



HAL
open science

Architectures et apports de systèmes de vision light-field pour la vision par ordinateur

Cécile Riou

► **To cite this version:**

Cécile Riou. Architectures et apports de systèmes de vision light-field pour la vision par ordinateur. Autre. Université de Haute Alsace - Mulhouse, 2017. Français. NNT : 2017MULH0498 . tel-02372168

HAL Id: tel-02372168

<https://theses.hal.science/tel-02372168v1>

Submitted on 20 Nov 2019

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.

UNIVERSITE DE HAUTE-ALSACE

UNIVERSITE DE STRASBOURG

Année 2017

N° d'ordre :

Thèse

Présentée pour l'obtention du grade de

Docteur de l'Université de Haute-Alsace

Ecole Doctorale : 269

MATHEMATIQUES, SCIENCES DE L'INFORMATION ET DE L'INGENIEUR

Discipline : Electronique, Optronique et Systèmes

Présentée et soutenue publiquement

par

Cécile RIOU

13/12/2017

Architectures et apports de systèmes de vision *light-field* pour la vision par ordinateur

Sous la Direction de : Directeurs de thèse : Christophe CUDEL et
Jean-Philippe LAUFFENBURGER et Co-encadrant : Bruno COLICCHIO

Jury : Jean-José Orteu, Professeur à l'Ecole des Mines d'Albi (Rapporteur)
Olivier Aubreton, Maître de conférences (HDR) à l'Université de Bourgogne
(Rapporteur)
Johan Debayle, Professeur à l'Ecole des Mines de Saint-Etienne (Président)
Christophe Cudel, Professeur à l'Université de Haute-Alsace (Directeur)
Jean-Philippe Lauffenburger, Professeur à l'Université de Haute-Alsace (Co-directeur)
Bruno Colicchio, Maître de conférences à l'Université de Haute-Alsace (Co-encadrant)
Manuel Flury, Maître de conférences (HDR) à l'INSA de Strasbourg (Membre invité)

Table des matières

Remerciements	7
Introduction.....	9
Contexte de travail et objectifs	9
Contributions de la thèse	10
Organisation du document.....	11
1 Etat de l'art.....	13
1.1. Introduction.....	13
1.2. Caméras de profondeur ou RGB-D.....	13
1.2.1. Imagerie active	13
1.2.2. Imagerie passive	18
1.2.3. Synthèse	25
1.3. Systèmes d'acquisition <i>light-field</i>	27
1.3.1. Historique	27
1.3.2. Fonction plénoptique et <i>light-field</i>	27
1.3.3. Systèmes d'acquisition <i>light-field</i>	29
1.4. Calibrage des systèmes <i>light-field</i>	39
1.4.1. Calibrage des systèmes multi-caméras	39
1.4.2. Calibrage des caméras plénoptiques.....	40
1.5. Domaines d'applications des systèmes <i>light-field</i>	43
1.5.1. Refocalisation numérique	43
1.5.2. Calcul de carte de profondeur.....	46
1.5.3. Autres exemples d'applications	49
1.6. Synthèse et positionnement des travaux.....	53
2 Géométrie des systèmes <i>light-field</i> et conception d'un système multi-vues plénoptique	55
2.1. Introduction.....	55
2.2. Des systèmes multi-caméras aux caméras plénoptiques.....	56
2.3. Prototypes de systèmes multi-vues et multi-caméras.....	58
2.3.1. Système multi-vues	58
2.3.2. Système multi-caméras	59
2.4. Développement d'un système multi-vues plénoptique.....	61
2.4.1. Modélisation géométrique.....	62

2.4.2.	Rappels d'optique de Gauss et d'optique matricielle	65
2.4.3.	Simulation et tracé de rayons du dispositif multi-vues plénoptique	69
2.4.4.	Caractérisation de notre système multi-vues plénoptique.....	73
2.4.5.	Exemples d'acquisition du système multi-vues plénoptique.....	78
2.5	Conclusion	84
3	Modélisation par l'homographie variable et calibrage	85
3.1	Introduction.....	85
3.2	Modèle sténopé	86
3.2.1	Transformation repère monde – repère caméra	87
3.2.2	Transformation repère caméra – repère capteur	87
3.2.3	Transformation repère capteur – repère image	88
3.2.4	Modèle sténopé complet	88
3.2.5	Prise en compte de la distorsion	89
3.2.6	Calibrage d'une caméra.....	89
3.3	Homographies	90
3.3.1	Projection d'un plan de l'espace sur le plan image d'une caméra : homographie	90
3.3.2	Homographies entre deux plans image de deux caméras	91
3.4	Homographie variable et mesures de profondeur à partir de la disparité	92
3.4.1.	Principe de l'homographie variable entre images	92
3.4.2.	Modélisation d'un système multi-vues	93
3.4.3.	Généralisation de l'homographie variable au dispositif multi-vues	94
3.4.4.	Expressions des homographies H_c et H	95
3.4.5.	Calcul de la profondeur Z	95
3.4.6.	Comparaison avec la triangulation.....	96
3.5.	Calibrage des systèmes multi-vues, modélisés par des homographies variables.....	98
3.5.1	Calibrage avec la méthode de Zhang	98
3.5.2.	Calibrage simple	100
3.5.3.	Calibrage amélioré.....	101
3.5.4.	Précision des mesures	102
3.5.5.	Critère de calibrage	105
3.6.	Comparaison avec le calibrage à partir de plusieurs vues [Zhan00]	106
3.7.	Conclusion	109

4	Imagerie multi-vues : intérêts et algorithmes.....	111
4.1.	Introduction.....	111
4.2	Rectification des images.....	112
4.3	De la rectification à la refocalisation numérique calibrée	113
4.4	Etude de la mesure du flou sur les images refocalisées.....	116
4.4.1	Critère de mesure du flou	116
4.4.2	Incertitudes de mesures et influence des paramètres.....	120
4.5	Exemples d'applications de la refocalisation numérique calibrée	126
4.5.1.	Mesures 2D : détections conventionnelles dans un plan de travail	127
4.5.2.	Vision 3D « agile »	129
4.5.3.	Conclusion	139
4.6.	Images de profondeurs par propagation de contours	140
4.6.1.	Introduction.....	140
4.6.2.	Calcul de la disparité avec le système à quatre mini-lentilles.....	141
4.6.3.	Résultats expérimentaux.....	147
4.6.4.	Conclusion	150
4.7.	Conclusion	150
	Conclusions et perspectives	153
	Conclusions.....	153
	Perspectives	154
	Production scientifique	157
	Bibliographie	159

Remerciements

Aux lectrices et lecteurs de ma thèse : ces remerciements,

Vous allez lire ces quelques lignes, soit parce que vous faites partie des gens qui seront probablement cités ultérieurement soit pour lire ma thèse par curiosité. A la première catégorie, un grand merci pour votre soutien ou à votre participation quel que soit le moment et la participation à cette thèse. Pour les seconds, j'espère que la lecture du reste du manuscrit sera, au moins agréable, et au mieux, instructive.

Cette thèse s'est effectuée en 3 ans au laboratoire MIPS à Mulhouse au sein des équipes IMTI et MIAM. Je tiens tout d'abord à remercier mes trois encadrants de thèse : mes directeurs de thèse Christophe Cudel et Jean-Philippe Lauffenburger et mon co-encadrant Bruno Colicchio. J'ai beaucoup apprécié de travailler avec vous trois, vous avez toujours essayer de privilégier les solutions et les directions les plus intéressantes pour ce projet, plus que pour la suite de vos carrières respectives et vos accointances personnelles. Particulièrement Jean-Philippe avec son regard plus extérieur sur la thèse, il m'a permis de prendre du recul. Pour Bruno, son aide a été précieuse de très nombreuses fois et toujours a toujours clarifié simplement les données. Enfin Christophe ta disponibilité et ton soutien permanent ont été un précieux atout, je t'en remercie.

Je remercie tous les membres de mon jury qui m'ont fait l'honneur et le privilège d'avoir accepté l'invitation pour cette soutenance de thèse. Merci au Pr. Johan Debayle d'avoir accepté de présider mon jury. Merci aux Pr. Jean-José Orteu et Pr. Olivier Aubreton d'avoir rapporté le présent manuscrit, vos commentaires et subjections ont permis d'améliorer sa qualité. Merci à MCF. Manuel Flury d'avoir accepté de participer à ce jury et de m'avoir conseillé sur certaines parties de ma thèse. De plus, ses conseils sur le doctorat au cours de ma dernière année d'école d'ingénieur ont été les premières graines qui m'ont poussées à faire une thèse.

Je tiens à remercier ceux qui ont entretenues et fait germer ces petites graines les membres de l'équipe du DOTA de l'ONERA Toulouse, et plus particulièrement, Christophe L., Nicolas R., Romain C., Erwan B., Karine A., Manuel C.C. et tous les autres qui m'ont procuré leurs conseils très précieux sur le choix d'une thèse et les différents écueils que l'on peut rencontrer en thèse. Au cours de ce travail, j'ai souvent pensé à vous en me remémorant ces conseils, ce qui m'a permis de supporter les difficultés rencontrées. Je te tiens aussi à remercier ceux qui m'ont soutenu pendant ma recherche de thèse, notamment pour leur enthousiasme, ils se reconnaîtront.

Aux membres du MIPS, je sais ce que vous vous dites : « elle nous a déjà oublié » et bien non, je ne vais vous remercier qu'à partir de maintenant : l'ensemble du laboratoire pour son accueil, son ambiance et plus particulièrement les équipes MIPS et MIAM. A Olivier H. merci pour les discussions scientifiques ou autres qui ont toujours été agréables et enrichissantes et pour tous les bons moments

(comme le marché de Noël, le Shamrock, les raquettes, le route des vins et j'en passe). Je te tiens à remercier Nicolas V. pour son aide très précieuse au cours des manip, sa joie de vivre et son humour. A Mathieu D., je le remercie pour son soutien dans la mise en œuvre de mes expérimentations, nos discussions sur de nombreux sujets. A Hakim D. : ta bonne humeur a toujours été source de regain d'énergie. A Ali M., merci pour nos discussions enrichissantes. A Djafar O.A. le partage et l'écoute de conférences notamment sur l'Algérie. A tous les membres de l'équipe IMTI, merci pour les moments échangés ensemble, la bonne humeur et les rires.

Aux membres de l'équipe MIAM permanents et doctorants, je tiens à vous remercier pour votre accueil même si je n'étais pas souvent présente. A Olivier D., les moments passés ensemble permettaient de se soutenir entre sudistes. Je tiens à remercier tout particulièrement mes partenaires de belotes du MIAM : Joël L., Jonathan L. et Luc G. Ces soirées ont été de réelles soupapes pour évacuer la pression. J'ai vu que vous continuez et c'est très bien. A Jonathan L. et Joël L., je vous remercie pour votre présence, votre aide au cours de la préparation et de la mise en œuvre de mes expériences Un grand merci à Joël L. pour ton accueil réconfortant, notamment la semaine de ma soutenance.

Pour en finir avec mes remerciements mulhousiens, citons les personnes rencontrées grâce à Pars-En-Thèse. Les soirées organisées (le bowling, les soirées jeux ...) ont été un excellent moyen de se vider la tête et de rencontrer d'autres doctorants qui ne sont pas de notre laboratoire. Un énorme merci à Jérémy W. pour son soutien et les bons moments passés ensemble. Un grand merci à Leïla H. ma Toulonnaise, à Mathilde S., à Agnès R., à Maxime L., à Liza M.. A Laura et Giorgio merci notamment pour les soirées glace, que du bonheur !

A l'ensemble des gens que j'ai rencontré au cours de cette thèse, j'espère que l'on pourra se revoir et partager de bons moments ensemble.

Passons aux soutiens indéfectibles, à Livia O. et Pierre O. mes grands-parents maternels, à mes cousins Georges M. et Danielle M. que je remercie grandement pour leur soutien avant et pendant ma thèse ainsi que pour leur présence à ma soutenance de thèse. A toutes les personnes du cercle familial ou amical qui m'ont soutenu pendant ma thèse, je les en remercie. A Erwan B., merci pour les conseils et les bons moments parisiens A Jean-Baptiste G., merci pour ton aide et ta bonne humeur.

Il reste une personne à remercier et non la moindre : Catherine O. ma mère. Ta patience et ton soutien ont été des éléments indispensables à la finalisation de cette thèse. Ta relecture et tes subjectives m'ont permis de produire un manuscrit que le jury a qualifié de pédagogique et je pense que c'est grâce à toi. Comme me l'on si bien dit certaines personnes, peu de parents se sont impliqués comme tu l'as fait et je t'en suis très reconnaissante.

Au cas où, j'aurais oublié des gens, je m'en excuse et vous suis reconnaissante. Je le suis également pour ceux qui ont pris le temps de lire ces quelques lignes. Aux lectrices et lecteurs de la suite du manuscrit, je vous souhaite une lecture agréable et constructive concernant les caméras *light-field*.

Introduction

Contexte de travail et objectifs

La mesure de profondeur présentant un enjeu dans de nombreux domaines en vision par ordinateur, comme la vision industrielle, la reconstruction 3D, la robotique ou encore la navigation par vision, elle a entraîné le développement des caméras 3D et des caméras de profondeur. Dans cette catégorie de dispositifs, les plus connus sont les systèmes de stéréovision, les caméras multi-vues, les caméras Time Of Flight ou les capteurs de type Kinect. Deux technologies peuvent être distinguées : les « passives » travaillant en illumination naturelle et les « actives » nécessitant l'ajout d'une illumination ou d'une autre source d'information (chaleur, son ...) pour extraire une information de profondeur d'une scène. Dans le cas où une source lumineuse annexe est ajoutée, une mire structurée ou aléatoire peut être contenue dans les images, les rendant ainsi inexploitable pour tout autre analyse par vision, comme la lecture optique de caractères (OCR).

Le but de ce projet de thèse est de s'intéresser aux caméras multi-vues et plénoptiques dans un cadre différent de celui de la photographie. Nous souhaitons notamment étudier de nouveaux domaines d'application pour la vision par ordinateur. La thèse s'articule autour de deux volets principaux : le premier relève de l'optique et le second de l'algorithmique.

Le premier volet de mes travaux concerne l'optique et la géométrie des caméras light-field. Si dans un premier temps nous avons travaillé avec un système multi caméras et une caméra multi-vues, nous étudions la possibilité de développer un dispositif multi-vues plénoptique. L'une des solutions envisagées est de développer un système de bague allonge composé des mini-lentilles que l'on pourrait insérer dans une caméra conventionnelle entre l'objectif et le capteur image. Pour effectuer cette partie, il a été nécessaire de développer un outil de calcul afin de simuler le comportement des rayons lumineux au sein du dispositif ainsi que de définir des critères nécessaires à la caractérisation de notre système.

D'un point de vue algorithmique, nous nous sommes intéressés à la modélisation et au calibrage de ces systèmes en exploitant les propriétés de l'Homographie Variable (HV). Ensuite, nous étudions leurs intérêts dans le domaine de la vision par ordinateur, en exploitant par exemple leur capacité de refocalisation numérique ou les informations 3D qu'elles capturent. Dans un premier temps, nous nous sommes intéressés aux propriétés de la « refocalisation numérique », et avons défini le concept de plan de travail, qui semble pertinent en vision par ordinateur. Ces travaux ont été réalisés soit avec un prototype de caméras quatre-vues soit avec un dispositif à quatre caméras, tous deux ayant été développés au laboratoire mais applicables au système multi-vues plénoptique et autres caméras plénoptiques. Enfin, nous avons proposé une méthode originale pour calculer des cartes de profondeur, ou des images de disparité. La méthode développée s'appuie sur le calcul de cartes de distances, pour mettre en évidence la parallaxe entre des contours coïncidents.

Contributions de la thèse

Les contributions de cette thèse peuvent être regroupées entre 3 grandes catégories :

- Géométrie des systèmes light-field
 - Développement de prototypes multi-caméras et multi-vues
 - Proposition d'un outil de simulation optique pour la caractérisation d'un système multi-vues plénoptique et développement d'un premier prototype
- Modélisation et calibrage des systèmes multi-vues et multi-caméras
 - Modélisation des systèmes multi-vues et multi-caméras basée sur l'homographie variable
 - Proposition d'une méthode de calibrage simple et aisée à partir du modèle précédent
- Intérêts et algorithmes des images acquises avec ces dispositifs
 - Calcul de cartes de disparité fondé sur une approche « contours »
 - Calcul d'images refocalisées numériques calibrées en profondeur
 - Utilisation de ces images pour faciliter des détections conventionnelles en vision industrielle ou la classification d'obstacles en zones de profondeur dans le domaine de l'automobile

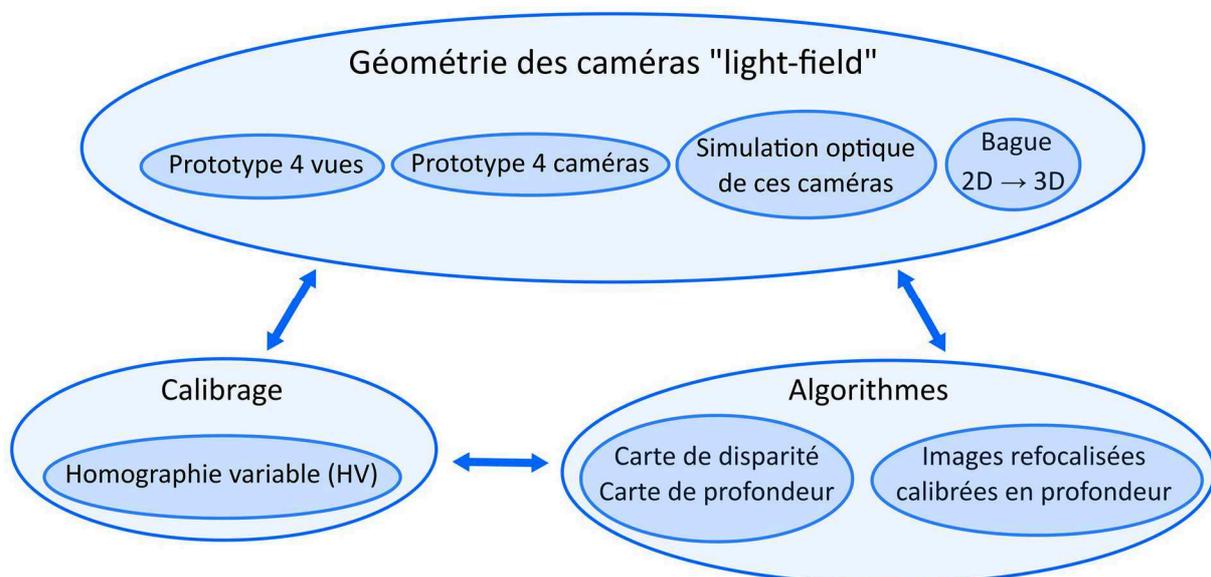


Figure 0.1 Résumé des grandes parties étudiées au cours de la thèse. Il suit la répartition des chapitres. Dans le chapitre 2 : Etude de la géométrie des caméras light-field, dans le chapitre 3 : calibrage des systèmes multi-vues et multi-caméras et dans le chapitre 4 : Développement d'algorithmes basés sur notre calibrage et dédié aux caméras étudiées et créés au laboratoire.

Organisation du document

Cette thèse est divisée en quatre chapitres :

- Chapitre 1** Présente l'état de l'art concernant ces travaux de thèse. Les différentes technologies de caméra de profondeur sont exposées. La fonction plénoptique et le light-field sont expliqués ainsi que les différentes approches d'acquisition de ces notions. Enfin les méthodes de calibrage et les domaines d'application des caméras *light-field* sont détaillés.

- Chapitre 2** Propose l'étude géométrique des caméras multi-vues et multi-caméras. Le développement des prototypes multi-vues et multi-caméras, mis en œuvre au cours de cette thèse est expliqué. Nous avons proposé un outil de simulation optique des systèmes multi-vues. Des résultats obtenus par simulation ainsi qu'expérimentaux sont présentés.

- Chapitre 3** Montre notre modélisation des systèmes *light-field* basée sur l'homographie variable. A partir de ce modèle, nous avons proposé une méthode simple et aisée de calibrage de ces dispositifs. De plus, une comparaison des résultats obtenus avec le processus de calibrage de Zhang est effectuée.

- Chapitre 4** Expose deux applications développées pour ces systèmes. La première consiste à calculer des cartes de profondeur en se basant sur une approche « contours ». La seconde application consiste à calculer des images refocalisées calibrées en profondeur. Des exemples concrets d'utilisation de ces images en vision industrielle et dans le domaine de l'automobile sont présentés.

1 Etat de l'art

1.1. Introduction

Cette thèse vise à étudier les caméras *light-field* (LF), s'inscrivant dans la famille des caméras 3D. L'état de l'art proposé dans ce chapitre a donc pour objectif de situer ces systèmes parmi les autres technologies d'imagerie. Les systèmes de vision 3D sont variés et adaptés à leurs champs d'utilisation. La recherche sur ce sujet a vu ses prémices au début des années 80 et est toujours d'actualité. Ainsi, de nombreuses technologies ont été développées pour effectuer ces mesures, la variété des techniques provenant du grand nombre de domaines d'application (médecine, industrie, robotique mobile, jeu vidéo...), et des spécificités relatives à ces domaines.

