

# Estimation de profondeur par caméra monoculaire

Jean Louchet\*†, Peter Villaert\*

\*Vision Systems Laboratory, Hogeschool Gent (Belgique) †Artenia (France)

jean.louchet@gmail.com

## Résumé

Dans le cadre du projet 3-SIS de caméra intégrée, cet article décrit un algorithme de stéréo mono-objectif faisant appel à un réseau de micropismes placé sur le capteur, et permettant d'entrelacer des images provenant de deux zones opposées de la pupille de l'objectif. La base stéréo étant limitée par le diamètre de pupille, les disparités mesurables sont très faibles; leur limitation à  $\pm 1$  pixel permet de les exploiter à l'aide d'un algorithme local n'impliquant que des pixels voisins, ce qui est bien adapté à un traitement SIMD au niveau du capteur. Une corrélation subpixel donne une information de profondeur peu précise mais quasi-dense à un coût réduit, ce qui constitue un indice supplémentaire dans le processus de segmentation d'image.

**Mots-clés:** stéréo monoculaire, capteurs, rétines.

## Abstract

In the framework of the 3-SIS project, this paper describes a single-lens stereo algorithm. Using microprisms on the sensor delivers a chessboard-like interlaced stereo pair by exploiting opposite areas of the lens pupil. As the stereo base is limited by lens aperture, suitable dimensioning will provide disparities between  $-1$  and  $1$ , which can be measured by only using local processing involving direct pixel neighbours, which is well suited to the parallel image processing primitives. Subpixel correlation allows low-cost, low precision but quasi-dense range information, which may be exploited as an extra clue into the image segmentation process.

**Keywords:** stereo, single-lens stereo, sensors, retinas.

## 1 Introduction

Les algorithmes classiques de stéréovision fournissent à partir d'au moins deux caméras, des résultats 3D précis [1], mais à un coût élevé. Une autre voie consiste à exploiter le *bokeh* (flou de mise au point) pour en déduire la profondeur à partir d'une seule caméra [2][3]. Or le flou de mise au point provient de l'intégration sur le capteur des contributions de l'ensemble de la surface de la pupille de l'objectif, ce qui est une perte d'information.

La technologie des micropismes et microlentilles [4], bien connue pour les systèmes de visée des caméras reflex, permet de séparer en les entrelaçant les contributions de différentes aires de la pupille. Elle a été transposée avec succès depuis les années 2000 sur les capteurs [6], notamment afin de compenser la forte directivité des capteurs. Dans cet article, nous montrons comment elle peut être exploitée sur le capteur d'une caméra intégrée, combinée à un algorithme parallèle qui permet d'extraire une information 3-D d'une image unique.

Dans le prototype de la caméra 3-SIS [5] nous nous intéressons à un canal muni d'une optique de focale  $f = 30$  mm, de nombre d'ouverture  $\omega = 7$  et d'un capteur  $480 \times 480$  donnant une distance hyperfocale [6] de  $12,8$  m. Les micropismes placés sur le capteur per-

mettent d'obtenir un entrelacement en damier des images provenant de deux zones opposées G et D de la pupille.

G	D	G	D
D	G	D	G
G	D	G	D
D	G	D	G

Fig. 1 - Disposition des sous-images entrelacées

En pratique, la distance entre les centres des zones de pupille exploitées par les prismes ne pouvant être aussi grande que le diamètre de pupille, nous considérerons l'hyperfocale (et donc la base stéréo équivalente) de l'ordre de  $10$  m. Cette valeur est très inférieure à ce qui est habituellement utilisé dans les systèmes stéréo qui disposent de plus grandes bases. En revanche, nous allons montrer que le fait de ne disposer que de disparités très faibles (typiquement entre  $-1$  et  $+1$  pixel) permet un calcul local parallèle rapide intégré au capteur, et donc d'obtenir à faible coût une information 3D. La faible valeur de la base stéréo (et donc de l'hyperfocale), habituellement pénalisante, peut être ici exploitée comme un avantage.

