

# Une caméra de profondeur plénoptique basée sur 4 mini-objectifs

## A four-lens based plenoptic camera for depth measurements

Cécile Riou<sup>1</sup>, Bruno Colicchio<sup>1</sup>, Jean-Philippe Lauffenburger<sup>1</sup>, Christophe Cudel<sup>1</sup>

<sup>1</sup>Laboratoire MIPS

Université de Haute Alsace,  
61 rue Albert Camus, 68093 Mulhouse Cedex, France  
cecile.riou@uha.fr

### Résumé

Ces travaux présentent un dispositif d'acquisitions multi-vues, modélisé au moyen de l'homographie variable. Cette caméra qui capture plusieurs vues sur le même capteur image, peut être considérée comme une caméra plénoptique. Des travaux précédents [1][2] ont montré les avantages de l'homographie variable dans un cadre de métrologie par vision. Nous avons étendu le modèle à cette caméra permettant ainsi une approche alternative à la modélisation de certaines caméras plénoptiques. Nous avons évalué l'incertitude de mesures de ce système au moyen de mesures ponctuelles et nous montrons que cette incertitude est inférieure à 5%, sur l'axe Z, pour une distance de travail de l'ordre du mètre. Nous montrons également que l'approche développée peut facilement s'interfacer avec une carte de disparité, pour transformer cette caméra plénoptique en caméra de profondeur.

### Abstract

This work presents a "multi-view" acquisition device, modelled using the variable homography. This camera, which captures several views on a unique sensor, can be considered as a plenoptic camera. Previous works [1][2] have shown the advantages of variable homography in a context of metrology by vision. We have extended this model to our multi-view camera, allowing thereby for an alternate approach for the model of some plenoptic cameras. The measurement incertitude of this system has been evaluated by punctual measurements and this incertitude is inferior to 5% on the Z axis, for a metre-scale working distance. The developed approach can easily interface with a disparity map, to transform this plenoptic camera into a depth camera.

### Mots Clef

Caméras plénoptiques, homographie, mesure 3D, caméra de profondeur

### Keywords

plenoptic cameras, homography, 3D measurements, depth camera

### 1 Introduction

En 1908, Lippmann [4] a posé les bases d'un nouveau type de photographie en capturant les rayons lumineux d'une scène au moyen d'un réseau de sténopés. L'intérêt pour cette idée n'a cessé de grandir ces dernières années, et les caméras « light-field » ou plénoptiques sont devenues des technologies très prometteuses dans le domaine de la photographie, mais aussi dans celui de la vision 3D. En utilisant plusieurs microlentilles ou un réseau de microlentilles, les données brutes capturées par ces dispositifs contiennent des informations de parallaxe et de focalisation. Même si les données brutes délivrées par la caméra plénoptique ne fournissent pas une image directement interprétable, il est possible d'accéder à la profondeur d'une scène. On peut identifier 3 approches :

- A partir des paramètres de refocalisation numérique : l'image brute permet de calculer des images pouvant être refocalisées par rapport à n'importe quel plan de la scène [5]. Connaissant le processus de formation d'images sur un capteur, il est possible d'établir des liens très étroits entre focalisation et orientation des rayons lumineux. De ce fait, une carte de disparité ou une image de profondeurs peut être établie avec les paramètres de refocalisation.
- A partir de l'information de parallaxe contenue dans les images de « sub-apertures ». Ces images de « sub-apertures » sont les différentes vues de la scène depuis les positions de chacune des microlentilles. Elles sont utilisées pour générer des

images dites « épipolaires » (EPI) [6]. L'intérêt de ces EPI est de mettre en évidence la parallaxe, en analysant la pente formée par les contours d'un même objet.

- Une méthode alternative consiste à travailler directement sur les données brutes et à calculer l'image de profondeur en utilisant des méthodes de triangulation [7][8]. Les concepts développés sont alors relativement similaires à ceux utilisés en stéréovision avec deux caméras parallèles.