Ce chapitre débute par la présentation des systèmes de vision 3D, classés selon deux catégories principales : (i) Les systèmes actifs sont composés d'une source lumineuse et d'un ou plusieurs capteurs ; (ii) Les systèmes passifs sont composés d'un ou plusieurs capteurs sans qu'aucune source lumineuse annexe ne participe à l'acquisition des données.

Dans un deuxième temps, nous décrivons plus précisément les systèmes LF et plénoptiques. La fonction plénoptique, le concept *light-field* et les caméras LF sont présentés. Enfin, dans une dernière partie, les principaux domaines d'application de ces caméras sont décrits.

1.2. Caméras de profondeur ou RGB-D

Les dispositifs utilisés en imagerie 3D peuvent être classés en deux catégories : les systèmes passifs ou les systèmes actifs. Les systèmes passifs acquièrent la lumière émise ou réfléchiée par l'objet étudié. Les systèmes dits « actifs » nécessitent l'ajout d'une source lumineuse annexe : par exemple, on peut utiliser les techniques de triangulation laser ou le temps de vol d'une onde lumineuse. Les systèmes décrits dans cette section sont listés ci-dessous :

- Imagerie active
 - Temps de vol
 - Triangulation laser (lumière structurée)
 - Défectométrie
- Imagerie passive
 - Approche sténopé (*structure from motion* et stéréovision)
 - Approche lentille (profondeur par la focalisation et profondeur par la défocalisation)
 - Approche lentille et ouverture (systèmes à ouverture codée et caméra plénoptique)

1.2.1. Imagerie active

Les caméras 3D dites « actives » sont caractérisées par deux éléments fondamentaux : une source lumineuse d'émission contrôlée dans l'espace et dans le temps ; un système pour analyser l'image de la scène illuminée. Le terme actif est utilisé parce qu'une source lumineuse est couplée à une caméra ou

un système d'acquisition. Il est à noter que la source d'illumination n'est pas obligatoirement dans le domaine spectral du visible. Il existe plusieurs approches, et on peut distinguer les caméras à temps de vol, les systèmes à triangulation ou illumination structurés. Suivant la nature des matériaux, l'illumination structurée permettra d'extraire une information 3D depuis des réflexions diffuses, ou bien spéculaires pour les surfaces réfléchissantes.

1.2.1.1. Temps de vol

La technique temps de vol (*Time Of Flight* ou *ToF*) utilise une source lumineuse pulsée (le plus souvent laser ou LED) synchronisée avec un détecteur. Le principe repose sur l'émission par la source d'une impulsion lumineuse qui est réfléchiée par les objets dans la scène. La distance source-objet est alors déterminée par télémétrie et il s'agit d'un système de mesure point par point. La distance entre la source et les objets peut être obtenue par la mesure du temps écoulé entre l'émission et la réception du signal lumineux. La distance d d'un point X qui est projeté en x a pour expression :

$$d = \frac{c \cdot t}{2} \quad (1.1)$$

où c est la vitesse de la lumière et t est le temps de propagation de la lumière.

La portée des systèmes basés sur le temps de vol dépend de la puissance de la source laser. La distance minimale est liée à la durée du pulse et des traitements mis en œuvre.

Pour mesurer le temps de vol, il existe plusieurs technologies. Les deux principales sont :

- La mesure directe du temps caractérisée par une illumination impulsionnelle
- La mesure de déphasage caractérisée par une illumination sinusoïdale continue modulée dans le temps

Pour la mesure directe, la distance mesurée correspond à celle de l'équation précédente.

Pour la mesure du déphasage, le signal émis et le signal reçu ont la même fréquence mais ce dernier a une amplitude plus faible. Il est également caractérisé par un déphasage φ qui dépend de la distance d . Pour éviter toute ambiguïté de phase, la période de modulation λ est grande par rapport au temps de vol et est faible par rapport au temps de pose afin de pouvoir répéter la mesure.

$$d = \frac{c\lambda}{4\pi} \varphi \quad (1.2)$$

où c est la vitesse de la lumière. [CCPR09]

Un exemple de système de ce type a été proposé en 2004 pour effectuer de la reconstruction de visages [ZhHu04]. La distance de travail dépend des modèles et des principes. En temps de vol direct, elle peut aller jusqu'à quelques kilomètres. Dans cet état de l'art, on s'intéresse surtout aux *ToF* modulées. Leur distance de travail est plus petite, de l'ordre de 5 à 10 mètres. Ces systèmes fonctionnent

de jour comme de nuit. Le principal inconvénient des systèmes *ToF* est l'utilisation d'un laser synchronisé avec le détecteur avec la nécessité de scanner la scène pour acquérir une image. De plus, ces dispositifs sont sensibles aux facteurs environnementaux, tels que la lumière ambiante et la diffusion (situation de brouillard), des facteurs dépendant de la scène comme les propriétés géométriques et celles du matériau. Plusieurs pistes sont envisagées pour réduire l'influence de ces facteurs. L'adaptation de la distribution lumineuse en fonction de l'environnement est à étudier tout comme le développement de l'imagerie Micro *ToF*. Les LIDARs peuvent être utilisés dans ce domaine. La Kinect v2 de Microsoft et la CamCube de PMDTech en sont deux exemples commerciaux.

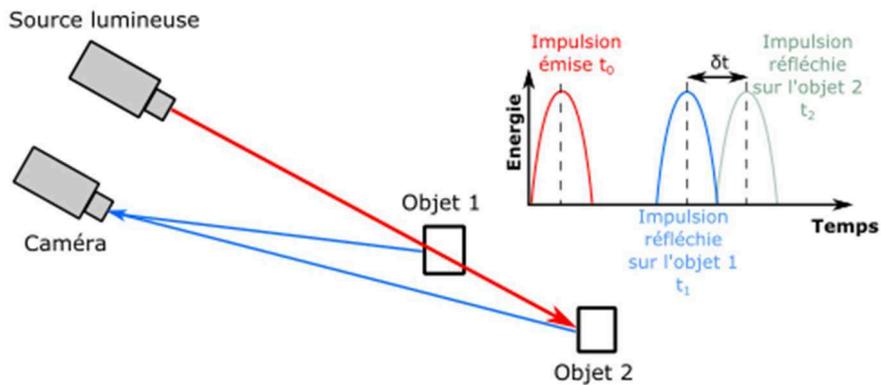


Figure 1.1 Principe du temps de vol

1.2.1.2. Triangulation laser

La triangulation est une technique qui peut être utilisée par les imageurs actifs et par les imageurs passifs. Son principe repose sur le théorème de Thalès. Il s'agit de déterminer la distance d'un point par rapport à la caméra en mesurant les angles entre ce point et deux points de référence. En illumination active, cette catégorie regroupe tous les systèmes faisant appel à des illuminations structurées, de la stéréoscopie ou de la triangulation par laser.

Dans le cas de la triangulation laser, le système est généralement composé d'un capteur linéaire ou matriciel et d'une source lumineuse émettant un faisceau en forme de point ou de ligne. Le principe de la mesure est décrit ci-dessous. La caméra et la source lumineuse doivent être calibrées.

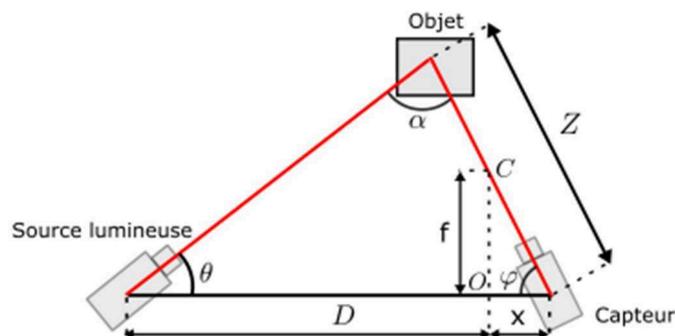


Figure 1.2 Principe de la triangulation laser

La focale du capteur est notée f , le centre de l'image O et le centre optique du capteur C . L'orientation θ entre le projecteur et le capteur est connue ainsi que leur distance D . La profondeur Z en fonction de la déviation x sur l'image est obtenue grâce à la loi des sinus, et a pour expression :

$$\frac{D + x}{\sin \alpha} = \frac{Z}{\sin \theta} = \frac{D + x}{\sin(\theta + \varphi)} \quad (1.3)$$

D'où

$$Z = (D + x) \frac{\sin \theta}{\sin(\theta + \arctan(\frac{f}{x}))} \quad (1.4)$$

L'avantage des systèmes de triangulation laser est leur moindre sensibilité à la lumière ambiante contrairement aux caméras passives. La résolution de ces systèmes dépend de l'éloignement entre la source lumineuse et le capteur, et de la distance à l'objet d'intérêt. Les distances de travail classiques sont comprises entre quelques millimètres et quelques dizaines de mètres. La résolution est quant à elle comprise entre quelques centaines de micromètres et quelques centaines de millimètres. Les principaux inconvénients de ces technologies sont leur portée limitée, les erreurs engendrées par les occultations de la scène et la limitation de leur fonctionnement aux objets opaques et structurés.

A partir de ce principe, beaucoup de méthodes dérivées ont été développées. Elles sont décrites dans Forest [Fore04] et regroupées suivant différentes catégories : *pose measurement*, *projective approach*, *euclidean approach*, *time multiplexing – switching the laser slit* et *lookup table generation*.

Caméras à lumière structurée

Les caméras 3D à lumière structurée ont été développées pour supprimer l'étape de balayage mécanique présente pour les systèmes de triangulation laser. Il s'agit donc d'une extension de cette technique. Ces caméras à lumière structurée se caractérisent par l'interprétation de la déformation d'un motif 2D projeté dans la scène afin d'accéder à l'information de profondeur. Elles reposent sur le principe de la triangulation et dans ce cas, le motif correspond à une multitude de points ou de lignes ayant des orientations différentes. Le nuage de point ainsi extrait, provient d'une unique acquisition. La structure du motif projeté détermine une codification spatiale. Cette codification doit permettre de différencier chaque point de ces voisins. Ainsi, plusieurs choix concernant la mire existent : on peut projeter des mires rectangulaires préférables aux mires continues. La mire choisie peut être projetée une seule fois ou plusieurs fois de manière séquentielle. L'inconvénient des mires séquentielles est le rallongement du temps d'acquisition. Cette technique nécessite une étape de mise en correspondance entre chaque point du motif et un unique point de l'image acquise par la caméra. Cette étape est complexe et a mené à l'étude de la modulation de la lumière structurée, soit en intensité, soit dans le temps, soit dans l'espace ou en couleur. Ces nombreuses variantes de mires et de modulation de mires

ont été envisagées par différentes équipes de recherche [SaPB04], dont la plupart ont été inventoriées par Geng [Geng11]. De plus, tout comme les systèmes de stéréovision, les caméras à lumière structurée nécessitent une étape de calibrage. La Kinect v1 de Microsoft est un exemple de système commercial à lumière structurée. Un autre exemple de commercialisation de système à lumière structurée est le scanner Comet 5 de Steinbichler. La projection d'un motif de lumière structurée présente plusieurs avantages tels que la distribution dense et uniforme de la reconstruction et des reconstructions calculées en ligne voire en temps réel. Enfin, cette technique présente de nombreux domaines d'application comme le contrôle qualité, la reconstruction d'empreintes digitales ou bien la reconstruction de sites archéologiques immergés.

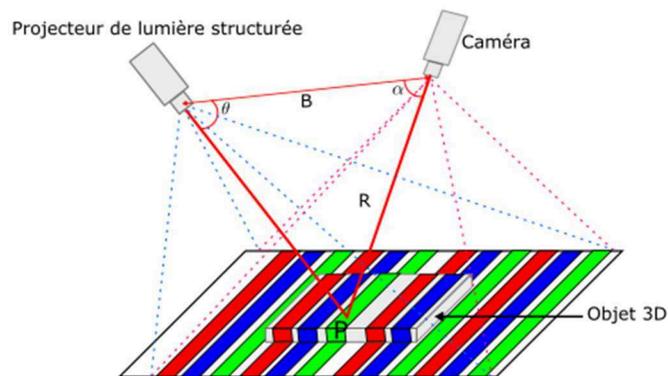


Figure 1.3 Principe des caméras à lumière structurée

1.2.1.3. Déflectométrie

La déflectométrie est une méthode permettant de mesurer un champ de pentes. Elle repose sur la constatation qu'une image d'une surface spéculaire est une déformation de son environnement pour l'œil humain. On peut estimer la forme d'une surface si l'on connaît le motif projeté (ses dimensions et ses caractéristiques) sur la surface spéculaire. De base, la déflectométrie va être en mesure de mettre en évidence un défaut de surface. Sa caractérisation 3D est plus complexe et nécessite une bonne connaissance de la géométrie générale de la scène.

Il existe plusieurs montages possibles [Surr02] : le plus courant est présenté par la Figure 1.4. La projection du motif doit être la plus diffuse possible pour prendre en compte la directivité des surfaces spéculaires. La projection de la mire, composée de franges noires et blanches, s'effectue sur un large écran dépoli. Cette technique repose sur l'extraction de la phase. Pour faciliter cette extraction, l'utilisation d'une mire sinusoïdale est préférable. Les franges correspondant au reflet déformé de la mire sont comparables à celles obtenues avec un système interférométrique. Cependant, contrairement à celles obtenues avec ce dernier, le système de déflectométrie est insensible aux vibrations et une plus grande profondeur de mesure peut être obtenue en réglant la sensibilité du système. Ce type de mesures est surtout employé pour faire de la mesure de planéité.

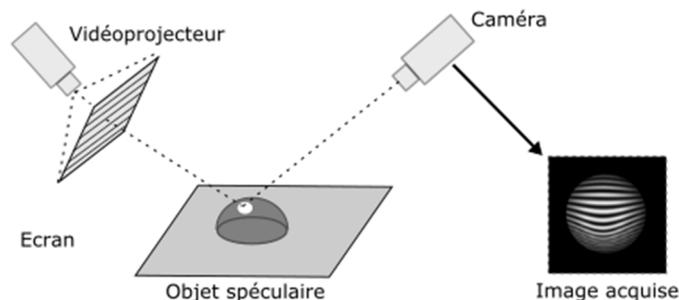


Figure 1.4 Principe de la déflectométrie.

Cette technique est employée pour les objets spéculaires [MoKu05]. Elle l'est aussi pour les objets transparents et il est ainsi possible de déterminer la forme de la surface de ces objets [HSKK96]. Une autre technique, appelée déflectométrie de Schlieren, est aussi utilisée pour observer les objets transparents [Sett01]. La déflectométrie est utilisée dans le domaine de la vision industrielle pour effectuer des contrôles sur des objets spéculaires et transparents. Une autre méthode dérivée de la déflectométrie appelé *shape from distortion* a été proposée par Tarini et al. [TLGS05] pour étudier les objets réfléchissants.

Notons que les réflexions spéculaires peuvent aussi être utilisées pour scanner des objets métalliques, dans des techniques de type *scanning from heating*. Dans ce cas, l'analyse de réflexions causées par une source infrarouge (IR) renseigne sur les aspects de surface des matériaux [ABVT13].

1.2.2. Imagerie passive

L'imagerie dite « passive » utilise le contenu de scènes naturelles et la géométrie du système de vision, pour déterminer des profondeurs dans une scène. Un éclairage peut être utilisé pour améliorer le contraste des images, mais il n'a aucun rôle pour la mesure 3D. Afin d'accéder à l'information utile pour le calcul de la profondeur, elles utilisent l'intensité lumineuse acquise par le capteur. Les techniques employées reposent sur différentes approches comme celles présentées dans le Tableau 1.1. Chacune des approches diffère dans sa manière d'acquérir un point de la scène et son chemin optique. La méthode la plus usitée est la stéréovision, mais la difficulté réside dans la mise en correspondance des points entre les deux images pour leurs zones homogènes [Cham05, SCDS06].

Approches	Sténopé	Lentille	Lentille et ouverture
Technologies	Structure from motion	Profondeur par la focalisation	Systèmes à ouverture codée
	Stéréovision	Profondeur par la défocalisation	Caméra plénoptique

Tableau 1.1 Les différentes approches d'imagerie passive et les technologies associées à chaque approche

1.2.2.1. Approche sténopé

L'approche sténopé repose sur le principe que, pour une caméra sténopé, chaque point de l'image correspond à un unique chemin optique. Ainsi, tous les points imagés par la caméra sont nets quelles que soient leurs profondeurs. Si on utilise deux points de vue différents, les deux images acquises sont différentes. Grâce à cette différence, l'information de profondeur peut être extraite. L'approche sténopé sera abordée d'un point de vue de sa modélisation dans le chapitre 3 Calibrage de ce document.

Les deux principales méthodes associées à l'approche sténopé sont la structure à partir du mouvement et la stéréovision.

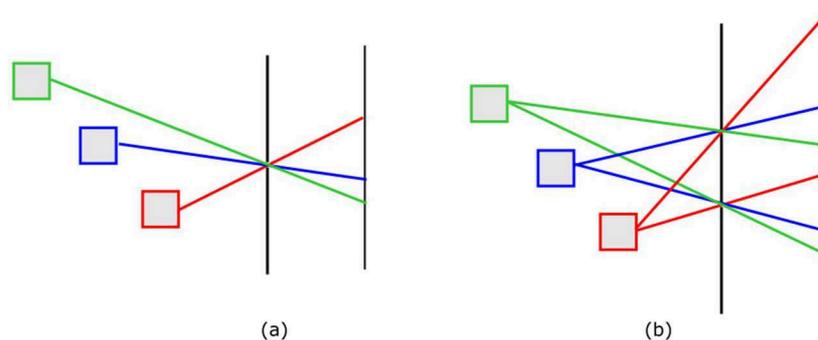


Figure 1.5 Modèle de caméra sténopé, (a) avec un point de vue ; (b) avec deux points de vue faisant apparaître deux images distinctes permettant d'extraire l'information de profondeur.

Structure à partir du mouvement / Structure from motion

En introduisant un mouvement lors de l'acquisition d'images, il a été montré qu'il est possible d'obtenir des informations de surface. La méthode *structure from motion* ou *shape from motion* mesure le mouvement relatif entre l'objet et la caméra [Rebe17]. A partir de cette mesure, il est possible de trouver l'information de surface. Cette technique nécessite une séquence d'images obtenues soit à partir d'une vidéo, soit à partir d'acquisitions successives d'un objet fixe par une caméra en mouvement, soit à partir d'acquisitions d'un objet mobile avec une caméra fixe. La recherche de points caractéristiques permet de calculer le déplacement entre les différents points de vue. La connaissance de ce déplacement permet de reconstruire l'objet [PVVV04]. L'analyse du mouvement apparent des points caractéristiques en fonction du mouvement entre deux acquisitions permet de retrouver l'information de profondeur. Par exemple, plus un point semble avoir un déplacement faible entre deux images, plus il est éloigné de la caméra. Une des applications les plus usitées est le SLAM (*Simultaneous Localisation And Mapping*), technique d'ego-localisation et de construction cartographique répandue en robotique mobile. Elle consiste à définir le mouvement d'une caméra fixée sur un robot pour reconstruire l'environnement ambiant en temps réel [Bail02]. Cette technique est passée dans le domaine grand public avec *Google Street View*. Cet outil permet de visualiser en 3D des villes grâce à des images acquises avec des caméras fixées sur un véhicule.

Stéréovision

Les systèmes de stéréovision sont composés de deux imageurs monovoies, séparés d'une distance connue. La stéréovision repose sur le principe de la triangulation présentée précédemment. Les caméras étant espacées, elles acquièrent deux images avec des points de vue différents. Cette différence est caractérisée par l'effet de parallaxe. Plus l'objet est proche des caméras et les caméras distantes l'une de l'autre, plus cet effet est important. Ainsi, en calibrant le système, il est possible de déterminer la distance à partir de la mesure de la parallaxe. Cette technique sera détaillée dans la section 1.5.2.

Les systèmes stéréoscopiques sont les systèmes passifs les plus étudiés. Le principe repose sur la reproduction du fonctionnement de la vision humaine. Deux caméras acquièrent deux points de vue différents de la même scène (Figure 1.6). A un point dans une image correspond une droite, appelée ligne épipolaire dans la seconde image. La matrice fondamentale F modélise ces correspondances à travers l'expression : $P_G(T) \cdot F \cdot P_D = 0$. Quand les deux caméras sont parallèles et alignées, les lignes épipolaires sont alignées. La *baseline* est la distance entre les centres des deux caméras.