## 2 Entrelacement et dimensionnement

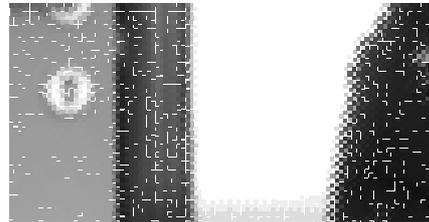


Fig. 2 - Détail de l'entrelacement E-W des images gauche et droite fournies par le réseau de microlentilles

Une mise au point à l'infini ne donnerait que des disparités positives. Dans notre cas il est plus adapté de traiter des disparités de valeurs faibles donc centrées autour de zéro. Si nous focalisons l'optique à l'hyperfocale  $h$ , alors la disparité sera, selon la profondeur  $z$  de l'objet:

$$-1 \text{ pour } z = \infty; \quad 0 \text{ pour } z = h; \quad 1 \text{ pour } z = h/2$$

Dans un cas plus général, si l'on s'intéresse à une fourchette de distances de  $z_{min}$  à  $z_{max}$  ( $z_{min} < z_{max}$ ), nous pouvons diviser l'espace entre les profondeurs  $z_{min}$  et  $z_{max}$  en trois tranches correspondant aux disparités  $-1$ ,  $0$  et  $1$  aux centres de ces tranches. Pour cela il faut régler le tirage de façon à focaliser sur la profondeur:

$$z_0 = \frac{2 z_{min} z_{max}}{z_{min} + z_{max}} \quad (1)$$

La disparité  $1$  sera atteinte à la profondeur  $z_1$  telle que  $\frac{6}{z_1} = \frac{1}{z_{max}} + \frac{5}{z_{min}}$  i.e.  $z_1 = \frac{6 z_{min} z_{max}}{z_{min} + 5 z_{max}}$ , et la disparité  $-1$  à la profondeur  $z_{-1}$  telle que  $\frac{6}{z_{-1}} = \frac{5}{z_{max}} + \frac{1}{z_{min}}$  i.e.  $z_{-1} = \frac{6 z_{min} z_{max}}{5 z_{min} + z_{max}}$ .

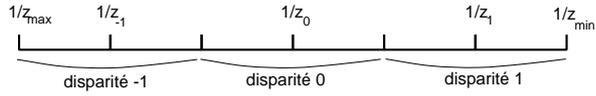


Fig. 5. Disparités sur une échelle en  $1/z$

Ces valeurs contraignent le choix de l'ouverture de pupille: l'hyperfocale  $h$  doit être telle que  $\frac{1}{h} = \frac{1}{z_1} - \frac{1}{z_0}$ ,

$$d'où h = \frac{3 z_{min} z_{max}}{-z_{min} + 4 z_{max}}$$

Pour une distance focale  $f$  et un pas interpixels  $p$ , comme  $h = f^2 / \omega p$ , le nombre d'ouverture  $\omega$  devra être:

$$\omega = \frac{f^2}{p} \frac{4 z_{max} - z_{min}}{3 z_{min} z_{max}}$$

et dans la pratique, légèrement inférieur (plus grand diamètre de pupille) pour des raisons de réalisation.

### 3 Traitement et interpolation

Avec la disposition en damier, lorsque  $x + y$  est pair, le pixel  $f(x, y)$  contient la valeur  $R(x, y)$  de l'image droite. Il doit être mis en correspondance avec l'une des trois valeurs de l'image gauche:  $L(x - 1, y)$ ,  $L(x, y)$ ,  $L(x + 1, y)$ ; mais parmi elles,  $L(x, y)$  doit être obtenue par interpolation locale alors que les deux autres sont directement contenues dans l'image damier:

$$L(x - 1, y) = f(x - 1, y)$$

$$L(x, y) \approx (f(x, y + 1) + f(x, y - 1)) / 2$$

$$L(x + 1, y) = f(x + 1, y)$$

	a	
b	x	c
	d	

Fig. 3 - Le pixel  $x$  de l'image gauche, et ses 4 voisins  $a, b, c, d$  qui appartiennent à l'image droite. La valeur de  $x$  sera comparée à  $b, \frac{a+d}{2}, c$  pour déterminer si la disparité vaut  $-1, 0$  ou  $+1$ .