Ces trois approches restent vraies pour les nombreux prototypes de caméras plénoptiques ou pour les modèles commerciaux proposés par Lytro© et Raytrix©. On peut mentionner que la plupart des caméras plénoptiques contiennent des microlentilles, mais que quelques-unes sont aussi constituées d'un réseau dense de caméras, comme la caméra ProFusion (PTGrey©) ou le réseau Multi-Camera de l'Université de Stanford [9].

Quelque soit le modèle de caméra, l'accès à la profondeur demande une étape de calibrage. Par exemple, on peut citer la référence [10] qui s'intéresse au calibrage des caméras plénoptiques 1.0, où la grille de microlentilles est placée dans le plan image de la lentille principale.

Les travaux présentés dans cette communication s'inscrivent dans une démarche visant à automatiser le calibrage des caméras plénoptiques, en s'appuyant sur des images de « sub-apertures ».

Dans des travaux précédents [1][2], nous avons étendu les principes de l'homographie variable, définie par Zhang et Greenspan [3], à la mesure de la hauteur de fibres émergentes sur des tissus en fibre de verre. Cette méthode a été proposée pour fonctionner avec des échantillons défilant sur un convoyeur. Deux images prises successivement étaient alors nécessaires pour réaliser la mesure de la hauteur de la fibre. Par la suite, nous avons réduit le nombre d'acquisition à une seule, en développant un système d'acquisition composé de quatre lentilles situées devant le capteur image. Ainsi, on obtient quatre sous-images projetées sur un unique capteur CCD. Le dispositif devient donc une caméra plénoptique, capturant plusieurs vues sur le même capteur image et fournissant directement les images dites de « sub-aperture » (Figure 1).

Dans ce travail, nous avons adapté la formulation de l'homographie variable à cette caméra plénoptique et nous proposons donc une nouvelle méthode pour

calculer la profondeur dans une scène, avec une méthode de calibrage simplifiée.

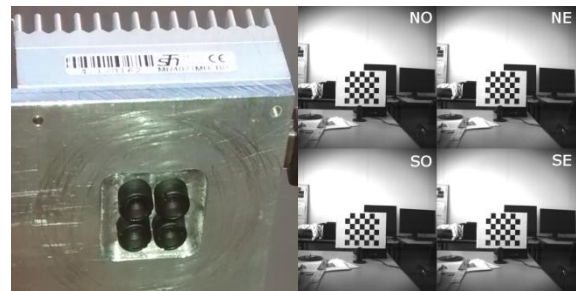


Figure 1 : (a) Caméra avec 4 mini-lentilles. (b) Sous-images correspondantes (Nord-Ouest, Nord-Est, Sud-Ouest et Sud-Est).

## 2 Homographie variable et mesures de profondeur

La figure 2 donne une représentation géométrique de notre système constitué de mini-lentilles et d'un plan image unique. Cette figure correspond à une section 2D représentant seulement deux projections voisines, numérotées  $i$  et  $i+1$  (on utilisera  $i=1$  pour les prochaines expressions). Avec quatre projections sur le capteur image, chaque sous-image a trois voisins possibles. Si l'index  $i$  représente la vue Nord-Ouest (NO), et l'index  $i+1$  peut être une vue de Nord-Est (NE), Sud-Ouest (SO) ou Sud-Est (SE). Ce schéma traduit ce dispositif comme un système de stéréovision à plans parallèles, pour lequel une triangulation pourrait être utilisée pour déterminer la profondeur du point  $P$  quand les paramètres intrinsèques et de distorsion sont connus. La situation idéale est rencontrée quand ces paramètres sont identiques pour chaque caméra, ce qui cependant n'est jamais le cas en pratique.