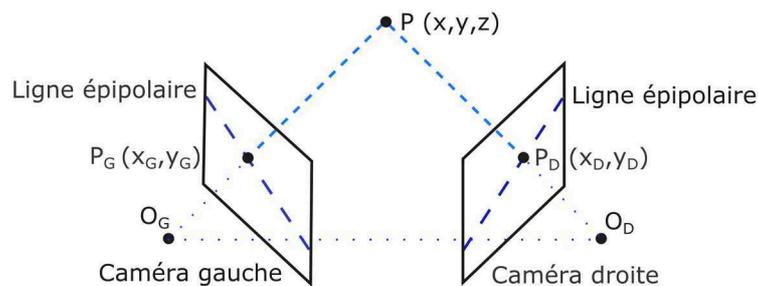


Figure 1.6 Figure schématique d'un système stéréoscopique composé de deux caméras.

La différence de point de vue permet d'extraire l'information de profondeur. Cette mesure n'est possible que si le système a été calibré préalablement. La principale méthode de calibrage consiste à déterminer les paramètres intrinsèques (focale, taille des pixels, ...) et les paramètres extrinsèques (orientation et position) de chacune des caméras. Grâce à la connaissance de ces paramètres, chaque image peut être rectifiée, pour être dans un même plan, au moyen d'une transformation homographique et ce, quel que soit l'orientation des caméras. Le calibrage correspond donc à l'estimation de la géométrie épipolaire du système. Les points 3D peuvent être calculés en utilisant la géométrie épipolaire [HaZi03] et un principe de triangulation. La profondeur Z dépend de la *baseline*, comme illustré par la Figure 1.7, et est définie par l'équation suivante :

$$Z = \frac{b \cdot f}{x_g - x_d} \quad (1.5)$$

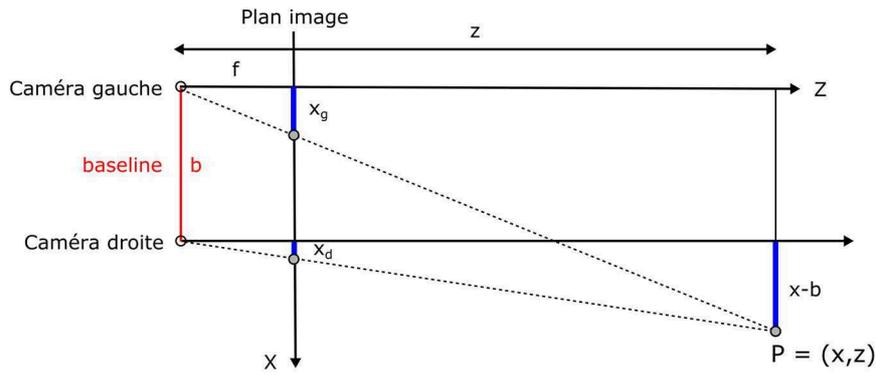


Figure 1.7 Illustration de la mesure de la profondeur Z avec un système stéréoscopique, composé des caméras gauche et droite.

L'information 3D d'une scène peut être établie dès que des points des images droite et gauche sont mis en correspondance. Cette étape est fortement dépendante de la texture des objets présents dans la scène (et donc de l'éclairage de la scène), des zones d'occultations potentielles. Ces appariements sont difficiles lorsqu'il y a de grandes zones uniformes et des occultations trop importantes [Szel10]. Afin de diminuer les zones homogènes et pour faciliter la mise en correspondance, les systèmes de stéréovision pour la vision industrielle ont souvent une source IR intégrée, permettant de projeter une texture aléatoire sur la scène. La Figure 1.8 présente un exemple d'acquisition stéréoscopiques et de carte de profondeur, obtenues avec la projection d'une texture aléatoire. Ce principe général peut être adapté suivant la nature des objets comme les travaux de Harvent et al. [HCBO15], améliorant la reconstruction 3D et préservant l'intégrité des contours.

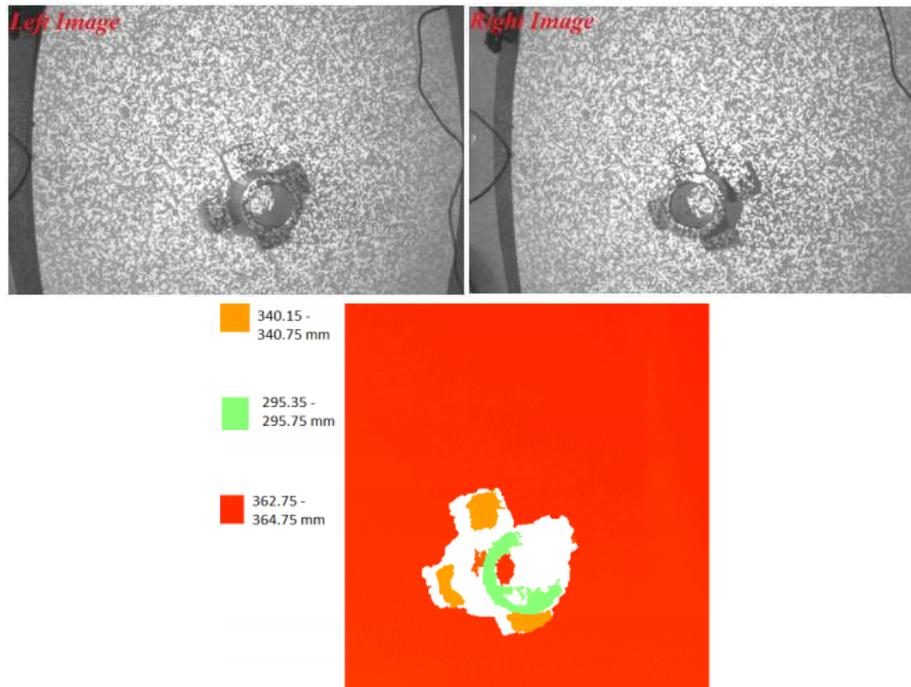


Figure 1.8 Exemple d'acquisition de stéréovision actif et carte de profondeur. L'illumination active sert à texturer la scène. De gauche à droite : les deux vues acquises et la carte de profondeur.

1.2.2.2. Approche lentille

L'approche lentille s'appuie sur l'étude de la netteté d'un point imagé par le système en fonction de sa profondeur. Elle repose sur la mesure de la variation de la réponse de l'imageur en fonction de la distance à mesurer et non pas sur une variation de la parallaxe. En considérant le système optique d'une caméra comme une lentille mince à focale fixe, seuls les objets présents dans le plan de focalisation seront nets. Les autres objets situés dans d'autres plans apparaîtront flous sur le plan image du capteur, comme on peut le voir sur la Figure 1.9. Le cas général pour une caméra est la présence d'ouvertures circulaires. Dans ce cas, l'image d'un point dans le plan focal est une tâche d'Airy et tout autre point en dehors du plan focal a une image qui correspond à un disque (section de cône) convolué avec la tâche d'Airy du plan focalisé. L'image du point focalisé représente la fonction d'étalement du système optique et dépend de la profondeur. Cette variation pourra être utilisée s'il est possible de mesurer la focalisation ou la défocalisation.

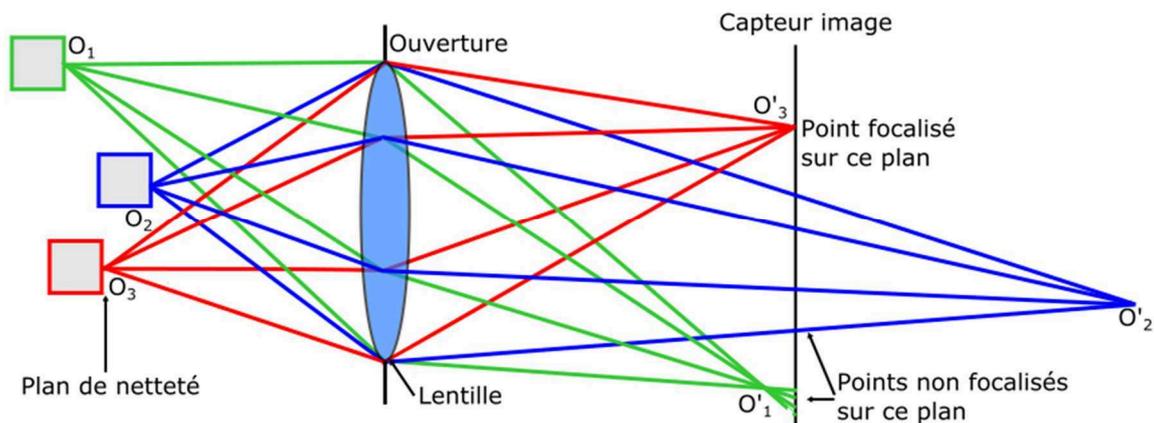


Figure 1.9 Modèle d'un imageur simple composé d'une lentille.

Il existe plusieurs méthodes pour effectuer ce type de mesure. Pour les deux méthodes présentées par la suite, il est nécessaire de réduire la profondeur de champ de l'imageur afin d'augmenter la précision de la mesure de profondeur. Il faut que la réponse de l'imageur en fonction de l'axe z varie le plus rapidement possible. Cependant, la variation de la réponse entraîne une dégradation de l'image acquise. Il est donc possible de mesurer la profondeur mais c'est la qualité de l'image qui est diminuée. La symétrie de part et d'autre du plan focal peut entraîner une incertitude sur la position d'un objet dont il est délicat de savoir s'il se situe avant ou après le plan de netteté. Il existe plusieurs manières d'extraire l'information de profondeur à partir de ces images. Parmi celles-ci, deux méthodes sont présentées ci-après : le *Depth from focus* et le *Depth from Defocus*.

Profondeur par la focalisation / Depth from focus (DFF)

La méthode *Depth from focus* a été introduite en 1989 [Naya89] et de nombreux travaux traitent de cette approche [PePG13]. Elle s'appuie sur un traitement d'images recherchant la netteté. C'est le même principe que l'autofocus des appareils photographiques numériques. Ce traitement associe à chaque pixel de la scène une image dans laquelle il est net. Grâce à l'acquisition d'une série d'images, une cartographie de l'espace objet peut être reconstruite et ainsi une image nette sur une grande profondeur de champ peut être calculée. L'acquisition consiste à balayer complètement la scène sur le plan focal. Elle présente l'avantage de nécessiter un unique imageur classique dont le détecteur peut être translaté. Ensuite, une évaluation de la netteté pour chaque pixel de la séquence d'images doit être effectuée. Cette étape permet de définir l'image pour laquelle chaque pixel est le plus net. Ensuite, une étape d'interpolation est nécessaire afin d'affiner la valeur de netteté associée à chaque pixel. Cependant, la position réelle du point dans la scène ne correspond pas toujours à la valeur maximale de netteté, selon la précision du système d'acquisition employé. Après cette étape, une carte de profondeur peut être générée à partir des différents points issus de l'interpolation. Lorsque le système est calibré métriquement, une profondeur métrique peut être associée à chaque point. Pour effectuer une reconstruction 3D de la scène, la carte de profondeur doit être projetée suivant trois axes et pour fournir un meilleur rendu visuel, la texture de la scène peut être appliquée sur la surface reconstruite. Le principal inconvénient de cette méthode est qu'il est nécessaire d'acquérir une série d'images, ce qui implique que la scène soit statique.

Profondeur par la défocalisation / Depth from defocus (DFD)

Une autre méthode a été développée pour supprimer cet inconvénient. Elle repose sur l'étude du flou d'une seule image. Elle a été proposée pour les systèmes imageurs à longueur focale fixe et se nomme *Depth from Defocus*. Elle nécessite une première étape pour déterminer a priori la Fonction d'Étalement de Point (FEP) (ou PSF : *Point Spread Function*) en mesurant la réponse de l'imageur pour des objets situés à des distances connues. Plus la PSF est importante, plus la distance objet-capteur l'est. Cette fonction a été modélisée de plusieurs manières : la fonction peut être considérée comme une fonction gaussienne [Pent87] ou bien être déterminée au moyen de la transformée de Fourier (TF), pour définir la fonction de transfert optique (OTF) [Subb88]. La fonction de flou peut être comparée à un filtre passe-bas dont la fréquence de coupure dépend de la focalisation et ce quelle que soit la modélisation choisie pour la PSF. Enfin, une opération de traitement est appliquée sur l'acquisition. Au cours de ce traitement, les objets sont projetés dans la base des FEP précédemment déterminée. Ainsi, chaque objet sera associé à une distance. Une cartographie 3D de la scène peut être générée à partir d'une image [EnLa93, TCL11]. Cette méthode peut être appliquée à n'importe quel imageur, ce qui représente un avantage par rapport à d'autres méthodes. L'étape de calibrage, consistant à caractériser les PSF en fonction de la profondeur, est une étape importante. De plus, la profondeur de champ de l'image reconstruite est faible contrairement à celle obtenue avec la méthode *Depth from focus*.

Les applications principales de ces deux méthodes sont le calcul de carte de disparité/profondeur et la reconstruction 3D, comme illustré sur la Figure 1.10 pour la DFF et sur la Figure 1.11 pour la DFD. De plus, comme pour l'exemple présenté pour la DFF, cette méthode peut être employée pour effectuer une extension de la profondeur de champ en microscopie.

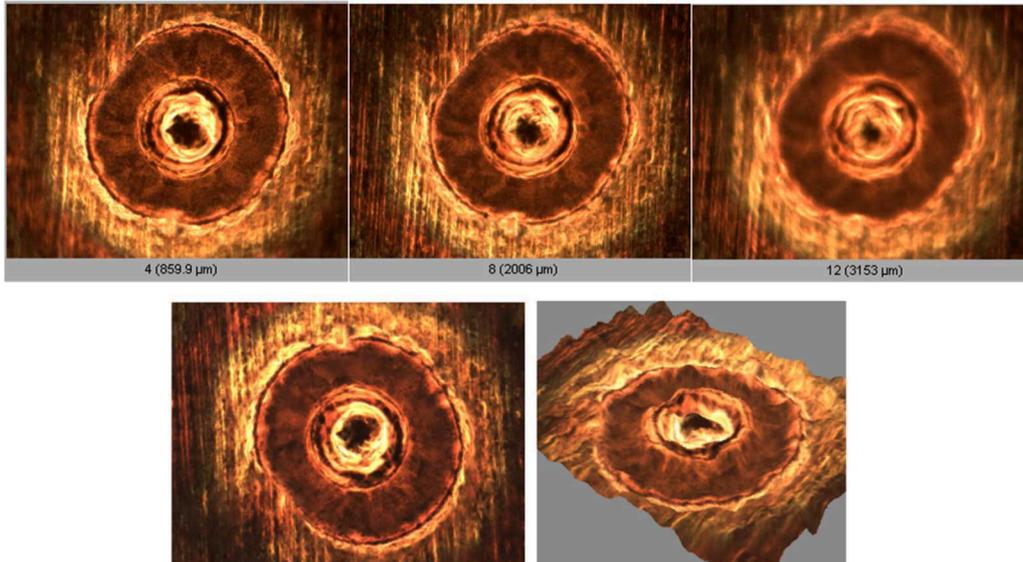


Figure 1.10 Illustration d'une application du DFF pour étendre le *Depth of Field* en microscopie [00]. Ici, on observe un impact de laser. En haut, trois images focalisées extraites de la pile des images focalisées. En bas : à gauche, image illustrant l'extension du *depth of field* et à droite, reconstruction 3D de l'impact laser.

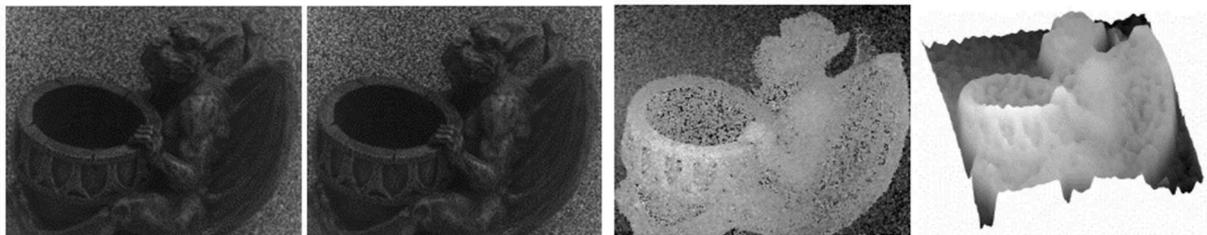


Figure 1.11 Exemples de carte de disparité et de reconstruction 3D en utilisant une technique de DFD [FaSo02]. Les deux premières images à gauche : deux images focalisées la première focalisée proche (à 0,70m) et la seconde focalisée loin (à 0,77m). Troisième image : profondeurs estimées représentées en niveau de gris et à image de droite : reconstruction 3D de la forme.

1.2.2.3. Approche lentille et ouverture

L'approche lentille et ouverture consiste à sélectionner une partie des chemins optiques passant par la lentille ce qui permet de réduire à la fois le flou et la quantité de lumière reçue. Cette sélection peut se faire par plusieurs approches. Les plus importantes sont regroupées dans la catégorie des caméras *light-field*, présentées dans la partie 1.3. Ces caméras servent à acquérir la fonction plénoptique, décrite dans la section 1.3.2.

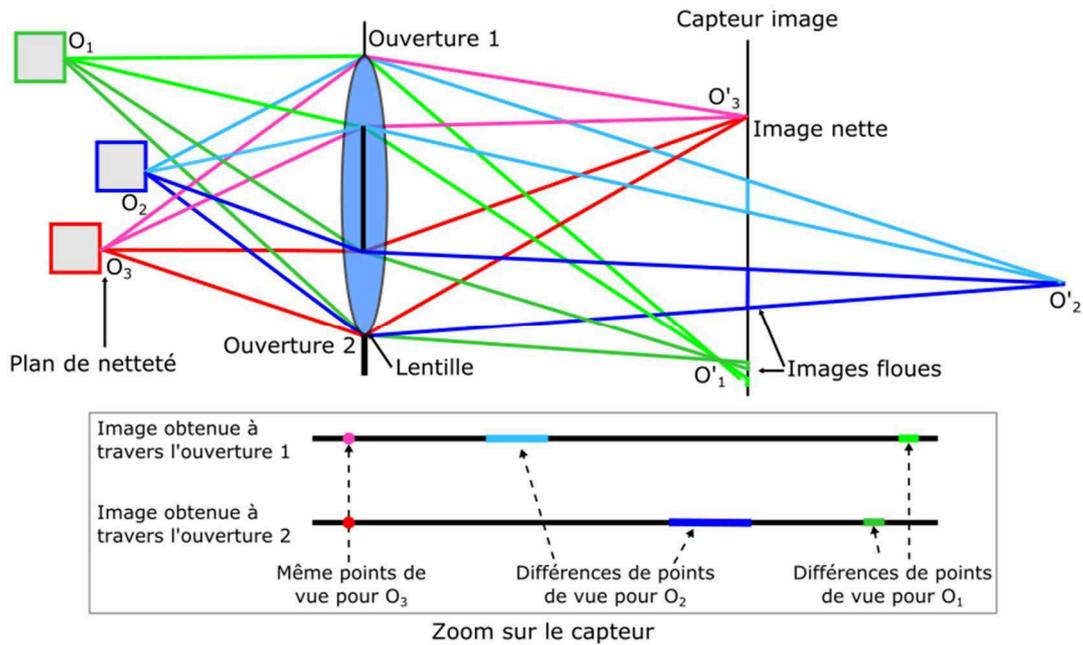


Figure 1.12 Modèle d'un imageur composé d'une lentille, avec deux ouvertures. En haut, représentation du tracé des rayons à travers les deux ouvertures. En bas, illustrations des différences de points de vue pour les différents objets entre les deux images obtenues chacune en visualisant une seule ouverture. Seul l'objet net O_3 a le même point de vue entre les deux images.

Caméras à ouverture codée

Dans le cas des caméras à ouverture codée, l'ouverture est modifiée à l'aide d'un masque. En fonction de la géométrie du masque utilisé, le flou dans la scène peut être interprété. A partir de cette étude, il est possible d'accéder à l'information de profondeur. Ce type de technique sera développé dans la section 1.3.3.

Caméras plénoptiques et leurs variantes

Les caméras plénoptiques sont des systèmes séparant les différents chemins optiques en des sous-faisceaux distincts les uns des autres et focalisés sur différentes parties du capteur. Plusieurs technologies de ce type de caméra seront détaillées dans la section 1.3.3.

1.2.3. Synthèse

Les principales techniques de vision 3D ont été énumérées dans ce chapitre. Nous présentons un tableau récapitulatif (Tableau 1.2), résumant ces techniques. C'est une version enrichie de la comparaison établie par Sansoni en 2009 [SaTD09]. Ainsi, en fonction des contraintes associées à l'application, la distance à mesurer et la précision désirée sont des éléments à prendre compte pour choisir le système d'imagerie 3D adapté.

