entre la scène observée et sa représentation sur le plan image c'est-à-dire par un modèle

de réponse impulsionnelle. En analyse d'images, deux modèles de connaissance sont principalement utilisés. Le modèle dit idéal se base sur les lois de l'optique géométrique. Le deuxième modèle dit parfait, tient compte du phénomène de diffraction mais, en étant plus complet, il devient particulièrement complexe. Aussi, on préfère le modèle idéal pour sa simplicité et sa fidélité dans la traduction du principe de formation d'images fortement défocalisées [1].

Pour assurer correctement la formation d'une image sur un plan, plusieurs lentilles forment les objectifs photographiques. En ce qui nous concerne, nous assimilons un objectif à une lentille convergente mince, de longueur focale équivalente en respectant l'approximation de Gauss. On définit alors le système optique par la loi de conjugaison :

$$\frac{1}{f} = \frac{1}{d_o} + \frac{1}{d_i} \tag{1}$$

où f est la distance focale, d_o est la distance plan principal - objet, et d_i est la distance plan principal - plan image.

Pour 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 appartenant aux objets présents sur les autres plans formeront des taches sur le plan image. L'expression suivante traduit la relation entre le diamètre s de la tache et la distance D_o de l'objet au plan principal en fonction de l'ouverture a et de d_o [2][5].

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

Cependant, des limitations en précision sont dues à des considérations purement physiques concernant la profondeur de champ et le tassement du flou optique. Il faut donc choisir judicieusement l'objectif photographique en fonction de la scène à analyser. Par ailleurs, les aberrations géométriques et les effets de diffraction limitent la résolution dans le plan image.

Pour calculer la profondeur d'un objet dans la scène pour un diamètre de tache donnée, on inverse la précédente équation et on obtient pour $\Delta d_o = D_o - d_o$:

$$\Delta d_o = \begin{cases} \dfrac{(f - d_o) d_o s}{d_o s - sf + af}, & 0 \le s \\ \dfrac{(d_o - f) d_o s}{-d_o s + sf + af}, & 0 \le s \le \dfrac{af}{d_o - f} \end{cases} \tag{3}$$

L'ambiguïté sur Δd_o est levée par le choix d'une valeur minimum ou maximum selon la scène à observer.

Les relations entre l'objet et l'image deviennent plus facilement exploitables si l'instrument optique peut être modélisé par un système linéaire. On décompose la fonction d'entrée en représentant les points de l'objet en une infinité de fonctions de Dirac pondérées par la luminosité. Le système optique transforme ce signal d'entrée en un signal sur le plan image à l'aide d'une fonction de transformation T :

$$\begin{aligned} g'(x,y) &= T\left\{ \iint_E g(x,y)\delta(x - x_0, y - y_0) dx_0 dy_0 \right\} \\ &= g(x,y) \otimes PSF(x,y) \end{aligned} \tag{4}$$

Cette fonction de transformation correspond à la réponse impulsionnelle du système optique ou PSF.

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 prises de vues (distance de l'objet, diaphragme, éclairage). Une forme connue de la PSF est la forme gaussienne 2D. L'écart type de cette gaussienne est proportionnel au diamètre de la tache floue sur le plan image.

L'extraction de l'information de profondeur est réalisée par l'identification du paramètre de la PSF pour chaque point d'intérêt dans une image défocalisée à l'aide de deux ou de plusieurs images de la même scène. Les points d'intérêt sont les points de contour de position connue que l'on obtient aisément à l'aide d'une image nette. L'estimation du flou est obtenue à partir du module des gradients de luminance sur les différentes images. Pour donner la même importance à chaque contour, on construit le rapport des modules des images nette et floue afin d'identifier le paramètre d'étalement de la PSF. En nous plaçant localement sur l'axe du gradient, nous obtenons une expression mono - dimensionnelle de son module :

$$\left|\vec{\nabla} g(x,y)\right| = \vec{\nabla} g(\bar{x}) \tag{5}$$

Il ne nous reste plus qu'à établir la correspondance entre le paramètre d'étalement et l'éloignement effectif des objets, obtenu à l'aide d'une phase d'étalonnage [2].

3. Approche Comportementale du Depth From Defocus

Le modèle de connaissance proposé au paragraphe précédent peut être considéré comme approché suite aux hypothèses simplificatrices prises lors de son élaboration. La justesse de ce modèle est liée à l'estimation des différents paramètres et, est tributaire de la phase d'étalonnage. De plus, les aberrations du système optique font que les données sont entachées d'une certaine imprécision. Aussi, un modèle flou trouve ici un terrain d'application adapté.

3.1. Modèle flou d'experts

Une première approche pour valider l'utilisation d'un modèle flou consiste à établir un ensemble de règles à partir d'une description d'experts sur le processus d'estimation de profondeur. Ces règles traduisent le comportement du système et sont exprimées sous forme de syllogisme "Si x_1 et x_2 et x_3 alors y" ou les x_i sont les entrées et y sont les sorties. Le principe du Depth from Defocus peut être traduit sous cette forme logique. Les antécédents des règles sont le gradient de l'image nette, le gradient de l'image floue, la distance focale, et la distance caméra - objet est la conséquence de cet ensemble de relations [4].