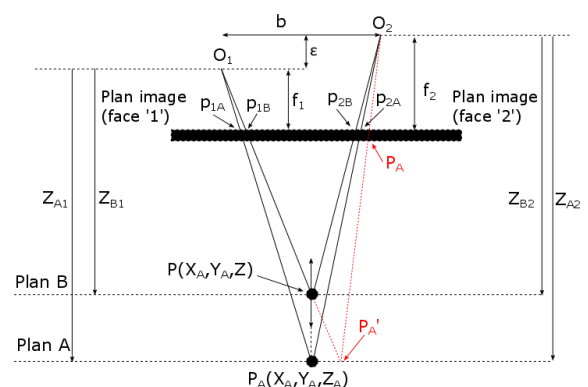


Figure 2 : Description géométrique du plan image des objectifs

Une approche alternative consiste à tirer profit de l'homographie variable pour modéliser ce dispositif. L'homographie variable introduit une matrice  $K$ , qui

fournit des propriétés intéressantes entre l'homographie  $H_A$  et  $H_B$  définies pour les plans parallèles A et B :

$$H_B(k) = K_i \cdot H_A \cdot K_{i+1}^{-1} \quad \text{Eq. 1}$$

$$\text{avec } K_{i/i+1} = \begin{pmatrix} k_{i/i+1} & 0 & (1 - k_{i/i+1})u_{0i/0i+1} \\ 0 & k_{i/i+1} & (1 - k_{i/i+1})v_{0i/0i+1} \\ 0 & 0 & 1 \end{pmatrix}$$

où  $u_{0i}$  et  $v_{0i}$  désignent les coordonnées des centres optiques projetés sur le capteur image. L'homographie variable introduit le ratio de distances  $k_{i/i+1}$  :  $k_{i/i+1} = Z_{Ai/Ai+1}/Z_{Bi/Bi+1}$ . On considère également le plan A comme plan de référence utilisé pour calibrer le dispositif.

A partir de la figure 2, trois relations homographiques entre projections réelles ou virtuelles sur la sous-image  $i+1$  peuvent être données :

$$\begin{cases} p_{2A} = H_A \cdot p_{1A} \\ p_{2B} = H_B \cdot p_{1B} \\ p_A = H_A \cdot p_{1B} \end{cases} \quad \text{Eq. 2}$$

En supposant que l'homographie variable résulte d'une matrice intrinsèque combinée à des matrices de rotation et de translation, nous pouvons exprimer  $H_A$  comme :

$$H_A = \begin{pmatrix} \frac{f_2 Z_{A1}}{f_1 Z_{A2}} & 0 & u_{0R} + \frac{b_u \cdot f_2}{Z_{A2}} - \frac{f_2 Z_{A1}}{f_1 Z_{A2}} u_{01} \\ 0 & \frac{f_2 Z_{A1}}{f_1 Z_{A2}} & v_{0R} + \frac{b_v \cdot f_2}{Z_{A2}} - \frac{f_2 Z_{A1}}{f_1 Z_{A2}} v_{01} \\ 0 & 0 & 1 \end{pmatrix} \quad \text{Eq. 3}$$

Les paramètres  $b_u$  et  $b_v$  sont respectivement les distances horizontales et verticales entre les centres optiques  $O_i$  et  $O_{i+1}$ . A partir de l'équation 1, l'homographie reliée au point P localisé dans le plan B est exprimée comme suit :

$$H_B = \begin{pmatrix} \frac{f_2 Z_{A1} k_2}{f_1 Z_{A2} k_1} & 0 & u_{02} + \frac{b_u \cdot f_2 \cdot k_2}{Z_{A2}} - \frac{f_2 Z_{A1} k_2}{f_1 Z_{A2} k_1} u_{01} \\ 0 & \frac{f_2 Z_{A1} k_2}{f_1 Z_{A2} k_1} & v_{02} + \frac{b_v \cdot f_2 \cdot k_2}{Z_{A2}} - \frac{f_2 Z_{A1} k_2}{f_1 Z_{A2} k_1} v_{01} \\ 0 & 0 & 1 \end{pmatrix} \quad \text{Eq. 4}$$