<i>Technologies</i>	<i>Atouts</i>	<i>Limitations</i>	<i>Gamme de mesures</i>	<i>Gamme de mesure</i>
Triangulation laser	Simplicité de mise en œuvre Performances peu influencées par la lumière ambiante Bonne densité d'acquisition	Sécurité liée au laser Volume de mesure limité Données manquantes (ombres et occlusions) Prix	5 à 60 μm	Résolution x taille du capteur
Lumière structurée	Acquisition dense Volume de mesure de taille moyenne Performances peu influencées par la lumière ambiante	Contraintes de sécurité (si source laser) Complexité du calcul Données manquantes (ombres et occlusions) Prix	De 20 à 200 μm	De l'ordre du mètre
Temps de vol	Volume de mesure de moyen à large Bonne densité d'acquisition	Prix Précision inférieure à la triangulation (à faible distance de travail)	1 mm	De 1 m à 10 km (suivant la technologie)
Stéréovision	Simple et bon marché Bonne précision sur des cibles bien définies	Calculs coûteux Données éparses Limitée aux scènes bien texturées Faible densité d'acquisition	Incertitudes de 10 à 150 μm	Du macroscopique aux mesures métriques
Profondeur à partir de la focalisation	Simple et bon marché DéTECTEURS disponibles pour l'inspection de surface et la microprofilométrie	Champs de vue limités Résolution spatiale non uniforme Performances affectées par la lumière ambiante (si système passif)		
Profondeur à partir de la défocalisation	Utilisable sur n'importe quel imageur Après calibrage, besoin d'une unique image	Mémoire importante nécessaire Calibrage indispensable Profondeur de champ faible		
Déflexométrie	Applicable aux objets spéculaires et transparents	Limitation aux objets plans ou peu incurvés	Du cm^2 à plusieurs m^2	De l'ordre du mètre
Caméras plénoptiques	Nombreuses applications en plus des cartes de profondeurs	Calibrage indispensable et complexe	0,1 mm	Du micron au mètre (suivant le domaine)

Tableau 1.2 Comparaison des techniques optiques de mesures 3D adapté de [SaTD09]

1.3. Systèmes d'acquisition *light-field*

1.3.1. Historique

C'est véritablement en 1908, que le concept de l'acquisition *light-field* est né avec la « photographie intégrale » de Lippmann [LIPP08] (sans doute inspiré du « stéréogramme de Eugene » Ives [Ives03]). L'idée de Lippmann étant de capturer une scène dans son intégralité, afin d'avoir une restitution plus complète qu'une vue 2D (Figure 1.13).

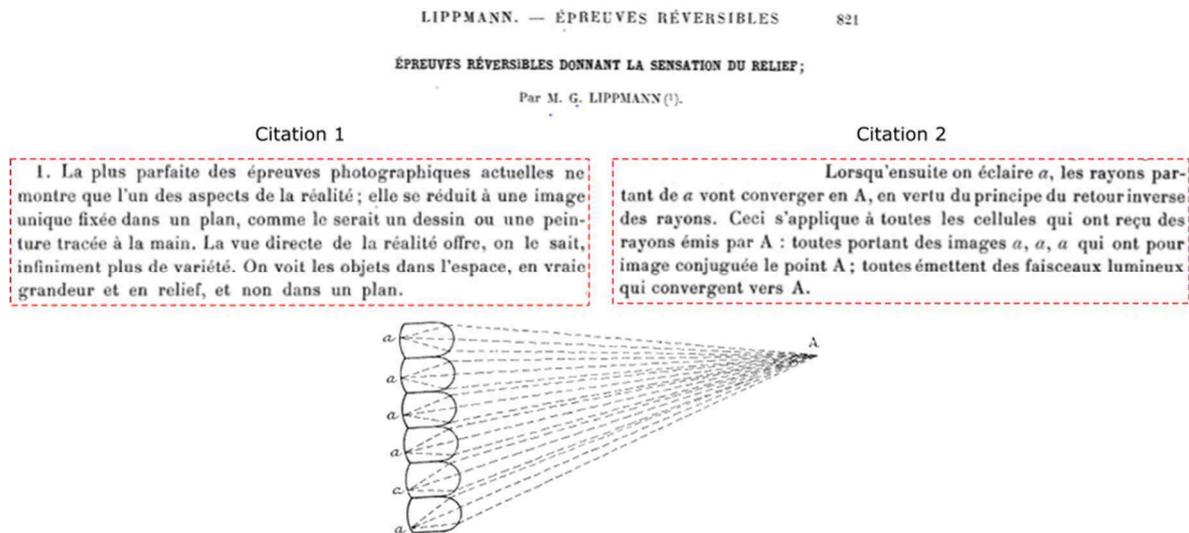


Figure 1.13 Illustration des épreuves réversibles et deux citations du papier de Lippmann [Lipp08]

Pour saisir une scène, Lippmann imagine une capture de plusieurs points de vue à l'aide d'un réseau de lentilles. Lippmann avait comme objectif d'exploiter la propriété de retour inverses des rayons, c'est-à-dire qu'en rétroéclairant tous les points de vue, il a imaginé une convergence des rayons vers la scène initiale. En 1930, Herbert Ives [Ives30] apporta une nouvelle amélioration par l'ajout d'une lentille principale. De plus, il proposa une méthode d'inversion pour estimer la profondeur. Cependant, du fait d'une limitation technique liée à la fabrication des microlentilles, ces systèmes sont restés théoriques. Par la suite, ce sont Adelson et Bergen qui, en 1991, ont introduit la définition de la fonction plénoptique, permettant de modéliser un rayon lumineux. En prolongement, c'est Levoy en 1996 [LeHa96], qui a proposé la définition actuelle du terme *light-field*.

1.3.2. Fonction plénoptique et *light-field*

La fonction plénoptique a été introduite par Adelson et Bergen en 1991 [AdBe91]. Le mot « plénoptique » vient du latin « plenus » signifiant complet ou plein. Ils ont défini une distribution géométrique totale de la lumière comme une fonction plénoptique, modélisant un rayon lumineux par ses positions et orientation, mais aussi par sa longueur d'onde et le temps. Le résultat est une fonction 7D définie comme suit :

$$P(\theta, \phi, \lambda, t, x, y, z), \quad (1.6)$$

où (θ, ϕ) sont les coordonnées angulaires représentant la direction du rayon, λ est la longueur d'onde, t est le temps, et (x, y, z) sont les coordonnées spatiales d'un point par lequel passe ce rayon.

Ce point est considéré comme l'origine de ce rayon. Cette fonction est donc basée sur un modèle *ray-based* pour la luminance énergétique. Représenter la lumière sous forme de rayons est une représentation habituelle en optique. La mesure de l'énergie de la lumière suivant un rayon donné est définie en radiométrie comme la luminance énergétique [BuPr66]. Mais, une autre modélisation de la distribution totale connue comme le 4D *light field*, a été introduite par Levoy et Hanrahan [LeHa96] et Gorthler et al. [GGSC96] (aussi nommé le *Lumigraph*). Le passage de 7D à 4D s'obtient en simplifiant la fonction plénoptique, avec les hypothèses et considérations suivantes :

- La propagation de lumière s'effectue en espace libre (pas d'occultation, même milieu)
- Illumination constante dans le temps (pas de t)
- Illumination monochromatique (pas de λ)

Les coordonnées angulaires (θ, ϕ) sont substituées par un vecteur (x, y, u, v) , où les coordonnées (u, v) et (x, y) sont respectivement les points d'intersections d'un rayon traversant deux plans parallèles, comme cela est illustré sur la Figure 1.14.

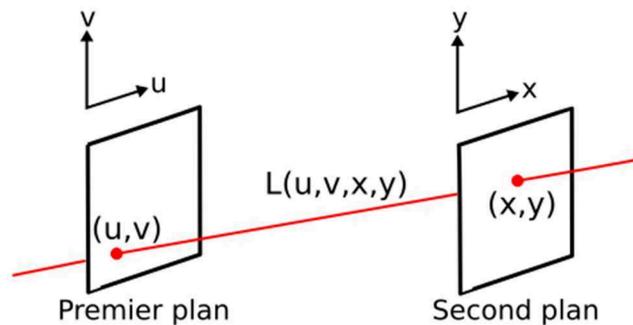


Figure 1.14 Paramétrage avec deux plans du light field [LeHa96]

L'écriture de la fonction plénoptique, est donc ramenée à une fonction 4D, appelée *light-field* :

$$L(u, v, x, y) \quad (1.7)$$

où la valeur de $L(u, v, x, y)$ correspond à la luminance, donc à la quantité de lumière d'un rayon lumineux monochromatique. On peut noter que dans le cas de la couleur, le vecteur *light-field* peut s'exprimer sous la forme d'un vecteur tridimensionnel RVB.

A partir de cette définition, Ng [Ng06] a proposé dans ses travaux de thèse, une représentation dite *ray-space*. Cette représentation dans un espace 4D (u, v, x, y) , est facile à aborder quand on la réduit à sa dimension horizontale, à travers le jeu de coordonnées 2D (u, x) (ou verticale avec (v, y)). Elle a alors pour mérite de mettre facilement en évidence des configurations optiques particulières, comme cela est illustré sur la Figure 1.15 :

- Figure 1.15 (a) : les rayons lumineux émanent depuis un ensemble de positions u régulièrement espacés vers un autre ensemble de x régulièrement espacés. Ainsi, le diagramme *ray-space* consiste en une matrice régulière de points.
- Figure 1.15 (b) : une collection de rayons traverse u et se focalisent en un seul point sur x . La focalisation d'un point se traduit donc par une ligne verticale dans le diagramme *ray-space*.
- Figure 1.15 (c) : on a une collection de rayons définie comme dans le cas précédent, mais qui ne focalisent pas sur x' . Le diagramme *ray-space* (u, x') se traduit par une ligne inclinée vers la droite.
- Figure 1.15 (d) : cas similaire, avec x'' , situé derrière x . Le diagramme *ray-space* se traduit par une ligne inclinée vers la gauche.

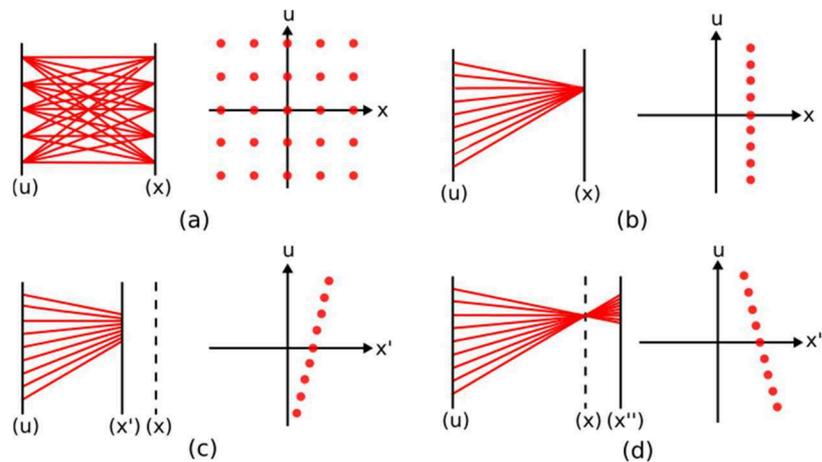


Figure 1.15 Exemples de représentations [Lam15]

Pour les exemples (b) et (c), si (x) et (x') représentent un axe image, et (u) la position d'une lentille mince, on peut facilement relier la pente de la droite de l'espace *ray-space* à la profondeur d'un point source diffusant ces rayons. De plus, pour l'exemple (c), la connaissance des coordonnées (u, x') amène à pouvoir calculer l'intersection des rayons sur l'axe (x) , et donc ainsi obtenir l'image focalisée du point source. Inversement, on peut également choisir un autre axe et calculer l'intersection du prolongement des rayons sur cet axe, et ainsi obtenir une image très dé-focalisée du point source. Ce concept expose donc ici à la fois capacités 3D d'un système *light-field*, et les bases de la refocalisation numérique. Au final, un système d'acquisition sera catégorisé comme *light-field*, dès que son architecture permettra un échantillonnage spatial des rayons lumineux. Notons également que ce mode d'imagerie appartient au domaine de l'imagerie computationnelle, puisque les données d'acquisition doivent être traitées pour fournir une image 2D ou 3D.

1.3.3. Systèmes d'acquisition *light-field*

Depuis les premiers travaux de Lipmann, beaucoup d'expérimentations ont été mises en œuvre, pour acquérir et échantillonner les rayons lumineux d'une scène. Les travaux les plus marquants sont ceux d'Adelson, qui en 1991 formalisait la fonction plénoptique et posait les bases de premières caméras plénoptiques avec une matrice de *pinholes* ou de microlentilles, pour échantillonner les rayons lumineux sous formes de vecteurs *light-field*. Par la suite, les travaux de Ng ont véritablement mis en avant

la première caméra plénoptique, et les concepts algorithmiques associés pour reconstruire des images 2D ou 3D. Parallèlement, le concept de réseaux de caméras c'est aussi développé dans différentes équipes de recherche, et ce sont les travaux de l'Université de Stanford qui sont les plus connus [WJVT05] dans le domaine. Un concept proche du réseau de caméras est proposé par Tanida [TKYM01] avec la caméra TOMBO, qui partage son ouverture entre plusieurs voies optiques. Les systèmes de caméras à ouverture codée, utilisés en général en imagerie à rayons X ou gamma, peuvent aussi être considérés comme une caméra *light-field* dans le domaine visible.

Le Tableau 1.3 ci-dessous résume les principales technologies qui seront décrites dans la suite de ce paragraphe.

<i>Approches</i>	<i>Principes</i>	<i>Commentaires</i>
<i>Multiplexage temporel</i>	Caméras à ouverture codées	Echantillonnage des rayons lumineux par un jeu de masques placé successivement dans le chemin optique d'une caméra. Avantage : assez simple à mettre en œuvre Inconvénient : plusieurs acquisitions nécessaires
<i>Multiplexage spatial</i>	Caméra plénoptique	Par définition, une caméra plénoptique est constituée d'un objectif principal et d'un réseau de microlentilles placé devant un capteur image. L'échantillonnage des rayons est dépendant du nombre de microlentilles du réseau utilisé. Avantage : système compact. Inconvénient : nombreux paramètres intrinsèques à déterminer durant le calibrage, résolution en Z limitée par les distances entre microlentilles et les dimensions du capteur image.
	Caméra multi-vues ou caméra TOMBO	L'échantillonnage des points de vue se réalise au moyen d'une matrice de lentilles placées devant le capteur image (typiquement de 4 à 16 vues). Un de nos prototypes est réalisé suivant ce principe, avec l'utilisation de 4 mini-lentilles placées devant un capteur image. Avantage : acquisition d'une seule image. Combinaison possible des vues avec des algorithmes de superrésolution, pour améliorer la résolution spatiale des sous-images. Inconvénient : résolution en Z limitée par les distances entre lentilles et les dimensions du capteur image.
	Réseaux de caméras	Principe d'acquisition identique à une caméra multi-vues, avec un encombrement plus important, mais aussi une <i>baseline</i> entre vues permettant de travailler à une plus grande distance. Avantage : grande profondeur de travail Inconvénient : complexité matérielle pour gérer n caméras à déclencher simultanément. La gestion des n images demande des ressources de calculs importantes.

Tableau 1.3 Principales technologies de caméras *light-field*, regroupées en deux approches : multiplexage temporel et multiplexage spatial.

1.3.3.1. Systèmes à ouverture codées

Les systèmes basés sur les ouvertures codées sont caractérisés par l'ajout d'un masque avec des ouvertures sur le trajet lumineux (Figure 1.16). Cette solution a été proposée pour différentes applications et notamment pour l'estimation de profondeur [LFDF07]. Les ouvertures codées ou pupilles codées sont des grilles, ou d'autres motifs de matériaux optiques que l'on place le plus souvent entre l'objectif et le capteur. Initialement, cette méthode était utilisée pour l'imagerie basée sur les rayons X et gamma en astrophysique [CSDN87] et pour l'imagerie médicale [AcGL01]. La lumière est donc bloquée ou non, par le motif choisi. Le changement de la forme de l'ouverture équivaut à décaler la perspective du point de vue.

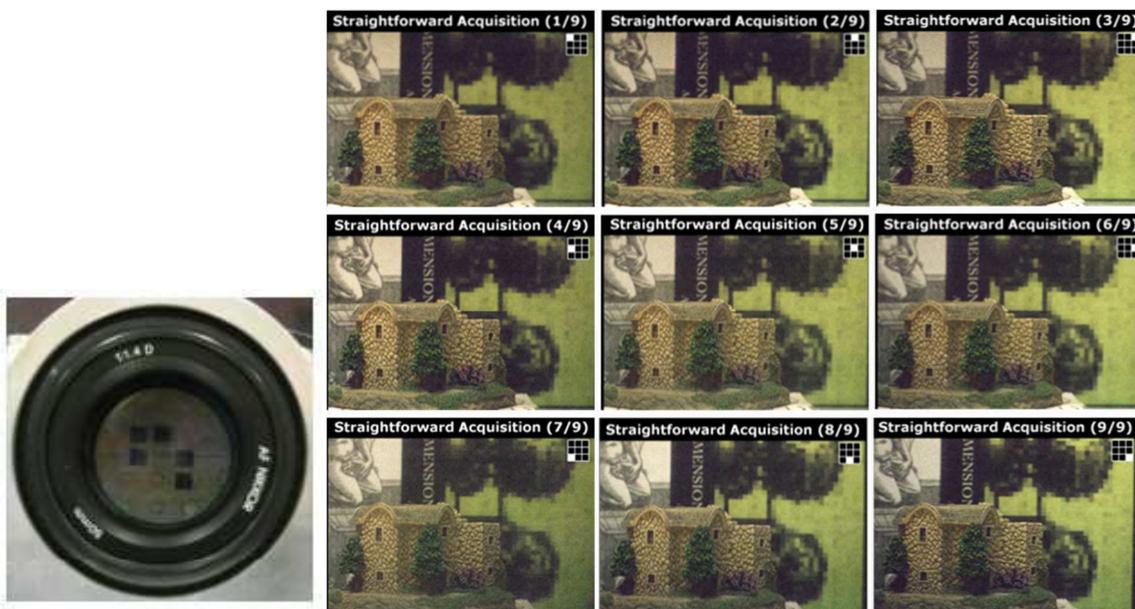


Figure 1.16 A gauche, exemple de masque d'une caméra à ouverture codée. A droite, exemples de masques et d'images acquises avec ces masques. En haut à droite de chaque vue, la forme du masque est représentée [LLWL08].

Liang et al. [LLWL08] ont proposé un système d'ouverture programmable dédié à l'acquisition de lights-field pleine résolution. Ces systèmes correspondent à des caméras à ouverture codée et programmable. Pour améliorer l'efficacité lumineuse, le masque utilisé a été défini en se basant sur le codage d'Hadamard. Pour le post-traitement, un nouvel algorithme de calibrage a été employé afin de réduire les effets de vignettage entre les différentes vues. Dans leur prototype, un écran faible coût à cristaux liquides (LCD) sans couche réfléchive a été introduit sur le chemin optique pour contrôler électroniquement l'ouverture. L'inconvénient est que la transmission du pixel LCD est faible et qu'il ne peut pas être complètement éteint. Des alternatives au LCD sont les technologies Liquid Crystal on Silicon (LCoS) [ZoNa06] et Digital Mirror Devices (DMD) [DaLD12]. Cependant, ces composants sont chers et leur intégration dans le système est complexe.

Levin et al. [LFDF07] ont proposé d'ajouter un élément occultant à motifs au niveau de l'ouverture de la caméra créant une ouverture codée. Un critère sur la profondeur a été utilisé pour déterminer le

motif privilégié, utilisé pour coder l'ouverture. L'information de profondeur et l'image totalement focalisée peuvent être retrouvées à partir d'une unique image acquise avec la caméra modifiée. Une carte de profondeur peut ensuite être extraite. L'image nette et la carte de profondeur peuvent être combinées pour des applications photographiques variées telles que la segmentation automatique de scène, la refocalisation ou le changement de point de vue.

Deux configurations de caméra à ouverture codée ont été proposées par Veeraraghavan et al. [VRAM07]. La première consiste à placer le masque proche de l'ouverture de la lentille et pour la seconde, le masque est proche du capteur image de la caméra. La première configuration est nommée *encoded blur*. Un masque grossier à large bande est utilisé. Pour obtenir une image nette, il est nécessaire de déconvoluer l'image flou avec le masque, dont la taille dépend du degré de flou. La seconde caméra est nommée caméra *light-field* hétérodyne. Le masque utilisé est basé sur un cosinus 2D à bande étroite pour avoir l'effet d'un décalage de la fréquence sur une porteuse. Pour obtenir le *light-field* 4D, les traitements sont effectués dans le domaine de Fourier. Ces deux configurations sont faciles à mettre en œuvre et leur coût est faible. De plus, elles permettent de s'affranchir des aberrations sphériques, chromatiques, de coma et de mauvais alignement. Enfin, le fait que le masque soit amovible permet de choisir d'ajuster la résolution spatiale et angulaire. D'une façon générale, les masques absorbent une partie de la lumière dans le chemin optique et réduisent également l'ouverture de la lentille, ce qui augmente dans les images, le flou causé par la diffraction. On peut également noter que les algorithmes de déconvolution utilisés pour la refocalisation numérique introduisent un bruit résiduel sur les résultats.