Une manière plus raffinée de calcul de la disparité, qui n'implique pas de lourds calculs ni de changements matériels, consiste à ajuster une parabole aux valeurs  $a, b, c$  pour réaliser une interpolation subpixel de la position du minimum de la fonction  $f(x) = Ax^2 + Bx + C$ . Si l'on écrit:

$$f(-1) = |b - x|$$

$$f(0) = \left| \frac{a + d}{2} - x \right|$$

$$f(1) = |c - x|$$

alors la position du minimum donne la disparité:

$$disp = \frac{-B}{2A} = \frac{f(-1) - f(1)}{2(f(-1) + f(1) - 2f(0))}$$

Les valeurs de disparité sont validées entre  $-2$  et  $+2$ , puis lissées itérativement à l'aide des valeurs de disparité valides des 4-voisins.

### 4 Résultats

La figure 4 montre une image originale simulée, le ré-

sultat de l'algorithme après le lissage des paramètres de la parabole, et les histogrammes des profondeurs trouvées sur 4 zones de l'image.

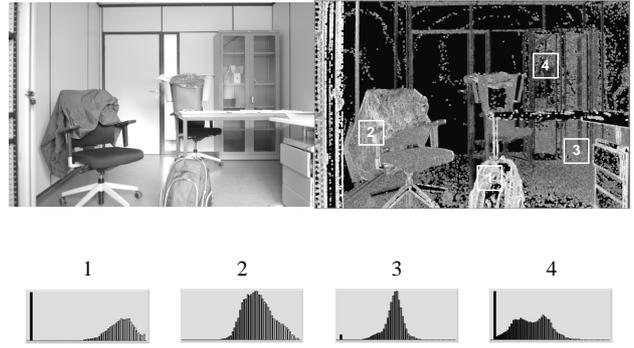


Fig. 4. Image originale (entrelaçage en damier simulé), résultat de l'algorithme et histogramme des profondeurs détectées sur 4 zones de l'image. Sur la zone 4, le pic secondaire correspond à une réflexion sur la vitre.

### 5 Conclusion

La méthode présentée ici permet, par addition d'un réseau de microlentilles au capteur, d'ajouter une fonctionnalité 3D fruste mais économique, pouvant contribuer à la segmentation ou à la détection d'obstacles en robotique. La perte de sensibilité est analogue à celle induite par les filtres de Bayer. Le traitement est par nature purement local et peut trouver sa pleine utilité avec des processeurs SIMD tels que les rétines électroniques. Une disposition plus complexe des microlentilles devrait permettre d'exploiter dans des conditions analogues 3 ou 4 zones de la pupille de l'objectif, afin d'exploiter une information plus riche analogue à la stéréo tri-oculaire,

### Références

- [1] R. Mohr, B. Triggs, Projective Geometry for Image Analysis, tutorial, ISPRS, 1996.
- [2] P. Favaro, S. Soatto, M. Burger, and S. Osher, Shape from defocus via diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(3):518-531, 2008.
- [3] Steffen Lösch, Depth from Blur: combining image deblurring and depth estimation, Bachelor's thesis, Saarland University, 2009.
- [4] Hyunsik Yoon, Sang-Guen Oh, Dae Shik Kang et al. Arrays of Lucius microprisms for directional allocation of light and autostereoscopic three-dimensional displays, Nature, 30 August 2011.
- [5] G. Yizengaw Belay, Y. Meuret, H. Ottewaere, P. Veelaert, Optical design of a multi-channel, multiresolution imaging system, Proc. SPIE, Vol. 8429-Optical Modelling and Design, 8429-10, May 2012.
- [6] J. Louchet, Capteurs et dimensionnement physique en vision, tutorial, RFIA2012 Reconnaissance des Formes et Intelligence Artificielle, Lyon janvier 2012.

### Remerciements

Je remercie l'équipe des laboratoires Vision Systems de Hogeschool Gent et IPI de Universiteit Gent. Ce projet a été financé par IWT (Flanders Agency for Innovation by Science and Technology), Belgique.