A partir de ces expressions, nous proposons de mesurer la profondeur d'un point P en mesurant la distance  $d$  définie par :

$$d = \text{dist}(p_{2B}, p_A) = \text{dist}(p_{2B}, H_A \cdot p_{1B}) \quad \text{Eq. 5}$$

Dans une première approximation, nous considérons que les distances  $f_1$  et  $f_2$  ont presque la même valeur. On peut montrer que cette approximation ramenée aux équations 3, 4 ou 5, revient à considérer que  $(1 + \varepsilon/f_1) = 1$ , ce qui est justifié vu les ordres de

grandeurs des distances focales. On peut alors simplifier l'expression de  $d$  :

$$d = \sqrt{\underbrace{\left( (k_2 - 1) \cdot \frac{c_u}{Z_{A2}} \right)^2}_{d_u: \text{ distance horizontale}} + \underbrace{\left( (k_2 - 1) \cdot \frac{c_v}{Z_{A2}} \right)^2}_{d_v: \text{ distance verticale}}} \quad \text{Eq. 6}$$

$$\text{avec } k_2 = \frac{Z_{A2}}{Z_{B2}} = \frac{Z_{A2}}{Z}$$

Cette relation montre que  $d$  est une fonction de la profondeur  $Z_{B2}$  et que les autres paramètres  $C_u$ ,  $C_v$  et  $Z_{A2}$  sont constants. En mesurant  $d$  sur les sous-images, nous pouvons donc utiliser l'équation 6 pour calculer la profondeur  $Z$  du point P.

En pratique, il est plus facile de séparer les distances horizontales et verticales pour obtenir  $Z$ . Par exemple, quand les sous-images sont verticalement alignées, comme NO-SO et NE-SE, c'est bien connu qu'il n'y a pas de parallaxe horizontale et que  $d_u$  est nul. Dans ce cas,  $Z$  peut être calculé à partir de  $d_v$ . Avec  $d_u$  ou  $d_v$ , la profondeur  $Z_{B2}$  peut être ainsi obtenue par :

$$Z_{B2} = \frac{C_{u/v}}{d_{u/v} + \frac{C_{u/v}}{Z_{A2}}} \quad \text{Eq. 7}$$

En supposant qu'un calibrage préliminaire détermine  $H_A$ ,  $C_u$ ,  $C_v$  et  $Z_{AR}$ , nous proposons un processus de mesure décrit par les étapes suivantes (résumées sur la figure 3) pour les points correspondants  $p_{1B}$  et  $p_{2B}$  :

- Sélectionner un point  $p_{1B}$
- Calculer le point (virtuel)  $p_A$  :  $p_A = H_A \cdot p_{1B}$
- Détecter le point  $p_{2B}$  le long de la droite épipolaire passant par  $p_A$ . Cette détection est une mise en correspondance de la paire de points ( $p_{1B}$ ,  $p_{2B}$ ). Ils ont donc des propriétés photométriques locales proches et obéissent à la géométrie épipolaire. Cette dernière condition peut être exprimé avec le produit vectoriel suivant :  $p_{2B} \times H_B \cdot p_{1B} = 0$ .
- Calculer la profondeur  $Z$  à partir de  $d_u$  ou  $d_v$

Les résultats intéressants de cette approche sont d'associer pour chacune des sous-images voisines, une homographie de référence  $H_A$  comprenant les paramètres intrinsèques et extrinsèques. Cela permet de ne pas calculer individuellement ces paramètres pour effectuer une mesure de profondeur, mais nous devons calculer  $H_A$ ,  $C_u$ ,  $C_v$  et  $Z_{A2}$ . Nous allons voir que cela peut s'effectuer avec une étape de calibrage

simple. Nous proposons deux méthodes pour calibrer la caméra plénoptique. La première est effectuée en considérant que des lentilles idéales sans distorsions optiques, tandis que la seconde est obtenue en modélisant les distorsions optiques.