Plus récemment, Reddy et al. [ReBR13] ont proposé une caméra *light-field* avec un masque externe et amovible. Ces auteurs se basent sur le fait qu'un masque externe est mathématiquement équivalent à un masque interne, qui serait placé proche du plan d'ouverture. Le paramétrage du *light-field* est alors obtenu avec le plan image u et celui du masque virtuel. Les pentes des droites obtenues sont alors directement liées à la profondeur des points de la scène. Cette approche avec un masque externe facilite la mise en œuvre des systèmes à ouverture codées. Néanmoins, les temps de reconstruction et de calcul restent importants, et le déploiement pour des applications temps-réel reste limité.

1.3.3.2. Caméras plénoptiques et leurs variantes

La seconde grande famille des systèmes *light-field* sont les caméras plénoptiques. Elles présentent la particularité d'avoir une grille de microlentilles qui est insérée dans le chemin optique d'une caméra, entre l'objectif principal et le plan image. Le rôle des microlentilles est d'échantillonner l'orientation des rayons passant dans le chemin optique, et donc ainsi de pouvoir caractériser une scène à travers une représentation paramétrique de ses rayons. Comme pour les caméras à ouvertures codées, nous sommes dans un cadre de *computational photography*, c'est-à-dire que les données brutes capturées sont inexploitable sans étape de traitement.

Ce paragraphe va présenter les principales familles de caméras plénoptiques, à savoir la version 1.0, la version 2.0, et deux variantes : la version 1.2 et la caméra kaléidoscope.

Caméras plénoptiques 1.0 (dites standard ou defocused plenoptic ou unfocused plenoptic)

La première configuration de la caméra est dite 1.0, standard, *defocused* ou *unfocused plenoptic*. Adelson et Wang l'ont proposée et Ren Ng l'a étudiée. Le détecteur est placé dans le plan focal des microlentilles, situées dans le plan focal de l'objectif principal de la caméra. Ainsi, la lumière passant par l'objectif est focalisée dans le plan des microlentilles, comme montré par la Figure 1.17. Tous les rayons issus d'un même point objet, passant par une certaine zone de la lentille, sont imagés par un unique pixel derrière la microlentille éclairée. Chaque pixel perçoit une direction particulière du cône d'incidence. Les pixels fournissent l'information angulaire de la fonction plénoptique (plan (u, v)) tandis que l'information spatiale (plan (x, y) de la fonction plénoptique) est donnée par la position des microlentilles. Chaque pixel représente alors un vecteur *light-field* $L(u, v, x, y)$, à l'image du vecteur de l'expression (1.7).

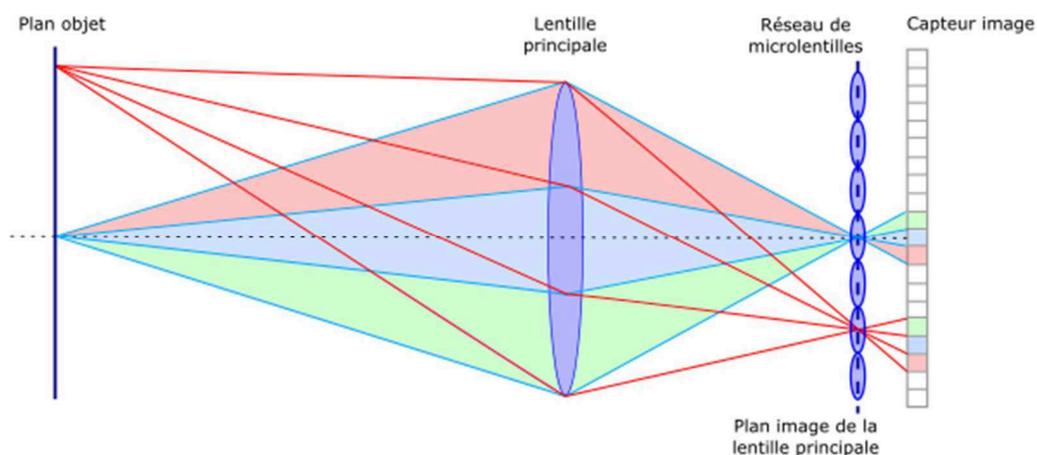


Figure 1.17 Schéma représentant la configuration 1.0 des caméras plénoptiques. Le réseau de microlentilles est placé dans le plan image de la lentille principale.

D'un point de vue géométrique, la pupille de l'objectif est subdivisée, ce qui équivaut à plusieurs caméras se partageant l'espace de la pupille. Avec cette configuration, le capteur image contient à la fois des informations spatiales, mais aussi des informations angulaires. La résolution spatiale d'une image, qui correspond classiquement à la résolution du capteur utilisé, se trouve donc naturellement réduite, au profit d'une information angulaire, qui est directement liée à la profondeur des objets dans une scène. Des systèmes commerciaux existent. Le plus connu est la caméra Lytro qui est utilisée principalement dans les domaines de la photographie et du cinéma.

Comme tous les dispositifs LF, l'image brute est inexploitable (Figure 1.18), et des algorithmes de reconstruction vont pouvoir être déployés pour exploiter l'information de profondeur à travers des cartes de disparité ou des images refocalisées. Les principes seront énoncés dans le paragraphe 1.4.

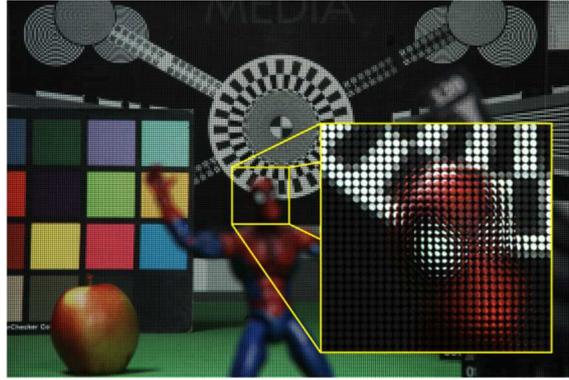


Figure 1.18 Exemple d'une image brute avec une caméra plénoptique 1.0 et zoom sur une partie de l'image brute [Hahn16].

Caméras plénoptiques version 2.0 (ou focused plenoptic)

En 2009, Georgiev et Lumisdaine [LuGe09] ont proposé une autre configuration de caméra plénoptique. Elle est nommée « 2.0 » ou caméra plénoptique focalisée. La différence avec la configuration 1.0 réside dans la position du capteur, qui coïncide avec le plan image des microlentilles, et le réseau de microlentilles, qui n'est plus positionné dans le plan image de la lentille principale, mais dans un plan intermédiaire. Cela donne des sous images refocalisées et l'intérêt est de bénéficier d'une meilleure résolution spatiale [GeLu09]. On peut ajouter que cela revient à juxtaposer des points de vue d'une même scène, qui ont des angles d'ouverture très faibles. L'introduction de ce plan intermédiaire fournit deux possibilités le concernant. La première lorsqu'il est situé entre l'objectif principal et le réseau de microlentilles, comme sur la Figure 1.19 (a). Elle est appelée « configuration télescopique » ou « plan intermédiaire réel ». Pour la seconde, le plan intermédiaire est situé après les microlentilles (Figure 1.19 (b)). Elle est aussi nommée « configuration binoculaire » ou « plan intermédiaire virtuel ». Cette seconde configuration permet de rendre plus compact le système. Le grandissement est positif. Cependant, un plus fort recouvrement entre les différentes sous-images est observable.

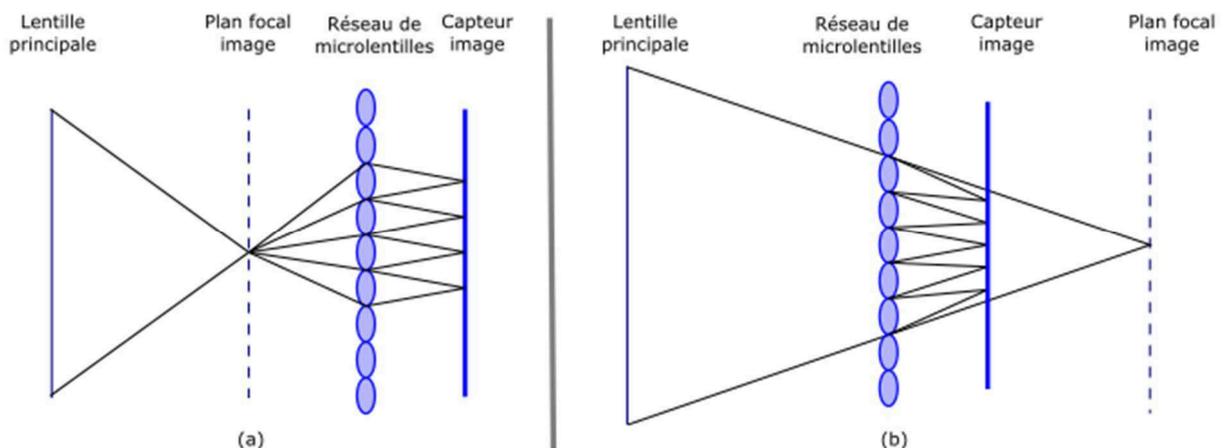


Figure 1.19 Les deux possibilités de positionnement du plan intermédiaire (a) « plan intermédiaire réel » ; (b) « plan intermédiaire virtuel »

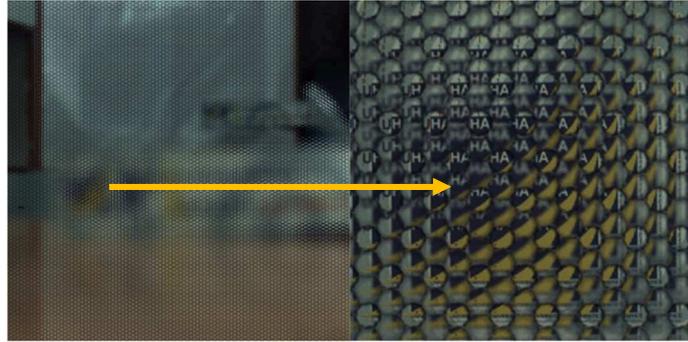


Figure 1.20 Image brute et zoom sur une image délivrée par une caméra plénoptique 2.0 Raytrix

La résolution spatiale correspond au rapport entre le nombre de pixels du capteur et le nombre de microlentilles imageant un même point sur le capteur. La résolution dépend donc de la matrice de microlentilles. La portée du système est définie par l'objectif. Les images acquises avec cette configuration présentent aussi du vignettage. Cependant, dans ce cas, il aura pour effet de faire chuter la luminosité en bord de champ comme pour une caméra conventionnelle.

Une évolution de cette configuration a été proposée afin d'augmenter les distances de travail. Les microlentilles sont prévues pour procurer des images nettes dans une gamme de 3 distances (proche, distance intermédiaire, loin). Pour cela, il suffit d'utiliser une nouvelle version de grille de microlentilles regroupant différents types de microlentilles avec des focales différentes [GeLu12]. La version commerciale, proposée par la société *Raytrix*, est composée de trois types entrelacés de microlentilles avec des focales différentes. Cet arrangement est judicieux car il y a toujours une microlentille de même type qui est voisine de celle étudiée. Ainsi pour obtenir l'information de profondeur, un point doit être imagé par plus de microlentilles que pour la configuration 2.0 classique.

Pour cette configuration, l'image finale de la scène se situe dans le plan focal des microlentilles. Ainsi, une partie de la scène est imagée par chaque microlentille avec un certain point de vue. Ce dernier est décalé entre chaque sous-image à cause de la parallaxe, dépendant des paramètres du système. Cette information sur la parallaxe est exploitée par les traitements.

Avec cette configuration, il est plus facile de sélectionner le ratio entre résolution spatiale et résolution angulaire. Il est préférable d'utiliser ce type de caméra pour les applications nécessitant à la fois une information de profondeur et une bonne résolution spatiale. Le vignettage ne génère pas de zones aveugles. La profondeur de champ peut être augmentée en utilisant une grille de microlentilles comprenant plusieurs types de microlentilles avec des focales différentes. Par contre, les traitements sont plus complexes que pour la configuration 1.0, et cette complexité est accrue par le vignettage qui diminue l'éclairement au bord des microlentilles.

Variantes des caméras plénoptiques

Sachant que les principaux défauts des systèmes plénoptiques sont la faible résolution angulaire et les effets de vignettage, d'autres concepts moins connus ont été proposés pour améliorer ces deux points.

Caméra plénoptique 1.2

Le concept de caméra plénoptique 1.2 a été proposé par Todor [GeLu10]. Cette architecture a été conçue pour remédier aux problèmes de résolution de la configuration 1.0 avant que la configuration 2.0 ne soit envisagée. L'architecture est constituée d'une grille de microlentilles qui est placée entre la scène et l'objet et d'une matrice de prismes, située derrière le réseau de microlentilles pour rediriger chaque micro image sur l'objectif. D'une microlentille à l'autre, l'angle du prisme varie. Dans ce système, l'image perçue par l'objectif est une image virtuelle formée par des microlentilles divergentes. L'image acquise par le détecteur est semblable à celle acquise par une matrice de caméras. Le champ de vue de la caméra dépend de la focale des microlentilles. Cette architecture permet de maintenir une meilleure résolution spatiale que la configuration 1.0 et de réduire les effets de vignettage, sous réserve que les angles de déviation de la matrice de prismes soient bien déterminés. L'ouverture de la caméra est également indépendante de son champ de vue. L'un des inconvénients réside dans la nécessité d'avoir un objectif présentant une bonne capacité de mise au point. Ce besoin est causé par la proximité de l'image virtuelle à imager (générée par la matrice de microlentilles) avec l'objectif. La matrice de prismes génère du chromatisme, dégradant ainsi l'image finale. L'objectif choisi et la scène observée déterminent les caractéristiques de la matrice de microlentilles ainsi que celles de la matrice de prismes. Cet inconvénient majeur explique la préférence pour la configuration 2.0 qui peut être mise en œuvre plus facilement et qui s'adapte à différentes configurations de scènes.

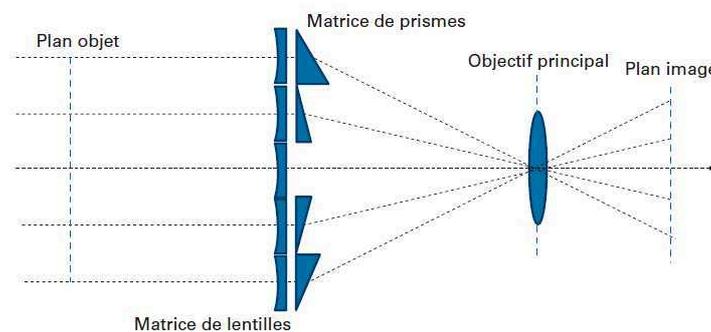


Figure 1.21 Schéma représentant l'architecture des caméras plénoptiques 1.2. Dans ce cas, la matrice de microlentilles est placée devant l'objectif principal et une matrice de prismes est positionnée entre les microlentilles et l'objectif principal [CDBV16].

Caméra kaléidoscopique

La caméra kaléidoscope est composée de miroirs plans placés avec une disposition kaléidoscopique. Ce dispositif a été proposé par l'équipe d'Ivo Ihrke [IRMS12]. La chambre contenant les miroirs a pour forme le tronc d'une pyramide. Elle est composée de plusieurs objectifs. Dans ce type de dispositif, les vues sont arrangées sphériquement autour de l'objet ce qui génère une haute variabilité des points de vue. Grâce à l'observabilité des écarts entre les miroirs et l'objet, il est possible de reconstruire l'enveloppe visuelle (*visual hull*) de l'objet, correspondant à la méthode *shape from silhouette*. Une étape de marquage de chaque pixel est nécessaire. Ce marquage consiste à associer chaque pixel avec le miroir plan pour lequel son rayon correspondant intersecte l'objet. Si la géométrie de l'objet est connue, le

marquage des pixels est trivial. En utilisant l'enveloppe visuelle, une vue virtuelle de l'objet peut être déterminée pour chaque pixel. Ces données multi-vues sont calibrées géométriquement et photométriquement. De plus, les données de sortie peuvent être directement utilisables avec des algorithmes standards de reconstruction multi-vues. En raison de la vaste gamme de vues réalisables avec ces systèmes, il est possible d'imager chaque point surface à partir d'un grand nombre de directions. Il pourrait être possible de réaliser l'évaluation géométrique et l'évaluation de la réflectance d'objets dynamiques. La profondeur de champ est ajustable en modifiant l'ouverture du système via le réglage de l'ouverture du second objectif. Le nombre de vues dépend de la longueur des miroirs. De plus, la résolution en profondeur dépend de l'encombrement du système.

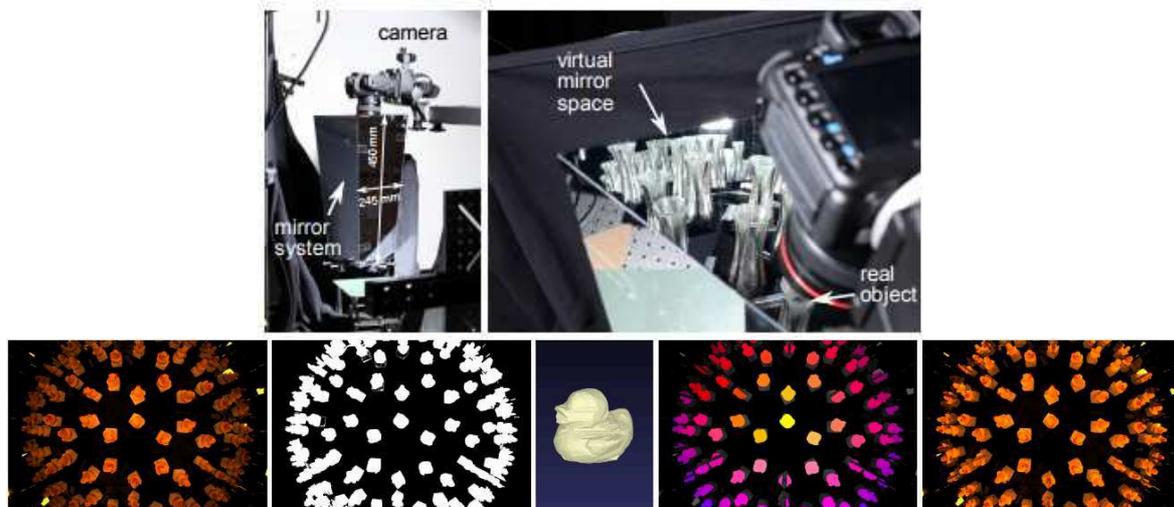


Figure 1.22 Illustrations d'une caméra kaléidoscopique. Sur la première ligne, à gauche, vue d'ensemble du montage et à droite, vue du dessus mettant en évidence les miroirs. Sur la deuxième ligne, de gauche à droite : image brute, image silhouette, reconstruction 3D de l'objet, segmentations, image compensée [IRMS12].

1.3.3.3. Systèmes multi-caméras et multi-vues

Ces systèmes sont caractérisés par un agencement permettant de capturer simultanément plusieurs vues d'une même scène, au moyen d'un réseau (matrice) de caméras ou bien de plusieurs objectifs placés devant un même capteur image (caméra multi-vues). Du point de vue de l'acquisition du *light-field*, il s'agit de l'une des méthodes les plus simples à mettre en œuvre afin de récupérer plus d'informations. Les caméras utilisées dans la majorité des cas sont des caméras 2D conventionnelles. Le nombre de caméras et leur disposition les unes par rapport aux autres dépendent du domaine d'utilisation et de l'usage des données acquises.

Les systèmes multi-caméras existent depuis longtemps, puisque qu'on pourrait inclure dans cette classification la stéréovision ou les systèmes trinoculaires ou tri-focaux. Cependant, ils ne sont pas étudiés sous l'aspect de systèmes *light-field*. Un des réseaux denses les plus connus et étudié dans un cadre de la vision *light-field*, a été proposé par Wilburn et al. [WJVT05] de l'université de Stanford. Ce réseau est composé de 10 par 10 caméras avec une résolution spatiale de 640 par 480 pixels pour chacune des caméras. Chaque capteur image est connecté à un FPGA et le flux vidéo est transféré à un PC par

une interface FireWire. En effet, ce réseau supporte des configurations physiques flexibles comme la séparation physique entre chaque caméra, le temps d'exposition et la synchronisation du déclenchement de l'obturateur. Avec des configurations différentes, le réseau est capable de capturer le *light-field* mais aussi des images à haute vitesse, haute dynamique, haute résolution et grand angle. Les auteurs ont aussi démontré qu'il est possible de voir des objets derrière des occlusions partielles comme des broussailles en sélectionnant de manière adaptative les rayons qui coupent les objets obturants ; cette technique est connue comme la photographie d'ouverture synthétique [VWJL04]. La modélisation entre les vues a été étudiée avec des homographies, cependant, le système n'était pas calibré en profondeur, même si la notion de disparité (mesurée en pixels) est bien présente.