Le gradient de l'image floue exprime l'information de profondeur. En conséquence, sa description doit être assez fine, et 4 termes linguistiques ont été choisis. L'univers de discours du gradient de l'image nette comporte 2 termes complémentaires. La distance focale est décrite par 3 termes linguistiques répartis sur l'univers de discours. Enfin, la profondeur comporte 5 termes linguistiques permettant d'obtenir une description précise de la sortie.

L'ensemble des règles proposées par les experts est donné dans le tableau suivant :

P	P	MP	MG	G
P	MP	MP	MP	MP
G	P	P	P	MP

M	P	MP	MG	G
P	M	MP	MG	MG
G	MP	M	M	M

G	P	MP	MG	G
P	MG	MG	G	G
G	M	M	MG	G

P = Petit ; MP = Moyennement Petit ; M = Moyen ; MG = Moyennement Grand ; G = Grand

Chaque matrice correspond à un terme linguistique particulier de la distance focale. Les lignes représentent les termes linguistiques du gradient de l'image nette et les colonnes représentent les termes linguistiques du gradient de l'image floue. L'information de profondeur sur chaque point d'intérêt de l'image doit être une valeur unique, et peut-être le résultat de plusieurs règles. Aussi, la méthode de défuzzyfication utilisée est le centre de gravité pour son caractère consensuel qui tient compte de toutes les propositions des différentes règles.

3.2. Modèle par équations de relation floue

Le modèle comportemental d'experts permet de vérifier l'applicabilité et la justesse de l'approche floue. Mais pour obtenir de meilleures performances, il est possible de l'améliorer en utilisant un système automatique de composition de règles pour lesquelles les paramètres sont optimisés. L'approche choisie est basée sur la résolution d'équations de relation floue [3]. Un système d'équation de relation floue prend la forme $R \circ X_c = Y$ où R est une matrice de valeurs définissant la validité des liens entre les observations floues X_c (gradient image nette, gradient image floue et distance focale) et les sorties floues Y (profondeur).

Les coefficients de la relation prennent des valeurs dans l'intervalle $[0,1]$ traduisant ainsi la validité de la règle qui est réalisée par une loi de composition de type T - conorme / T - norme. Les coefficients de la relation R peuvent être obtenus en minimisant l'erreur entre $R \circ X_c$ et Y, sous la contrainte $Def_y[F(y)] = y$. Le modèle flou est basé sur le produit cartésien. La loi de composition est obtenue par le produit et somme bornée/produit. Pour assurer la contrainte de bijection de la défuzzyfication, la partition floue de la sortie Y est limitée à deux termes triangulaires et la défuzzyfication est réalisée par le centre de gravité [6]. Les descriptions linguistiques des entrées restent les mêmes que pour le modèle expert.

La relation floue obtenue par cette méthode comprend deux lignes de 24 coefficients comme indiqué dans le tableau ci-dessous :

		Gradient Net * Gradient Flou * Distance Focale					
		P.P.P	P.P.G	P.MP.P		G.G.P	G.G.G
Profondeur	P	0.03	0.12	0		0	0.61
	G	0.67	0.59	0.22		1	0.96

Contrairement au modèle expert les valeurs de la relation ne permettent pas une interprétation linguistique stricte, mais assure une meilleure justesse du modèle [6].

4. Expérimentation

Pour évaluer les performances des différents modèles, nous avons utilisé une caméra de type commercial CCD 640*580. Deux images de la même scène obtenues avec des ouvertures différentes sont nécessaires afin d'appliquer les modèles. Actuellement l'image nette et l'image floue sont obtenues séquentiellement et les modifications potentielles d'intensité sont éliminées par un filtre de densité neutre pour l'image prise avec la plus large ouverture.

Une première scène est constituée de bandes de contraste à des profondeurs graduellement croissantes (Figure 1).

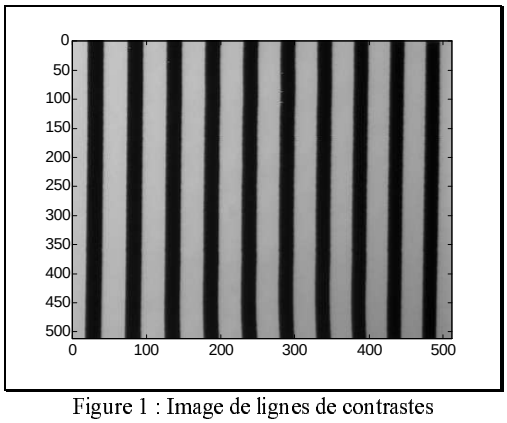


Figure 1 : Image de lignes de contrastes

La longueur focale choisie est de 60mm et pour éviter toute ambiguïté sur Δd_o, les images sont focalisées en arrière-plan. L'opérateur gradient est réalisé par un masque de Sobel de taille 3*3. L'application des trois modèles montre que l'erreur maximum relative obtenue avec le modèle flou optimisé est de 1.5%, elle passe à 3% avec le modèle expert. Le modèle de connaissance ne donne de bons résultats qu'avec des profondeurs proches, les performances s'estompant très rapidement avec l'accroissement de la profondeur (Figure 2).