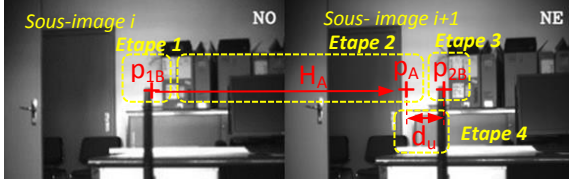


Figure 3 : Les 4 étapes de la mesure de profondeur

### Calibrage avec des lentilles idéales

Nous proposons un processus de calibrage simple ne demandant que deux acquisitions d'une mire de type « échiquier ». La première acquisition est effectuée à la profondeur de référence  $Z_{A2}$ . La seconde est faite à une profondeur  $Z_{B2} = Z_{A2} + dz$ . Ces deux acquisitions sont suffisantes pour déterminer les paramètres de calibrage.

- 1<sup>ère</sup> acquisition :  $H_A$

Les coins de l'échiquier sont détectés sur toutes les sous-images et utilisés pour calculer l'homographie  $H_A$  entre les sous-images. Nous notons  $d_A$  la distance en pixels entre le coin en haut à gauche et le coin en bas à droite.

- 2<sup>ème</sup> acquisition :  $Z_A, C_u, C_v$

Les coins sont également détectés et notons  $d_B$  la distance en pixels entre le coin en haut à gauche et le coin en bas à droite.  $Z_{A2}$  est obtenu par triangulation et peut être exprimée comme suit :

$$Z_{A2} = \frac{d_z \cdot d_B}{d_B - d_A} \quad \text{Eq. 8}$$

Cette seconde acquisition est aussi utilisée pour estimer les constantes  $C_u$  et  $C_v$ . Les coins détectés sur les deux images constituent un jeu de données des points  $p_{1B}$  et  $p_{2B}$ , permettant de calculer facilement  $C_u$  et  $C_v$  avec l'équation 7.

### Calibrage avec la correction de la distorsion des lentilles

Notre prototype de caméra plénoptique utilise des lentilles bas coût, présentant un coefficient de distorsion proche de 1,5% (valeur issue de la fiche technique de la lentille). Pour améliorer la précision de nos mesures, nous proposons de prendre en compte la distorsion dans notre processus de calibrage. En utilisant des approches standards, comme les corrections en utilisant des modèles de

distorsions radiales, la position précise des centres optiques de chaque sous-image doit être déterminée. Comme ces paramètres intrinsèques sont combinés dans  $H_A$ , ils ne sont pas directement accessibles dans notre processus de calibrage. Nous proposons donc une approche alternative. Nous considérons que les effets de la distorsion sont comparables à un facteur de grossissement qui serait non uniforme sur le plan image. Cette approche permet d'ajuster le calcul de  $Z_{B2}$  en fonction de la position spatiale des correspondants dans les sous-images. Ce qui peut être fait en utilisant la relation qui existe entre le grossissement, la distance focale et les paramètres  $C_u$  ou  $C_v$  (éq. 6). Cette correction est effectuée en utilisant l'opérateur  $\Gamma$ , qui transforme le vecteur représentant le point P de coordonnées (u,v,1) en « shifted coordinates » [11]  $\tilde{P}$  :

$$P(u, v, 1)^T \rightarrow \tilde{P} = \Gamma(P, P) \quad \text{Eq. 9}$$

$$\tilde{P} = (u^2 \quad v^2 \quad u \cdot v \quad u \quad v \quad 1)^T$$

En utilisant  $p_{B2}$  comme position de référence,  $C_u$  et  $C_v$  sont alors calculés comme les fonctions quadratiques définies par :

$$\begin{pmatrix} C_u \\ C_v \end{pmatrix} = \begin{pmatrix} a_u & b_u & c_u & d_u & e_u & f_u \\ a_v & b_v & c_v & d_v & e_v & f_v \end{pmatrix} \cdot \tilde{p}_{B2} \quad \text{Eq. 10}$$

Les coins collectés durant le calibrage présenté précédemment, constituent un jeu de données suffisamment bien répartis dans l'image, pour calculer les coefficients polynomiaux en utilisant une méthode des moindres carrés.