Caméra TOMBO

Le concept, proposé par Tanida [TKYM01] et son équipe, a été développé en 2001. Il se nomme TOMBO pour *Thin Observation Module by Bound Optics*. Contrairement aux caméras plénoptiques, ce dispositif ne comporte pas d'objectif principal. Ce système est composé de $N \times N$ voies identiques qui permettent de former $N \times N$ images de la même scène sur un unique capteur. L'image finale est reconstruite à l'aide d'algorithmes de super résolution à partir des imagerie. Sa résolution peut être plus importante que celle des imagerie. Pour générer $N \times N$ images non redondantes, il faut que la période de la microlentille soit multiple du pas d'échantillonnage. Un effet de parallaxe permettant de remonter à l'information de profondeur existe grâce à l'écartement entre les différentes voies optiques. Plus le champ de vue du système est important, meilleure est l'ouverture de la caméra. Pour ce système, la parallaxe dépend de la taille du capteur. Ainsi, il est préférable d'utiliser une caméra plénoptique si on désire un dispositif à champ de vue.

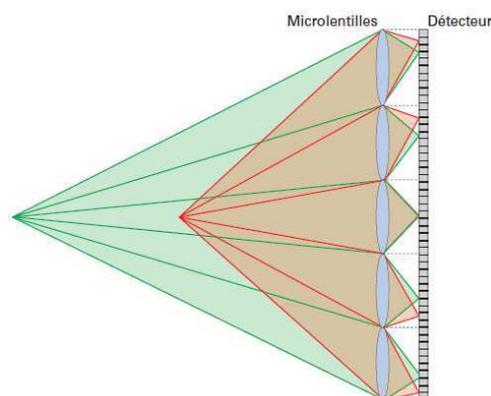


Figure 1.23 Illustration de l'architecture d'une caméra TOMBO. Sa principale différence avec une caméra plénoptique provient du fait que le dispositif ne comporte pas d'objectif principal [CDBV16].

La caméra *Picam* est la version commerciale la plus aboutie. Il s'agit d'une caméra pour téléphone portable [VLDM13]. Ce système permet de capturer un *light-field* 4D dense en une unique acquisition. Afin de rendre le module aussi fin que possible, un réseau de 4 par 4 lentilles est utilisé. Comme cette caméra ne peut pas ajuster la focale, l'ouverture de chaque lentille a été réduite afin d'avoir un large

champ de vue allant de quelques millimètres à l'infini. L'utilisation de petites ouvertures rend le rapport signal sur bruit faible. La qualité de l'image est ainsi limitée par la diffraction. Grâce à des algorithmes de superrésolution, précédés de prétraitements tels que la normalisation et la mesure de la parallaxe, une image de haute résolution avec peu de bruits peut être générée à partir des 16 images. Cependant, la limite de diffraction des lentilles individuelles ne peut pas être dépassée. Ce dispositif est très compact (taille d'une pièce de monnaie), mais les traitements d'images nécessaires pour reconstruire des images 2D ou 3D sont encore une charge de calculs importante des processeurs de mobile. C'est son principal inconvénient.

1.4. Calibrage des systèmes *light-field*

Pour calibrer les systèmes *light-field*, il existe différentes méthodes dont les principales vont être présentées en fonction de la catégorie de systèmes auxquels elles s'appliquent. Pour rappel, le calibrage d'une caméra conventionnelle consiste à modéliser la formation de l'image sur le capteur. Il sert à relier les coordonnées monde avec les coordonnées du point image sur le capteur. Lorsqu'il s'agit de systèmes multi-vues ou multi-caméras, le calibrage est utilisé pour définir les paramètres intrinsèques de chaque caméra/vue et les paramètres extrinsèques situant les caméras/vues dans le repère monde. Comme présenté dans la partie 1.3.3, la plupart des caméras plénoptiques contiennent des microlentilles, la principale différence avec les caméras conventionnelles est d'accéder au centre des microlentilles. Certains des systèmes *light-field* sont constitués d'un réseau dense de caméras, telles que la caméra ProFusion (entreprise PTGrey) ou la Stanford Multi-Camera Array [VGTA07, VWJL04]. Ces dernières nécessitent l'acquisition de l'image pour chaque caméra. Quel que soit le type de caméra plénoptique utilisé, la modélisation et le calibrage sont des étapes essentielles pour récupérer la profondeur dans une scène.

Cette section est divisée en deux parties : une première traitant du calibrage des systèmes multi-caméras et la seconde présentant le calibrage des caméras plénoptiques.

1.4.1. Calibrage des systèmes multi-caméras

Afin d'avoir un rendu *light-field* avec des dispositifs multi-caméras, différentes solutions pour les modéliser et les calibrer ont été proposées. En 1999, Koch et al. [KHPV99] ont proposé une approche géométrique basée sur le *Structure from Motion* pour le calcul du calibrage. Les données de calibrage sont calculées directement à partir des données de l'image par suivi d'éléments caractéristiques. Plus tard, Ueshiba et al. ont proposé une méthode de calibrage nécessitant un motif plan de référence pour calculer des matrices homographiques entre les images des caméras [UeTo03]. Elle est basée sur la méthode de Zhang, définie pour le calibrage d'une caméra [Zhan00]. Cependant, cette méthode nécessite des informations additionnelles pour surmonter le fait que la matrice de mesures a une échelle inconnue. Une méthode plus pratique est proposée par Svoboda [SvMP05]. Au lieu d'utiliser un motif

plan, elle utilise comme objet de calibrage un spot lumineux bougeant librement. Néanmoins, cette méthode se limite aux systèmes multi-caméras statiques. En 2004, Vaish et al. [VWJL04] ont développé une méthode de calibrage pour les grands réseaux de caméras. Elle se base sur une représentation plan + parallaxe de l'environnement. Cette représentation permet de calculer des matrices homographiques qui relient les différentes caméras entre elles, mais elle nécessite de choisir une référence. L'avantage de cette modélisation est qu'elle permet de s'affranchir du calcul des paramètres intrinsèques de chaque caméra. Si les caméras se trouvent dans un plan parallèle au plan de référence, la mesure de la parallaxe permet de retrouver la position des caméras. On peut comparer cette méthode à la rectification des systèmes de stéréovision. Nous verrons que notre méthode, présentée dans le Chapitre 3, basée sur l'homographie variable se rapproche de celle-ci. En 2015, Xu et al. ont proposé une méthode de calibrage pour les systèmes multi-caméras mobiles [XMNT15]. La représentation spatio-angulaire du *light-field* est calculée en premier et le calibrage est basé sur la méthode de Zhang [Zhan00] pour définir les paramètres intrinsèques initiaux. Ensuite, en utilisant une méthode d'optimisation globale, les paramètres intrinsèques sont affinés. Cette méthode est flexible puisqu'elle permet de choisir le nombre de points de vue et l'affinement des paramètres intrinsèques est optionnel.

1.4.2. Calibrage des caméras plénoptiques

Les paramètres intrinsèques des caméras plénoptiques 1.0 et 2.0 sont listés dans le Tableau 1.4 et certains sont représentés par la Figure 1.24. Les paramètres intrinsèques entre les deux configurations sont identiques sauf pour la distance entre microlentille – capteur qui est égale à la focale des microlentilles pour la plénoptique standard mais qui ne l'est pas pour la version focalisée. De plus, chacune des méthodes proposées ne fournit pas le même nombre de degrés de liberté à chacun des paramètres qu'elle utilise, sachant que certains ne sont pas employés. Malgré ces similitudes, les approches pour calibrer ces caméras sont différentes et dépendent de leur configuration.

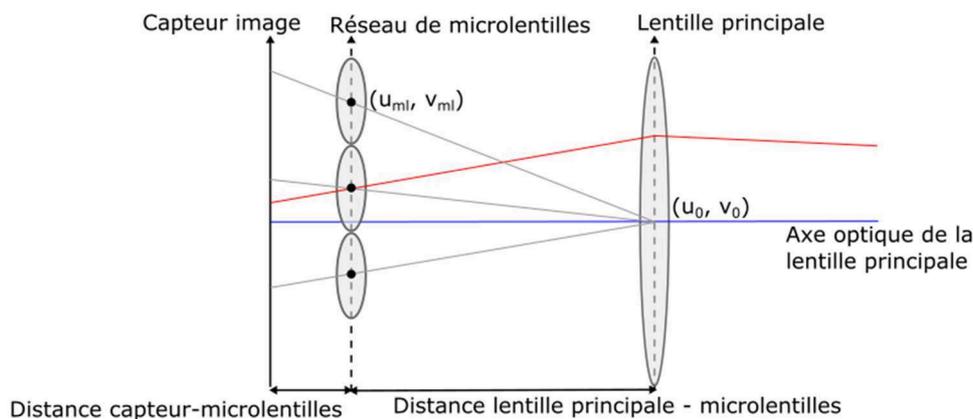


Figure 1.24 Représentation de certains des paramètres intrinsèques de caméras plénoptiques quel que soit leur configuration. (u_{ml}, v_{ml}) sont les coordonnées d'un des centres des microlentilles et (u_0, v_0) sont les coordonnées du centre de la lentille principale.

Éléments	Grandeurs	Plénoptique 1.0	Plénoptique 2.0
Lentille Principale	F_{LP}	Oui	Oui
	u_0	Oui	Oui
	v_0	Oui	Oui
	Distorsion	Oui	Oui
Micro lentille	u_{ml}	Oui	Oui
	v_{ml}	Oui	Oui
	Période u	Oui	Oui
	Période v	Oui	Oui
	Distorsion	Oui	Oui
	Distance microlentille - capteur	f_{ml}	Distance microlentille - capteur
	Distance lentille principale - microlentille	Oui	Oui
	Alignement de la grille des microlentilles	Oui	Oui

Tableau 1.4 Paramètres intrinsèques des caméras plénoptiques 1.0 et 2.0. Les 12 paramètres sont regroupés par éléments auxquels ils se réfèrent et montrent la seule différence qui existe entre les types de configurations. Celle-ci correspond à la distance entre les microlentilles et le capteur : pour la version standard, elle est égale à la focale des mini-lentilles, tandis que pour la version focalisée, elle correspond bien à cette distance.

Concernant les caméras plénoptiques standards, les approches de calibrage sont assez classiques. Elles vont réduire le nombre de paramètres intrinsèques en éliminant les moins influents tels que la distorsion ou l'alignement. C'est le cas des travaux de Cho et al. [CLKT13]. Ils ont montré l'importance de la connaissance des paramètres de calibrage pour la reconstruction d'images de haute qualité. Bok et al. [BoJK16] ont proposé une approche assez similaire à celle de Cho, la différence principale réside dans le fait d'utiliser une représentation sous forme de segments. Pour ces caméras, les travaux les plus significatifs sont ceux de Dansereau [DaPW13]. Au lieu d'utiliser une représentation point à point, il a utilisé une représentation coordonnées image vers vecteur *light-field*, comme représenté par la Figure 1.25. Cette méthode regroupant 1 paramètre de distorsion et 12 paramètres intrinsèques est résumée par l'équation (1.8) :

$$\begin{bmatrix} s \\ t \\ U \\ V \\ 1 \end{bmatrix} = \begin{bmatrix} h_{1,1} & 0 & h_{1,3} & 0 & h_{1,5} \\ 0 & h_{2,2} & 0 & h_{2,4} & h_{2,5} \\ h_{3,1} & 0 & h_{3,3} & 0 & h_{3,5} \\ 0 & h_{4,2} & 0 & h_{4,4} & h_{4,5} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} i \\ j \\ k \\ l \\ 1 \end{bmatrix} \quad (1.8)$$

Avec (s,t) positions du rayon, (U,V) : directions du rayon, (i,j) : coordonnées du pixel et (k,l) : index de la caméra (microlentille).

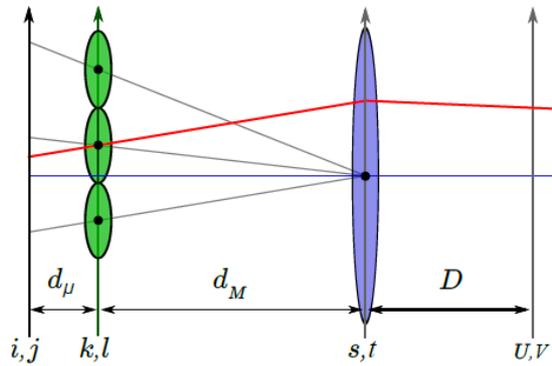


Figure 1.25 Représentation de la caméra plénoptique. Dans ce modèle, la lentille principale est représentée comme une lentille mince et les microlentilles comme un réseau de *pinholes*. Les rayons tracés en gris représentent les centres des images des microlentilles [Dans14]. La distance microlentilles – capteur est d_μ , la distance lentille principale – microlentilles est d_M .

Concernant les caméras plénoptiques 2.0, on recense plusieurs approches de calibrage. Johannsen [JHGP13] a proposé un modèle à 15 paramètres intrinsèques pour la version focalisée de caméras plénoptiques. La détection des centres optiques pour la lentille principale et les microlentilles n'est pas nécessaire. Par contre, les distorsions radiales et latérales sont corrigées et cette méthode permet une analyse métrique de la scène.

Une seconde approche modélise la caméra plénoptique sous forme de n caméras *pinholes* identiques translatées les unes par rapport aux autres [ZeQS14]. Avec ce modèle, seuls 4 paramètres intrinsèques sont nécessaires et la distorsion est prise en compte. Néanmoins, cette approche semble inadaptée pour des focales longues de la lentille principale.

Un modèle à 10 paramètres intrinsèques, prenant en compte le désalignement des microlentilles, a été proposé par Zhang [ZhJW16]. Ces paramètres sont la distance entre la lentille principale et les microlentilles L , celle entre les microlentilles et le capteur image l , le désalignement des microlentilles $x_m, y_m, (\theta, \beta, \gamma)$, la focale de la lentille principale F et le décalage des coordonnées image (u_0, v_0) . Ce calibrage estime ces paramètres à partir des images brutes ce qui garantit une solution unique. Cette méthode présente une bonne stabilité et une faible complexité algorithmique. Les résultats obtenus présentent une erreur inférieure à 0,6 pixel.

Heinze et al. [HSHP15] ont proposé une méthode robuste automatique de calibrage des caméras plénoptiques multi focus. Elle introduit un modèle plus précis des microlentilles permettant d'effectuer des mesures 3D plus précises et de réduire le bruit sur les cartes de disparité. Les aberrations des microlentilles sont corrigées.

Un autre modèle de caméra plénoptique focalisée considère la distorsion latérale de l'intensité de l'image comme une distorsion virtuelle de la profondeur [ZeQS16]. L'utilisation d'une mire 3D pour trouver les paramètres d'initialisation du modèle est rendue nécessaire. Finalement, elle est inadaptée aux objets situés à moins d'1 m de la caméra.

Une équipe s'est intéressée au mauvais alignement des microlentilles [ThTF14] et a mis en place une méthode pour calculer la position et l'orientation de la grille de microlentilles.

Une méthode a été définie pour exploiter les propriétés uniques des matrices de la fonction de la pupille image ou PIF (*Pupil Image Function*) caractérisant les caméras plénoptiques [PaTB14].

Hall et al. [HFTG17] se sont intéressés au calibrage volumétrique des caméras plénoptiques basé sur une fonction de cartographie polynomiale.

L'utilisation d'une caméra plénoptique est soumise à la détermination de l'ensemble des paramètres intrinsèques, dont le nombre dépend de la méthode employée. Cette phase de calibrage est difficile à mettre en œuvre pour certaines applications. Nous verrons que notre calibrage avec nos prototypes est plus aisé à mettre en œuvre. De plus, il nécessite de calculer moins de paramètres que les autres méthodes.

1.5. Domaines d'applications des systèmes *light-field*

Les applications et les domaines d'application des systèmes *light-field* sont nombreux. Cette partie présente deux cadres qui s'inscrivent directement dans le champ d'application de ces travaux de thèse : la refocalisation numérique et le calcul de cartes de profondeur. D'autres secteurs expérimentaux sont cités afin de mettre en perspective d'autres applications potentielles.

1.5.1. Refocalisation numérique

Comme le *light-field* 4D de la scène est capturé numériquement par la caméra *light-field*, il peut permettre de reconstruire une image dans un plan synthétique par une projection 2D. Cette opération est connue comme la refocalisation numérique. En d'autres termes, avoir le *light-field* 4D permet de simuler numériquement la refocalisation optique. Cependant, cette méthode est parfois confondue avec d'autres approches basées sur le traitement du signal comme l'application Android® *Lens Blur* fournie par *Google Camera* [16]. Celle-ci floute de manière sélective le contenu de l'image en fonction de l'information de profondeur pour produire l'effet Bokeh ou faible effet de profondeur. La connaissance de la profondeur est le plus souvent estimée à partir d'une séquence d'images focalisées [JaBL12, SuHS15] ou des images multi-vues [BASH15] d'une scène statique. Par contre, la refocalisation numérique *light-field* requiert seulement une unique acquisition avec une seule caméra.

Dans la photographie conventionnelle, le focus de l'appareil photo doit être fixé avant que la photo ne soit prise afin de s'assurer que les éléments choisis dans la scène soient dans le plan de focalisation. La refocalisation numérique peut être attrayante pour les photographes amateurs. Lors de la capture d'un moment sans ajustement précis de la focale, l'objet d'intérêt peut être flou sur la photo ce qui est un problème. Ng et al. [NLBD05] ont démontré pour la première fois la refocalisation numérique utilisant les données *light-field* 4D capturées à partir d'une caméra *light-field* basée sur un réseau de microlentilles. Ces travaux ont finalement conduit à un produit commercial, la caméra *Lytro* pour laquelle le principal intérêt est la refocalisation intégrée. Comme montré sur la Figure 1.26, les images 2D reconstruites à partir de la caméra *Lytro* sont focalisées numériquement sur une région d'intérêt.

Avec une caméra *light-field*, la refocalisation devient une étape de post-traitement et un processus interactif de divertissement.



Figure 1.26 A gauche, refocalisation numérique sur un objet proche ; au milieu, refocalisation numérique sur un objet lointain ; à droite, une image synthétique *all-in-focus* [Shan16].

En utilisant plusieurs lentilles ou un réseau de microlentilles, les données brutes capturées avec des caméras plénoptiques contiennent des informations sur la parallaxe et la focalisation. Même si les données brutes ne fournissent pas une image compréhensible, les informations capturées permettent de calculer des images numériques refocalisées [NLBD05]. Dans le domaine de la photographie computationnelle, la refocalisation numérique est une application répandue et il existe plusieurs méthodes de calcul des images refocalisées. Deux principaux types d'images refocalisées sont calculés, soient des images refocalisées en un point donné, soient des images refocalisées sur tout le champ de l'image appelées aussi *full refocused image* ou image totalement refocalisée.

En fonction des dispositifs *light-field*, des approches et des algorithmes différents ont été proposés pour calculer des images refocalisées à une profondeur. Pour les caméras plénoptiques version 1.0 (comme les caméras *Lytro* ou *Cafadis*), les principes consistent à calculer des vues obtenues sur un plan virtuel, se déplaçant parallèlement au plan du capteur image. Deux principaux concepts ont été proposés : la somme de rayons lumineux et l'utilisation de la transformée de Fourier 4D. La première approche, présentée par R. Ng [Ng06], consiste à utiliser la représentation vectorielle 4D des rayons lumineux capturés par chaque pixel. Grâce à des opérateurs de décalage et de dilatation, le cône de rayons passant à travers la lentille principale peut être sommé sur des plans virtuels parallèles. Quand le cône est bien sommé sur le même pixel de l'image virtuelle, l'objet correspondant apparaît ainsi bien focalisé sur cette image. A cause de ce dispositif, des effets d'aliasing peuvent être réduits en sur-échantillonnant, en interpolant les sous-images localisées sous chaque microlentille. Une autre méthode est proposée par Fiss [FiCS14] pour équilibrer les effets d'aliasing. Elle correspond à un algorithme de *plane sweep* utilisant un noyau *splat* fixé appliqué à chaque rayon. Le noyau *splat* doit être

assez large pour rendre invisible les artéfacts d'*aliasing* dans l'image refocalisée. La seconde approche initiée par R. Ng en 2005 est nommée algorithme de *Fourier Slice refocusing* (refocalisation par coupe dans Fourier). Cette méthode consiste à précalculer la transformée de Fourier du *light-field*. Ensuite, le plan focalisé désiré doit être sélectionné dans le domaine de Fourier en extrayant la coupe dilatée de Fourier (*dilated Fourier slice*) de la transformée de Fourier pré calculée. La coupe dépend d'un facteur lié à la profondeur. Enfin, l'image refocalisée est obtenue en calculant la transformée de Fourier inverse. La convolution effectuée dans l'espace de Fourier génère de l'*aliasing* dans les images résultats, pouvant être atténué en filtrant et en sur-échantillonnant. Comparé à la première approche, la qualité de l'image est bien meilleure et le calcul plus rapide. Dans la continuité de ces travaux, Danseureau [DaBr03, DaPW15] a montré que le mouvement de parallaxe (*motion parallax*) a des conséquences dans le domaine de Fourier et a proposé d'appliquer des filtres 4D *hyperfans* afin de sélectionner des zones de focalisation.