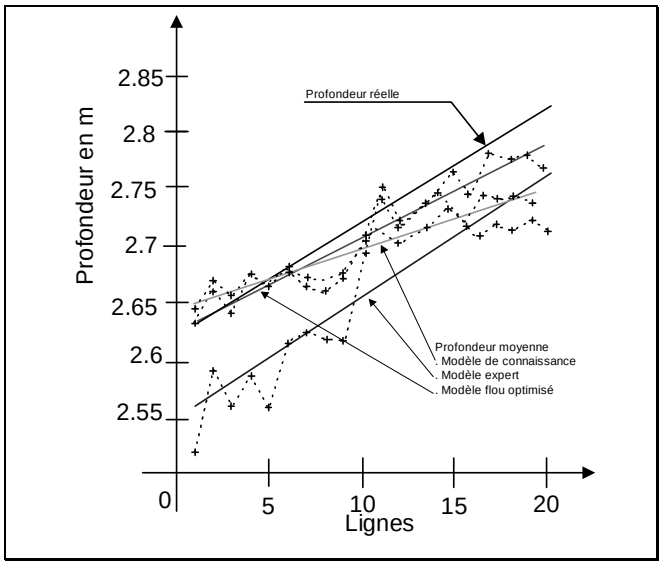


Figure 2: Evaluation de la profondeur par les trois modèles

Une seconde scène est constituée de cubes placés à des profondeurs différentes sur un fond noir. L'erreur maximum obtenue est alors de 3.5% (Figure 3a, 3b).

5. Conclusion

L'approche décrite se veut une première technique d'un dispositif simple permettant d'accéder à la troisième dimension pour des problèmes industriels aussi bien dans le domaine du contrôle produit et équipement, de la robotique mobile, que de la sécurité. Ce dispositif monoculaire peu encombrant, peut être embarqué sur de petits systèmes. En revanche, le développement du système optique de défocalisation reste ouvert.

L'approche par modèle flou optimisé permet de s'affranchir en partie des particularités du système optique, ce qui rend l'outil plus juste et précis. Néanmoins, comme le modèle de connaissance, il demande une phase d'estimation liée aux conditions de prise de vue, ce qui peut réduire son champ d'action à des scènes ne nécessitant pas de modifications de paramètres optiques.

En terme de temps de calcul, il est particulièrement rapide et peut être implanté sur processeur pour des applications en temps réel.

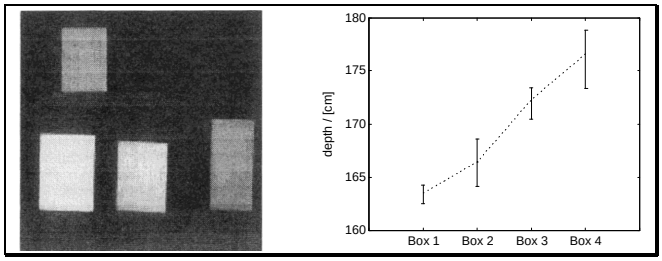


Figure 3 : Image de Cubes (a) ; Evaluation des profondeurs (b)

6. Références

[1] A.P. Pentland, « A new sense for depth of fields », IEEE Transaction on Pattern Analysis and Machine Intelligence, 9(4) :523-531, July 1987.

[2] T. Simon, « Appréciation de la profondeur depuis des images nette et floue en vision 3D passive », Thèse de Doctorat en Automatique, CRAN, Université Henri Poincaré Nancy 1, 1991.

[3] A. Voisin, L. Rondeau, E. Levrat, J. Brémont, « An interpretation of a fuzzy relational model derived from data », in Computational Engineering in Systems Applications (CESA 96 IMACS Multiconference), Lille, France, July 9-12, 1996.

[4] Hönig J., « Visual Extraction of shape characteristics on the basis of the fuzzy sets theory », Phd thesis, Universität Karlsruhe, Institut für Elektrotechnik und Mestenichnik, 1997.

[5] Schneider G., « Numerical analysis of monocular depth sensing principles and their applications in robotics », Phd thesis, Universität Karlsruhe, Institut für Elektrotechnik und Mestenichnik, 1995.

[6] Rondeau L., « Identification des systèmes par modèles flous linguistiques. Prise en compte des aspects numériques et symboliques », Thèse de Doctorat en Automatique, CRAN, Université Henri Poincaré, Nancy 1, 1997.