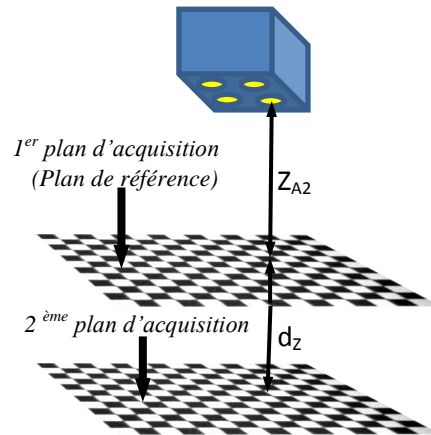


Figure 4 : Calibrage avec deux acquisitions

## 3 Mesures expérimentales

Les premiers résultats expérimentaux (Figures 5 et 6) sont obtenus en utilisant un banc de mesure, où nous avons fait reculer un échiquier par pas de 1 cm entre 2 acquisitions. La figure 5 représente le

contexte expérimental. La profondeur reportée sur la figure 6 est celle calculée sur le premier coin détecté pour chaque acquisition. La courbe obtenue est très régulière et est conforme à la profondeur  $Z_{B2}$  donnée par l'équation 7.

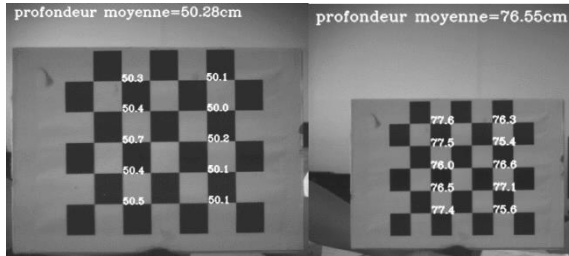


Figure 5 : Illustration des mesures expérimentales pour tracer  $Z_{B2}=f(d_u)$ .

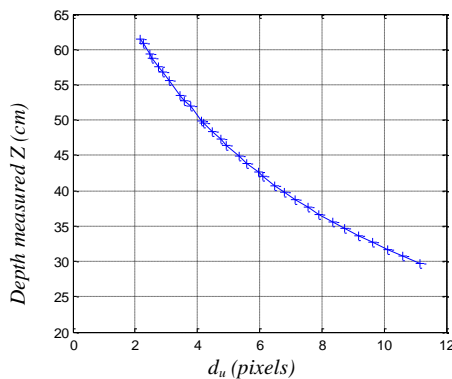


Figure 6 : Mesures expérimentales de  $Z_{B2}=f(d_u)$ . La courbe obtenue suit bien l'expression de l'éq. 7

Une seconde expérimentation utilise également un échiquier et un détecteur de coins. Ici, la distance entre l'échiquier et la caméra est constante et près de mille points de mesures sont collectés pour tester la répétabilité de la méthode. L'échiquier a simplement été déplacé dans le plan (x,y). Nous comparons les deux méthodes de calibrage sur trois gammes de profondeur. La distance caméra-échiquier est successivement fixée à 4 cm, 33 cm et 88,5 cm. Les profondeurs mesurées correspondent à la distance séparant l'échiquier et le centre optique de la lentille  $i+1$ .

En travaillant à courte distance, les images capturées présentent des importantes distorsions optiques. A cette distance, la seconde méthode offre une amélioration significative, fournissant un écart-type inférieur à 1%. Cette seconde méthode est aussi capable de corriger la valeur moyenne des valeurs mesurées. Pour des distances plus grandes, l'amélioration est moins significative, mais l'écart-type reste inférieur à 1% à 33 cm et inférieur à 5% à 88,5 cm. Ces résultats confirment l'intérêt de l'homographie variable comme principe de mesures

3D. Au final, le facteur principal qui influencera la qualité de ces mesures sera la précision de la mise en correspondance entre les points des différentes images.