Pour les caméras plénoptiques version 2.0, aussi nommée *focused plenoptic* (comme les caméras *Ray-trix*), une image brute est constituée par un ensemble de micro-images focalisées correspondant à chaque microlentille. En fonction de la distance de travail et des caractéristiques du dispositif, les objets peuvent être totalement ou partiellement dupliqués au voisinage des micro-images. Les principes de la refocalisation des images a été décrit par Georgiev [GeLu10], et consistent en une moyenne (« un mélange ») du même point spatial à travers les différentes images des microlentilles. Le découpage et l'assemblage de patches extraits à partir de ces micro images permet de l'obtenir. La taille de ces patches dépend de la distance focalisée désirée. Liu [LiZL15] a proposé une extension de cette méthode pour calculer des images haute résolution. Dans cette méthode, la taille des patches dépend d'un facteur de focalisation et d'une taille spécifique d'ouverture. Le facteur d'ouverture fournit la capacité de contrôler la profondeur de champ de l'image refocalisée.

Dans le cas des caméras à ouverture codée, le vecteur *light-field* 4D peut être retrouvé en réarrangeant les transformées de Fourier 2D des images acquises pour chaque masque [VRAM07]. Dans ce cas, les algorithmes de refocalisation calibrée sont similaires à ceux utilisés pour les caméras plénoptiques version 1.0. Quelques variantes existent, par exemple Reddy [ReBR13] a présenté un dispositif à ouverture codée composé d'un masque amovible externe devant une seule lentille principale. Afin de calculer des images refocalisées à différentes profondeurs, le *light-field* est déformé (*warp*) vers une position virtuelle et ensuite intégré sur fenêtre synthétique d'ouverture.

Dans le cas des multi-caméras ou des caméras multi-vues, les méthodes exploitent la parallaxe existante entre les différents points de vue. Ceux-ci sont projetés dans le plan d'une image de référence, en utilisant un ensemble d'homographies. Les images refocalisées sont calculées en moyennant les images projetées, et elles sont ensuite focalisées dans le plan de référence. Dans leurs travaux, Vaish et al. [VGTA07] définissent aussi la transformation homologique pour passer du plan de référence à n'importe quel plan de la scène, généralement parallèle au plan du capteur image. Dans ce cas, les images peuvent être focalisées sur n'importe quel plan de la scène.

Le rendu final des effets de focalisation ou de défocalisation de ces méthodes est différent. Quelle que soit la technique utilisée, la plupart des méthodes présentent des résultats pour lesquels les images sont focalisées sur un plan proche, au milieu ou un plan lointain, sans aucune référence métrique.

Pour les mêmes paramètres optiques, comme la focale et la taille de l'ouverture, Ng et al. [NLBD05] ont démontré expérimentalement que la caméra *light-field* peut fournir une plus large profondeur de champ (*Depth Of Field* DOF) que celle obtenue avec une caméra conventionnelle alors que les deux types de caméras ont le même taux d'échantillonnage spatial. L'extension de la profondeur de champ est obtenue par manipulation des images refocalisées obtenues par projection numérique du *light-field* 4D en images 2D dont les paramètres de refocalisation varient. Pour synthétiser artificiellement une image avec une profondeur de champ étendue, un algorithme *graph-cut* [ADAD04] a été utilisé sélectionnant de manière adaptative les objets focalisés à partir des images refocalisées et les fusionnant en une image all-in-focus. Pour quelques applications de vision pour lesquelles les objets sont répartis sur une large gamme de profondeur, la caméra a besoin d'un large champ de focalisation. Réduire l'ouverture de la caméra n'est pas possible quand la lumière ambiante est faible.

Une extension possible de la refocalisation est le changement de point de vue : si la refocalisation est pensée comme le déplacement de l'observateur dans les deux sens suivant l'axe optique, alors il est possible de déplacer l'observateur latéralement en projetant le *light-field* d'une façon différente. Cet effet est nommé effet de parallaxe (*parallax effect*) [WaGo14]. Ce type d'images peut être calculé avec les caméras plénoptiques 1.0 et 2.0.

Les images refocalisées sont rarement calibrées en profondeur. En 2016, Hahne et al. ont proposé une solution pour le faire à partir d'images acquises avec une caméra plénoptique standard [HAVF16]. Cette méthode, basée sur l'utilisation du tracé de rayons et des caractéristiques géométriques du dispositif, permet de prévoir précisément la distance à laquelle est refocalisée l'image calculée. Cependant, ces images sont uniquement exploitées dans le cadre de la photographie pour des aspects esthétiques et artistiques.

1.5.2. Calcul de carte de profondeur

Déduire la profondeur des objets à partir d'images a été un champ d'investigation très actif dans le domaine de la vision par ordinateur. Le calcul d'une carte de profondeur revient à attribuer à chaque pixel d'une vue de la scène une valeur correspondant à la profondeur de la surface des objets dans la scène. La première caméra portable *light-field* a été développée pour détecter la profondeur et est connue comme une caméra stéréo avec une seule lentille [AdWa92]. Zhou et Nayar [ZhNa11] ont analysé théoriquement le principe de la détection de profondeur avec une caméra *light-field* et ont conclu que le résultat estimé devrait être plus fiable que celui obtenu avec une caméra stéréo. Les trois principales méthodologies suivantes ont été envisagées : l'utilisation des lignes épipolaires (EPI), l'information de refocalisation et les approches stéréo multi-vues.

Comme la refocalisation et l'information de champ lumineux sont fortement corrélées avec la profondeur des objets capturés par la caméra, une carte de disparité ou une image de profondeur peut être calculées en utilisant les paramètres de refocalisation [SSDS14].

Plusieurs variantes des méthodes basées sur les lignes épipolaires existent. Une première utilise le phénomène de parallaxe en calculant des images « sous-ouverture » [BiFa11, WaFJ11]. Ces images de « sous-ouverture » sont utilisées pour générer des « images épipolaires » (EPI) permettant d'estimer la profondeur en analysant la pente des lignes apparues dans les EPI [KZPS13, WaFJ11]. Une autre méthode nécessitant la définition d'un espace à échelle de profondeur a été présentée par Tomic et al. [ToBe14]. Elle propose de détecter les rayons dans les EPI en trouvant les extrema. Les occultations sont traitées en analysant le recouvrement des rayons détectés. Elle fournit une carte de profondeur grossière pour les régions uniformes, et une fine pour les régions texturées. Plus récemment, Shan et al. [Shan16] ont produit deux cartes de disparité à partir des lignes épipolaires : la première pour l'axe horizontal (u) et la seconde pour l'axe vertical (v). Ces deux cartes sont fusionnées et affinées afin de générer la carte de disparité finale. Un exemple d'images épipolaires est illustré par la Figure 1.27. En regardant l'image épipolaire du haut (pour l'axe vertical), on constate que la ligne épipolaire au niveau du buisson présente une forte pente ce qui signifie que le buisson est loin de la caméra. La ligne épipolaire au niveau de l'œil du perroquet est verticale ce qui indique que le perroquet est proche de la caméra.

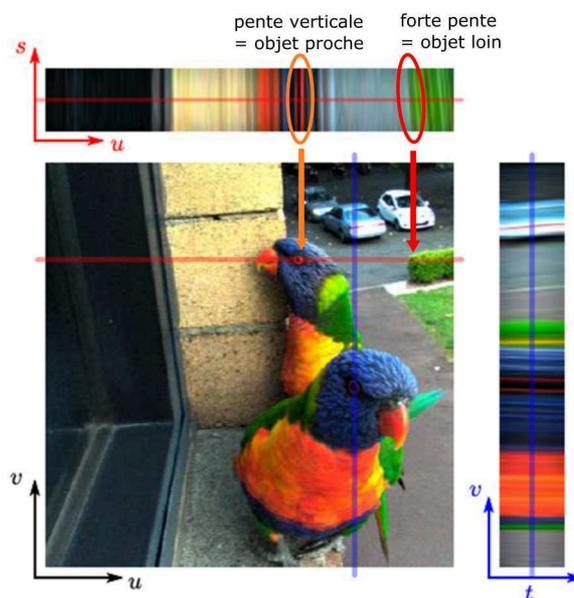


Figure 1.27 Au centre : un exemple de sous-image et les deux images associées (en haut et à droite) représentent des images épipolaires de cette sous-image [Dans14].

Tao et al. [THMR13] ont combiné deux cartes de profondeur générées à partir de la défocalisation et de la mise en correspondance. La première est obtenue en calculant la variance horizontale (spatiale) après l'intégration verticale (angulaire) des lignes épipolaires et la seconde l'est en calculant la variance verticale (angulaire). Le calcul de carte de profondeur nécessite une unique acquisition. Une optimisation a été proposée par Kim et al. [KiOK14]. La différence réside dans le poids accordé à chaque pixel

lors de la combinaison. La première utilise directement des poids dits de confiance tandis que la seconde définit des pixels « ancrés » pour lesquels la profondeur associée est fiable en exploitant leurs propriétés en termes de correspondances et de flou. Les autres pixels avec une profondeur non fiable sont corrigés en supposant que les pixels voisins de même couleur ont la même profondeur. Plus tard, Wanner et al. [WaGo14] ont proposé d'analyser l'orientation des lignes épipolaires 2D : la pente de chacune est proportionnelle à la disparité et donc à la profondeur (cf. Figure 1.27). Le résultat final est optimisé par filtrage variant.

Une méthode alternative consiste à travailler directement avec les données brutes et à calculer les images de profondeur en utilisant des méthodes de triangulation [USKK13]. Les concepts développés sont similaires à ceux connus en stéréovision avec deux caméras parallèles.

Ces trois types d'approches restent vrais pour de nombreux prototypes existants de caméras plénoptiques et pour des modèles commerciaux proposés par les compagnies *Lytro* et *Raytrix*.

D'autres méthodes utilisent les images de sous-ouverture sans calculer les EPI. Une méthode est basée sur la mise en correspondance stéréo des images de sous-ouverture. Pour les déplacer, elle emploie le théorème du retard de la transformée de Fourier. Ce décalage, de précision sous-pixellique, doit être efficace dans le domaine de Fourier. Le calcul de la carte de disparité dense avec une précision sous-pixellique s'effectue de manière itérative en utilisant les mises en correspondance comme contraintes du modèle d'optimisation pour corriger la carte de profondeur. Cette dernière est raffinée à l'aide d'une fonction quadratique locale [JPCP15]. Une autre propose un modèle variationnel nécessitant la mesure de la rotation de points caractéristiques entre certaines images de sous-ouverture. A partir de cette mesure, il est possible d'accéder à l'information de profondeur et après une phase d'optimisation, la carte de profondeur optimale est générée [HeRP13].

Bien que l'estimation de la profondeur représente un problème d'optimisation multidimensionnelle, un compromis entre la complexité et, la précision ou la cohérence, doit toujours être fait. Par exemple, un marquage de la profondeur globalement cohérent, exécuté directement sur la représentation *light-field* 4D, permet d'obtenir des cartes de profondeur denses [WaGo12, WaGo14]. En outre, l'estimation de la profondeur est améliorée et la complexité est réduite en utilisant une méthode basée sur les structures géométriques comme les segments linéaires 3D [YGLL13]. D'autres méthodes réduisent la complexité en utilisant des contraintes locales plutôt que globales et produisent une carte de profondeur éparsée. Ces cartes fonctionnent uniquement pour les régions texturées [BiFa11, PeWi12].

Les problèmes d'occultation pour le calcul de carte de profondeur constituent un pan de recherche à eux seuls. Yu et al. [YYLG13] ont proposé une méthode itérative pour les prendre en compte. Elle est basée sur le calcul de la variance. Si elle est minimale, la profondeur associée correspond à sa valeur optimale. Can et al. [CLYB14] ont aussi démontré qu'en cas d'occultations importantes, la profondeur véritable d'un pixel correspond à un minimum de la courbe d'uniformité en fonction de la profondeur. Plus récemment, Johannsen et al. [JoSG16] ont pris en compte le fait qu'une certaine combinaison des informations verticales et horizontales des EPI permet d'estimer correctement la disparité aux niveaux des occultations.

1.5.3. Autres exemples d'applications

1.5.3.1. Microscopie

Levoy et al. [LNAF06] ont été les premiers à proposer un système de microscopie *light-field* en insérant un réseau de microlentilles sur le chemin optique d'un microscope. Cela a permis de surmonter les limites d'un microscope conventionnel (vue limitée à la vue orthographique, champ de vue peu profond). Dans leur prototype, une lentille relais est employée pour faire correspondre l'ouverture numérique de la lentille objectif et celle du réseau de microlentilles. Leur prototype est capable de produire une pile d'images avec différents plans focaux et différentes perspectives d'échantillons biologiques. Pour les échantillons semi-transparents, ils ont aussi réalisé la reconstruction en volume 3D par une déconvolution 3D avec une PSF. En introduisant une autre microlentille et un miroir digital (DMD), Levoy et al. [LeZM09] ont développé un microscope 4D *light-field* avec illumination contrôlée. Ils ont aussi démontré expérimentalement la correction modérée des aberrations par le contrôle de l'illumination *light-field* 4D. Au lieu d'utiliser l'optique géométrique, Broxton et al. [BGYC13] ont proposé un modèle de transport *light-field* basé sur l'optique ondulatoire. Les hypothèses sont que la lumière est émise isotropiquement à partir de tous les fluorophores dans le volume et qu'elle se déplace directement vers l'objectif du microscope sans diffusion ou réfraction. En utilisant l'approximation numérique du modèle proposé, le volume d'un échantillon biologique fluorescent est reconstruit par une déconvolution 3D améliorée. Dans leur expérimentation, pour avoir un taux d'échantillonnage angulaire suffisant, il y a 27x 27 pixels derrière chaque microlentille. Ils ont démontré que la reconstruction volumétrique d'un grain de pollen pour lequel la résolution spatiale est améliorée par un facteur de 2 à 4 dépendant de la distance à laquelle le plan de reconstruction est décalé par rapport au plan de focalisation suivant l'axe optique. Pour atténuer la non uniformité de la résolution latérale suivant l'axe optique, Cohen et al. [CYAB14] ont proposé d'utiliser le front d'onde codé obtenu en plaçant des masques de phase dans le chemin optique du microscope *light-field*. Ils ont montré qu'il est possible d'obtenir une MTF plus plate suivant l'axe optique que celle obtenue dans la référence [BGYC13]. Pour capturer le *light-field* haute-résolution, Lu et al. [LuMF13] ont proposé d'utiliser un système hybride constitué d'un microscope *light-field* et d'un microscope conventionnel avec séparateur de faisceau. L'inconvénient de ce système est qu'il est plus compliqué et que l'efficacité lumineuse est réduite de moitié.

1.5.3.2. Détection de front d'onde

La détection de fronts d'onde (*wavefront sensing*) est une méthode permettant de mesurer les aberrations d'un front d'onde lumineuse. Elle est communément employée en optique adaptative [Tyso10] appliquée à l'astronomie. Ces dernières années, elle a été étendue aux domaines de la microscopie [Boot07] et de l'ophtalmologie [MSCS14].

Actuellement, les caméras plénoptiques composées d'une matrice de microlentilles sont capables de mesurer la forme du front d'onde au plan de la pupille de sortie de la lentille principale. Rodriguez-

Ramos et al. [RCNF08] et Clare et Lane [CILa04] ont démontré que les *light-field* 4D peuvent être converties en front d'onde. Cependant, les considérations en termes de design pour les capteurs de front d'onde et les caméras plénoptiques avec une matrice de microlentilles sont mutuellement opposées [CILa04]. Les coordonnées angulaires et spatiales sont échangées entre les capteurs de front d'onde et les caméras plénoptiques (plénoptique 1.0). C'est pour cette raison que les résultats de détection de front d'onde [CILa04, RCNF08] obtenus avec une caméra plénoptique ont une faible résolution spatiale sur le front d'onde.

1.5.3.3. Correction des aberrations optiques

L'utilisation de la technologie *light-field* pour supprimer les aberrations des lentilles a été proposée par Hanrahan et Ng [HaNg06]. L'idée est de corriger numériquement la terminaison des rayons afin de corriger les aberrations. Au lieu d'intégrer suivant l'axe vertical pour obtenir l'intensité de l'image dans le *light-field* 4D, l'intensité de l'image sans aberration peut être obtenue en intégrant suivant une courbe qui peut être déterminée à partir du tracé de rayons basé sur un modèle de système de lentille spécifique. Cependant, cela requiert un capteur *light-field* avec un taux d'échantillonnage angulaire suffisant. D'autre part, la résolution spatiale du capteur doit être supérieure à un certain niveau pour que l'aberration optique permette de limiter la résolution spatiale. Pour prendre en compte les deux critères, la résolution du capteur image doit être plus large que celles des capteurs employés actuellement pour les plus grandes caméras professionnelles. En considérant la technologie actuelle des capteurs et leur coût, l'utilisation de la déconvolution avec les bons paramètres et la mesure de la PSF [HRHL13] pour réduire les aberrations optiques est une meilleure solution que l'approche basée sur le retracé de rayons *light-field*.

1.5.3.4. Reconnaissance d'objet

La reconnaissance d'objets est un sujet bien exploré dans le domaine de la vision par ordinateur. Cependant, les descripteurs actuels tels que les algorithmes SIFT (scale-invariant feature transform = transformation de caractéristiques visuelles invariante à l'échelle) [Lowe99] ou SURF (Speeded Up Robust Features = caractéristiques robustes accélérées) [BETV08] ne fonctionnent pas avec les objets transparents. Pour résoudre ce problème, Maeno et al. [MNST13] ont développé un descripteur *light-field* basé sur le fait que les objets transparents ont la caractéristique unique de déformer le fond ou les objets présents derrière eux (phénomène causé par la réfraction). Selon leur expérimentation, le taux d'identification a été amélioré de manière significative pour les objets transparents.

1.5.3.5. Vélocimétrie d'image de particule

La vélocimétrie d'image de particule (VIP) est une méthode d'imagerie qui visualise l'écoulement utilisée principalement pour des expériences de dynamique des fluides. Le fluide est enrichi de petites

particules qui peuvent être illuminées par une source. Le mouvement des particules est supposé représenter le champ de vitesses locales 3D. Belden et al. [BTAT10] ont démontré qu'un réseau de 3x3 caméras peut résoudre des champs vectoriels 3D pour des fluides abondamment enrichis en utilisant l'ouverture synthétique. Plus tard, Lynch et al. [LyFT12] et Garbe et al. [GaVS12] ont proposé d'utiliser des caméras *light-field* composées d'un réseau de microlentilles avec des configurations optiques différentes pour mesurer la vitesse en trois dimensions à trois composants (3D3C).

1.5.3.6. Analyse de *light-field* (*probe light-field*)

Wetzstein et al. [WeRH11] ont utilisé un système composé d'une caméra et d'un dispositif spécifique d'illumination pour visualiser optiquement le changement de réfraction dans un objet transparent grâce à une technique connue comme la photographie Schlieren. Dans leur expérimentation, le composant clé est le capteur *light-field*, constitué d'une source lumineuse uniforme, un film transparent imprimé avec un masque spécifique et un réseau de microlentilles. En appliquant différents types de masque sur le film transparent, le dispositif lumineux est capable de produire différentes formes de *light-field* avec aussi bien des variations colorimétriques que d'intensité. L'idée est d'étendre la reconstruction à la surface d'objets transparents [WRHR11]. Ils ont proposé d'utiliser les gradients couleurs et les contraintes géométriques pour trouver les correspondances précises rayon à rayon, ainsi que de calculer la normale de la surface avec la loi de Snell-Descartes.