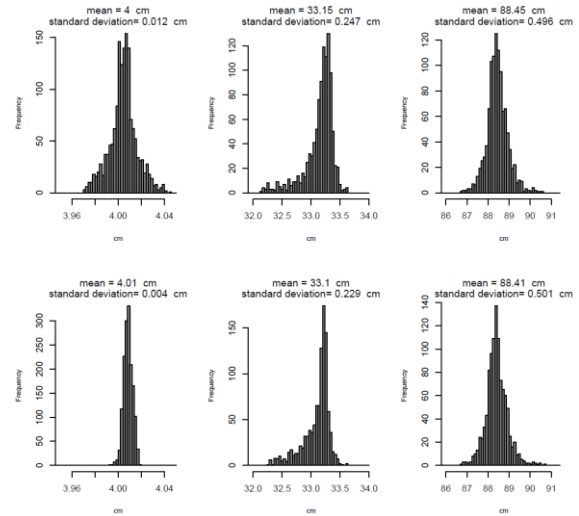


Figure 7 : Distribution des mesures de profondeur pour les trois gammes de mesures : 4 cm, 33 cm et 88,5 cm. En haut : résultats avec la première méthode de calibrage (lentilles idéales). En bas : avec la seconde méthode de calibrage (modélisation des distorsions).

#### 4 Résultats préliminaires en mode « caméra de profondeur »

D'un point de vue général, nos travaux visent à proposer des modèles de caméras de profondeur basées sur notre système plénoptique, et pouvant travailler en lumière naturelle. Il est très simple de coupler cette méthode de mesure avec une carte de disparité pour aboutir à une véritable caméra de profondeur, grâce à l'équation 7, reliant  $Z_{B2}$  à  $d_u$  ou  $d_v$ .

Les figures 8 et 9 montrent les résultats préliminaires sur des images expérimentales, pour lesquelles la profondeur des objets est codée en couleur. Sur la figure 8, les 3 sous-images NO, NE et SO permettent d'établir une carte de disparité, où les zones uniformes contiennent la disparité relative à leurs contours. Dans un second temps, on peut coupler la méthode de mesure par homographie variable à cette carte de disparité, et ainsi obtenir soit l'image en profondeur de la scène, soit l'image des hauteurs des objets par rapport à la distance de référence  $Z_A$ . La figure 9 présente ici l'image des hauteurs des objets dans la scène. D'un point de vue « benchmarking », notre méthode fonctionne à 15

images par seconde sur un ordinateur standard (processeur intel i5).

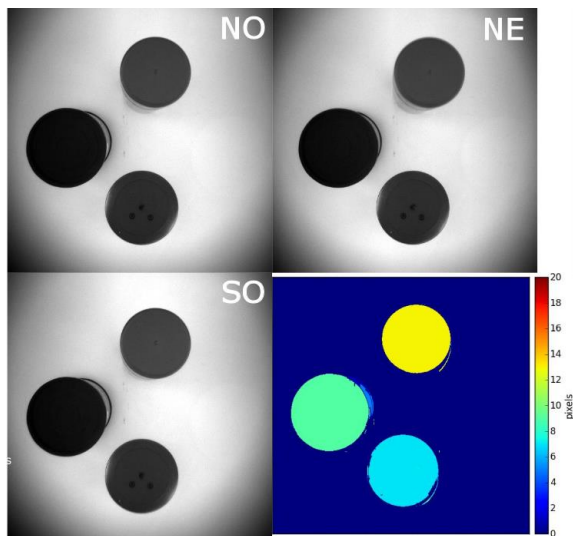


Figure 8 : Carte de disparité obtenue en lumière naturelle, illustrant les possibilités de la caméra de profondeur basée sur une caméra plénoptique

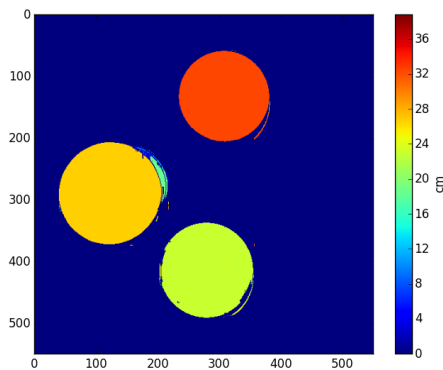


Figure 9 : Carte de profondeur en cm

## 5 Conclusion

Ce papier présente une méthode alternative pour modéliser et calibrer les caméras plénoptiques. Notre premier prototype utilise seulement 4 mini-lentilles, et les premiers résultats sont très prometteurs. Le calibrage peut être effectué avec une méthode simple, pour laquelle les paramètres intrinsèques et extrinsèques ne sont pas déterminés individuellement. Cela représente un réel avantage comparé aux autres méthodes, qui requièrent la connaissance de la distance entre les différents éléments optiques.

Les perspectives de ce travail sont nombreuses. Par exemple, le premier prototype pourrait être amélioré en utilisant des lentilles plus petites, et nous pensons travailler sur une version composée de 16 mini-lentilles. Nous voudrions améliorer certains aspects

des caméras de profondeur en tirant avantage des sous-images fournies par ce dispositif. Cela pourrait réduire le phénomène d'occultation horizontale, qui apparaît dans les systèmes classiques de stéréovision, utilisant seulement deux vues de la scène.

## Références

- [1] Xu J., and al., "Original method to compute epipoles using variable homography: application to measure emergent fibers on textile fabrics," *J. Electron. Imaging* 21(2), 1–12 (2012).
- [2] Xu J., and al., "Using variable homography to measure emergent fibers on textile fabrics," *Proceeding Tenth Int. Conf. Qual. Control Artif. Vis.*, 1–8 (2011).
- [3] Zhang S., and Greenspan M., "Variable Homography Compensation of Parallax Along Mosaic Seams," *Image Anal. Recognit.* 4633, 271–284 (2007).
- [4] Lippmann G., "Epreuves réversibles. Photographies intégrales," *Acad. Sci.* 146, 446–451 (1908).
- [5] Ng R., and al., "Light Field Photography with a Hand-Held Plenoptic Camera," *Tech Report CSTR 2005-02, Stanford University Computer Science* (2005).
- [6] Wanner S., Fehr J., and Jähne B., "Generating EPI Representations of 4D Light Fields with a Single Lens Focused Plenoptic Camera," *Adv. Comput. Vis.* 6938, 90–101 (2011).
- [7] Bishop T. E., and Favaro P., "Full-Resolution Depth Map Estimation from an Aliased Plenoptic Light Field," *Comput. Vis. – ACCV 2010* 6493, 186–200 (2011).
- [8] Wilburn B., and al., "High performance imaging using large camera arrays," *ACM Trans. On Grap.* 24(3), 765-776 (2005).
- [9] R. Ueno and co., "Compound-eye camera module as 8.5x8.5x6 mm for 26 k-resolution depth map and 2-Mpix 2D imaging", *IEEE photonics journal*, Vol.5, n°3, august 2013
- [10] D. G. Dansereau, O. Pizarro and S. B. Williams, "Decoding, Calibration and Rectification for Lenselet-Based Plenoptic Cameras", *Computer Vision and Pattern Recognition, 2013 IEEE conference, Portland (USA), 23-28 June 2013*, pp 1027-1034.
- [11] Barreto J. P., and Daniilidis K., "Epipolar Geometry of Central Projection Systems Using Veronese Maps," *IEEE Proceeding Comput. Vis. Pattern Recognit. CVPR* 1, 1258–1265 (2006).