1.5.3.7. Odométrie visuelle

L'odométrie visuelle est une méthode estimant la trajectoire d'un mobile (robot, etc.) à partir d'images captées en continu. Cette estimation est effectuée en utilisant un ensemble de relevés de points relatifs. Les recherches concernant ce domaine ont commencé dans les années 80. Récemment, une caméra plénoptique 1.0 a été utilisée par Dansereau et al. [DMPW11] dans ce domaine. Ils ont proposé trois solutions capables d'estimer la trajectoire à 6 degrés de liberté : une approche modulaire qui peut être adaptée aux applications spécifiques, une solution plus intégrée basée sur une généralisation point par point du flux plénoptique et une intégration totale de l'équation plénoptique fournissant une solution à une étape. Toutes les méthodes opèrent sans extraire d'éléments distinctifs, efficacement et de manière constante dans le temps indépendamment de la complexité de la scène. Les résultats obtenus avec les deux dernières méthodes sont confirmés par la vérité terrain. Une caméra plénoptique focalisée permet d'effectuer de l'odométrie visuelle de plusieurs manières [ZeQS15, ZeQS16]. La première méthode fusionne les données de profondeur obtenues à partir d'un algorithme SLAM monoculaire et celles reçues à partir d'une caméra plénoptique focalisée. Elle est adaptée aux caméras avec un faible champ de vue. Pour les objets très distants, l'information de profondeur est tellement améliorée qu'elle peut être utilisée pour faire de la segmentation d'objets. La seconde méthode obtient une carte de profondeur avec une acquisition de la caméra plénoptique qui est fusionnée avec les données de profondeur déterminées en trouvant les correspondances avec les images synthétiques

totalemment focalisées. En outre, les capacités de suivi sont accrues avec ce système, grâce à une méthode de calcul inverse de la carte de profondeur. Les résultats obtenus sont plus exacts et plus fiables que ceux obtenus avec la profondeur estimée à partir d'une seule image *light-field*. L'inconvénient réside dans le temps (effort) de calcul. Avec ces deux méthodes, l'échelle de la scène peut être observée. En outre, le suivi métrique et la cartographie sont aussi réalisables.

1.5.3.8. Reconstruction 3D

Persini et al. [PeST00] se sont intéressés à la reconstruction dans le domaine culturel et historique, en utilisant les informations fournies par un système multi-caméras. La méthode de reconstruction est basée sur les correspondances stéréo et est indépendante de la géométrie du système d'acquisition. La reconstruction 3D avec l'utilisation d'un réseau de multi-caméras a aussi été étudiée par Koch et al. [KFEW02] mais appliquée à des scènes extérieures. L'approche est basée sur une méthode non calibrée de *structure from motion*. Kim et al. [KZPS13] ont utilisé leurs cartes de profondeur générées à partir des EPI pour calculer un maillage 3D, généré par la triangulation individuelle des cartes de disparité et en les moyennant dans un unique modèle. De plus, ils peuvent reconstruire une scène à partir des données d'entrée. Les résultats sont directement visualisables comme un nuage de points 3D colorés.

1.5.3.9. Superrésolution

Avec la configuration 1.0, il est possible de calculer des images résolues pour certaines distances de refocalisation. Le choix de la distance de refocalisation est primordial car dans ce cas, les objets (en pixel sur le détecteur) sont décalés d'une portion de pixels les uns par rapport aux autres. Pour obtenir une image mieux résolue, il faut recombinaison et entrelacer tous les pixels du détecteur. Ce type d'images peut aussi être calculé à partir des données capturées avec la caméra plénoptique 2.0. Le principe repose sur l'existence de plans intermédiaires de focalisation pour lesquels le décalage des objets entre chaque pixel équivaut à une portion non entière de pixel. A partir des images brutes, il est possible en combinant les pixels de calculer une image superrésolue [GeLu09]. Cependant, à cause du vignettage, cette application ne peut pas être utilisée sur les bords des microlentilles. Les deux méthodes présentées sont les plus communes. D'autres approches ont été envisagées pour chacune des configurations. Elles peuvent être basées notamment sur les 8 possibilités suivantes : la prise en compte du processus de formation des images incorporant des distributions Lambertiennes et de textures [BiFa12, BiZF09], un modèle fondé sur l'optique ondulatoire [SaKG16], une extension de la méthode Fourier Slice [PPRM12], la régularisation de graphes [RoFr17], une méthode utilisant la transformée en ondelettes discrète [RRYB13b], la solution d'un problème inverse [WaGo12a, WaGo14], un modèle gaussien (GMM) [MiVe12] ou l'utilisation d'un modèle préalable gaussien 3D [LeDu10].

1.6. Synthèse et positionnement des travaux

Différentes structures de caméras *light-field* ont été présentées dans ce chapitre, et nous résumons ces architectures dans le Tableau 1.5. D'un point de vue optique, il existe des équivalences entre les architectures de caméras *light-field*. Par exemple il est possible à partir de l'image brute fournie par une caméra plénoptique, de calculer des points de vue décalés d'une scène, similaires à ceux obtenus avec une caméra multi-vues ou un réseau de caméras. Ces équivalences ont été étudiées par Georgiev et al. [GeLG11], et plus récemment par Loïs Mignard, qui étudie dans le cadre de sa thèse l'unification des modèles de caméras *light-field* au travers de réseaux de caméras équivalents [MiRI17].

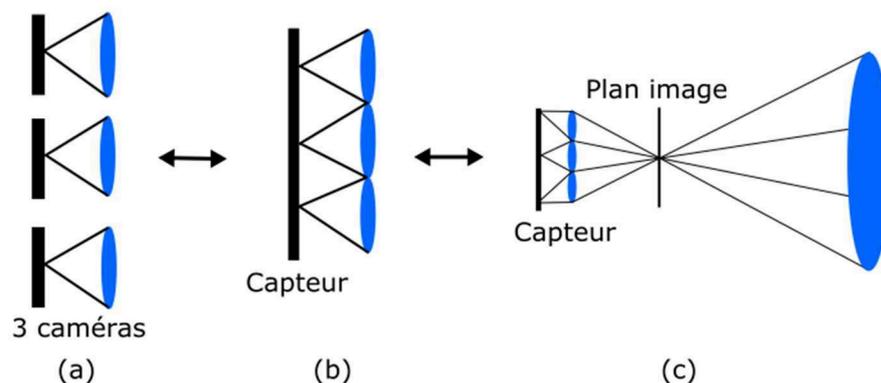


Figure 1.28 Trois schémas de dispositifs *light-field* (a) Réseaux de caméras ; (b) Caméra multi-vues ; (c) Caméra plénoptique

A partir des architectures décrites dans le Tableau 1.5, les questions qui ont animé nos travaux de thèse sont de 3 natures :

- Nous avons cherché à établir si un modèle de caméras à 4 vues apporte une alternative en termes de système de vision 3D. Une caméra à 4 vues c'est 6 couples de caméras stéréo, mais c'est aussi une caméra *light-field* avec peu de points de vue, comparée aux autres systèmes *light-field*.
- D'un côté, on trouve des caméras plénoptiques fournissant des imagerie à faibles angles d'ouverture, et de l'autre des réseaux de caméras (ou caméras multi-vues), où les images sont plutôt à grande ouverture : quel serait l'intérêt d'étudier une caméra plénoptique à grande ouverture ? ou à 4 vues ? Le dispositif présente-t-il un bénéfice dans le domaine de la vision ?
- Peut-on aborder l'imagerie 3D différemment, en travaillant avec des images refocalisées d'un dispositif à 4 vues ? Peut-on amener une plus-value dans le domaine de la vision industrielle en travaillant avec une caméra *light-field* ?

Les chapitres 2 et 3 répondent aux 2 premières questions tandis que le chapitre 4 propose quelques approches algorithmiques, utilisant les propriétés des images refocalisées pour la vision industrielle.

Approches	Intérêts	Limitations	Nb vues équivalentes
Ouverture codée	Possibilité d'utiliser un LCD programmable Absence de vignettage (en fonction de la configuration)	Acquisition séquentielle Vignettage (en fonction de la configuration)	9 vues (extrait des principales publications, montrant des jeux de 9 masques)
Plénoptique 1.0	Traitements d'image simples Bonne résolution en profondeur	Faible résolution spatiale Zones aveugles dues au vignettage	Matrice de 100 000 microlentilles (Lytro 1)
Plénoptique 2.0	Meilleur compromis entre résolution spatiale et résolution en profondeurs Pas de zones aveugles	Traitements d'images plus complexes Présence de vignettage complexifiant les traitements	Matrice de 79x79 microlentilles (caméra Raytrix)
Plénoptique 1.2	Bonne résolution spatiale Pas de problèmes de vignettage	Chromatisme Peu adapté à des systèmes à focale variable Peu adapté à des scènes proches	100 images à partir de 20 lentilles
Kaléidoscope	Pas de problème de vignettage Profondeur de champ ajustable	Encombrement proportionnel à la résolution en profondeur souhaitée	-
Multi caméras	Ecartement des caméras contrôle la parallaxe Adapté à des scènes proches ou lointaines	Encombrement Synchronisation des caméras nécessaire Sensibilité aux vibrations	Nb de vues = nb de caméras
TOMBO, multi vues	Compacité : meilleur candidat pour le marché du mobile	Faible résolution de chacune des vues avec le système à 16 vues de la caméra PiCam, nécessitant de la superrésolution Traitements encore trop lourds pour des processeurs mobiles	Matrice de 4x4 lentilles

Tableau 1.5 Synthèse des avantages et des inconvénients des différentes approches.

2 Géométrie des systèmes *light-field* et conception d'un système multi-vues plénoptique

2.1. Introduction

Au cours des travaux de thèse de Jun Xu [Xu12], l'emploi de caméras *light-field* pour effectuer des mesures 3D dans le domaine de la vision industrielle et plus particulièrement pour la détection de fibres émergentes a été envisagé. Au cours de cette thèse, un système à deux mini-lentilles a été développé au laboratoire. L'une des perspectives a été d'étendre le nombre de mini-lentilles pour un meilleur échantillonnage du *light-field*. Sur la base des conclusions de cette thèse, des essais ont été effectués avec une caméra plénoptique version 2.0 (Raytrix). Les résultats associés nous ont interrogés sur la puissance de calcul nécessaire à l'acquisition et aux traitements de ces données. L'alternative que nous proposons dans ces travaux concerne le développement de caméras multi-vues, dans un premier temps pour des applications de vision industrielle. Puis une extension à des grandes distances a été envisagée et mise en œuvre.

A partir de ces résultats, nous avons cherché s'il n'existe pas une solution, facilitant l'intégration d'un système multi-vues dans une caméra 2D. Nous proposons un principe reposant sur l'insertion d'une bague "optique" entre objectif et capteur image. Nous avons réalisé la modélisation optique du système, associé au développement d'outils de simulations et de caractérisation. La caractérisation doit aider à la co-conception géométrique, afin que le dispositif soit capable d'obtenir un jeu d'images où nos approches de refocalisation, que nous proposons dans ces travaux, puissent être appliquées.

Plusieurs méthodes de modélisation des caméras *light-field* ont été proposées dans la littérature. Perwass et Wietze [PeWi12] ont défini un modèle de Caméra Plénoptique Standard (CPS), dont certaines caractéristiques sont définies grâce aux triangles semblables. Plusieurs approches basées sur le tracé de rayons ont été proposées en 2016. Un premier modèle, développé par Shan [Shan16], propose un environnement *open-source* aidant à la conception de systèmes *light-field* et plus particulièrement les caméras plénoptiques standards. Une autre approche a pour but de modéliser le *light-field* des CPS en traçant suffisamment de rayons lumineux distincts. Les premiers résultats expérimentaux associés ont été publiés dans [DaPW13]. Ce modèle a été affiné par Hahne [Hahn16, HAHV14]. De plus, ce modèle permet aussi de définir exactement les plans objets qui peuvent être focalisés numériquement. Mignard et al. [MiRI17] ont récemment proposé de modéliser les caméras *light-field* comme un réseau équivalent de caméras. Cela consiste en l'abstraction de ce type de dispositif au moyen de réseaux de caméras virtuelles observant l'espace objet. Il permet de décrire les caractéristiques de la caméra, telles que la profondeur de champ ou l'angle de vue, applicables à la plupart des caméras *light-field*. Dans un premier temps, les raisons de passer de systèmes multi-caméras aux caméras plénoptiques sont expliquées. Dans une seconde partie, les différents dispositifs développés au laboratoire sont présentés : un dispositif multi-vues et un multi-caméras. Pour chacun d'eux, un cahier de charges est défini regroupant une gamme de distance de travail. La plupart des dispositifs ont été développés pour fonctionner dans le domaine de la vision industrielle, hormis le dispositif multi-caméras. L'étude et le

développement du dispositif multi-vues plénoptique, nécessite un outil de simulation plus complexe que nous avons mis en place ainsi que la définition de caractéristique. Des expérimentations sont menées pour confirmer notre simulation et notre caractérisation. Ce modèle multi-vues plénoptique constitue à la fois la dernière partie étudiée de ce chapitre et de cette thèse.

2.2. Des systèmes multi-caméras aux caméras plénoptiques

Les travaux de Georgiev [GeLG11] ont montré l'équivalence des modèles optiques des dispositifs *light-field*. Cette équivalence est facile à établir entre un réseau de caméras (caméra polydioptrique) (Figure 2.1 (a)) et une caméra multi-vues (Figure 2.1 (b)). Dans ce cas de figure, c'est la baseline entre les centres optiques qui sera déterminante pour fixer l'étendue des mesures en Z. La résolution en Z du système est fonction des résolutions des capteurs images mis en jeu. Entre la caméra multi-vues et la caméra plénoptique 2.0, on peut considérer que les deux systèmes font l'acquisition de plusieurs points de vue d'une scène. Dans le premier cas avec un large recouvrement entre les sous-images et dans le second cas, avec peu de recouvrement et des angles de vues très faibles (Figure 2.2).

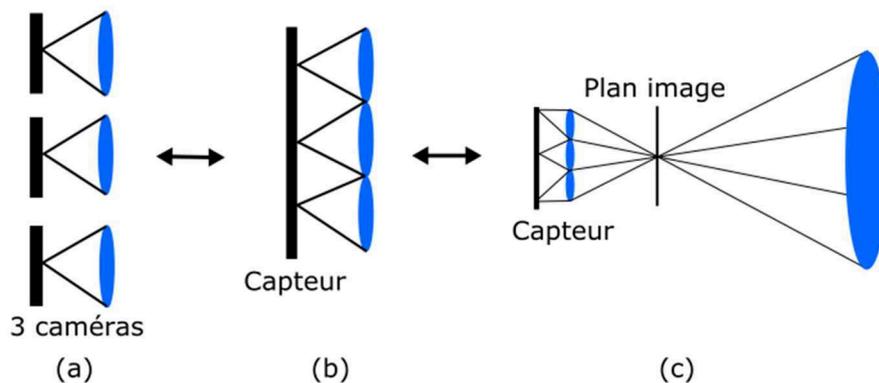


Figure 2.1 Trois schémas de dispositifs *light-field* (a) Réseaux de caméras ; (b) Caméra multi-vues ; (c) Caméra plénoptique.

Si les capacités 3D des caméras plénoptiques sont très prometteuses, l'expérience du laboratoire, montre que la technologie reste encore difficile à mettre en œuvre dans un certain nombre de domaines, comme par exemple la vision industrielle. D'une part, la méthode de calibrage demande un soin particulier, et d'autre part, la puissance de calculs reste un frein au déploiement de cette technologie pour le moment. Dans ce travail de thèse, le chapitre 4 s'attache à démontrer que l'on peut adopter des approches algorithmiques, basées sur la refocalisation numérique avec un système de caméra multi-vues ou avec un réseau de caméras. D'un point de vue optique, nous nous sommes posé la question de savoir s'il était possible de disposer d'un dispositif de vision plénoptique, ayant des caractéristiques proches d'un système multi-vues, c'est-à-dire fournissant un jeu de quatre images, avec un taux de recouvrement important, afin d'appliquer nos algorithmes de refocalisation numérique, comme décrit au chapitre 4. Nous utilisons ici le terme plénoptique, car dans ce cas de figure, nous souhaitons insérer un jeu de mini lentilles entre l'objectif principal et le capteur. La Figure 2.3

montre une grille de microlentilles collée à quelques centaines de micromètres d'un capteur image. L'assemblage doit être précis et reste ensuite irréversible.

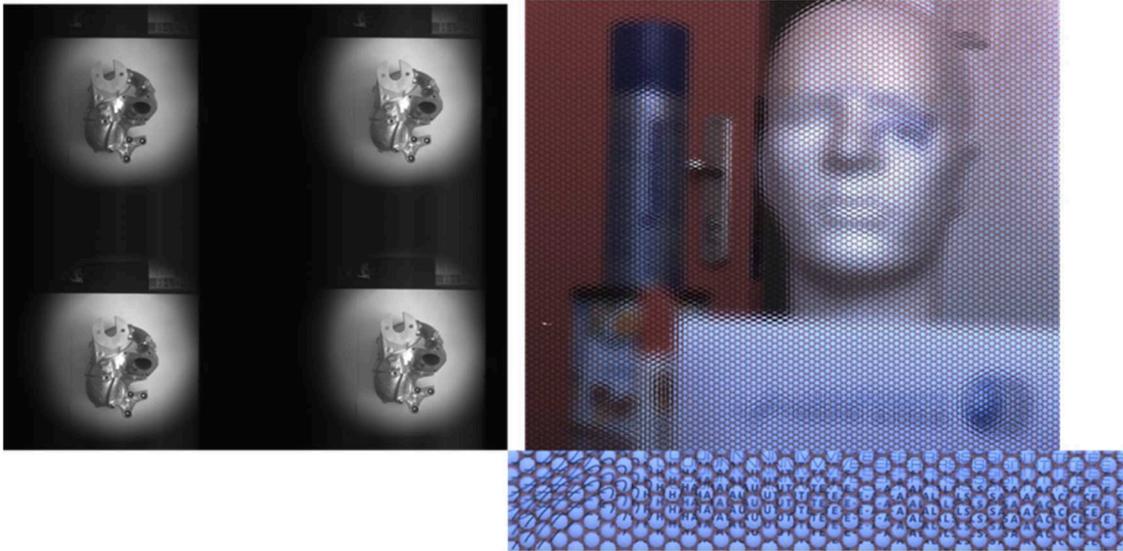


Figure 2.2 Image d'une caméra multi-vues et d'une caméra Raytrix (focused plenoptic).

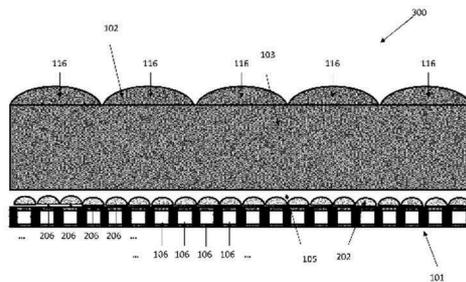


Figure 2.3 Grille de microlentilles collée à quelques centaines de micromètres d'un capteur image.

L'idée principale consiste à étudier s'il est possible d'établir un schéma de transformation d'une caméra 2D classique en une caméra à quatre vues, c'est-à-dire un système de vision 3D, qui puisse être réversible, par l'ajout ou le retrait d'une bague, comme cela est illustré par la Figure 2.4.



Figure 2.4 A gauche, illustration d'une caméra conventionnelle et à droite, illustration de notre prototype multi-vues plénoptique composé d'une caméra conventionnelle couplée à une bague de mini-lentilles.

Ainsi, dans la section suivante, la conception des différents systèmes répondant à ces attentes est présentée.

2.3. Prototypes de systèmes multi-vues et multi-caméras

Afin de maîtriser les paramètres associés aux images acquises, des prototypes multi-vues et multi-caméras ont été créés. Leur conception est présentée dans cette section.

Le premier dispositif présenté est composé d'un capteur et de quatre mini-lentilles, appelé système multi-vues. Sa gamme de travail est comprise entre quelques centimètres et 2 m. Il est développé pour fonctionner dans le cadre de la vision industrielle. Par contre, le second prototype a été mis en œuvre pour fonctionner dans le cadre de la perception automobile. La gamme de travail est donc comprise entre 2 m et plus. Pour répondre à cette problématique, il avait été envisagé de développer un autre dispositif multi-vues mais les études préliminaires nous ont amenés à concevoir un système multi-caméras.

2.3.1. Système multi-vues

Au cours de la thèse de Jun Xu [XCKF12, XCKF13, Xu12], un dispositif à deux lentilles a été développé pour la mesure 3D de défauts textiles. Ces travaux ont utilisé l'homographie variable, proposée par Zhang et Greenspan [ZhGr07], afin d'effectuer des mesures de parallaxe. A partir de ces travaux, il a été envisagé de proposer un nouveau dispositif multi-vues afin de réduire les problèmes d'occultations associées à un système à deux lentilles.

Dans un premier temps, une solution à quatre vues a été envisagée. Du point de vue du développement, nous avons décidé d'utiliser le même capteur que celui employé pour le dispositif à deux vues issu de [Xu12], puisque les images des deux lentilles utilisées couvrent la moitié du capteur. Cette solution à quatre lentilles permet d'utiliser une grande partie de la surface du capteur.

Pour prévoir les résultats escomptés de la mesure de profondeur en fonction de la disparité, l'équation suivante est employée, résultant de la triangulation :

$$Z = \frac{bf}{d} \quad (2.1)$$

Avec Z : la profondeur calculée, b : la distance entre les centres optiques des deux objectifs, d : la disparité et f : la focale des objectifs.

La Figure 2.5 représente le tracé de la profondeur en fonction de la disparité avec les caractéristiques suivantes : $f = 8$ mm, $b = 9$ mm et des pixels de $7,4 \mu\text{m}$. Dans ce cas, pour une disparité comprise entre 0 et 15 pixels, la profondeur est comprise entre 0,6 m et 10 m.

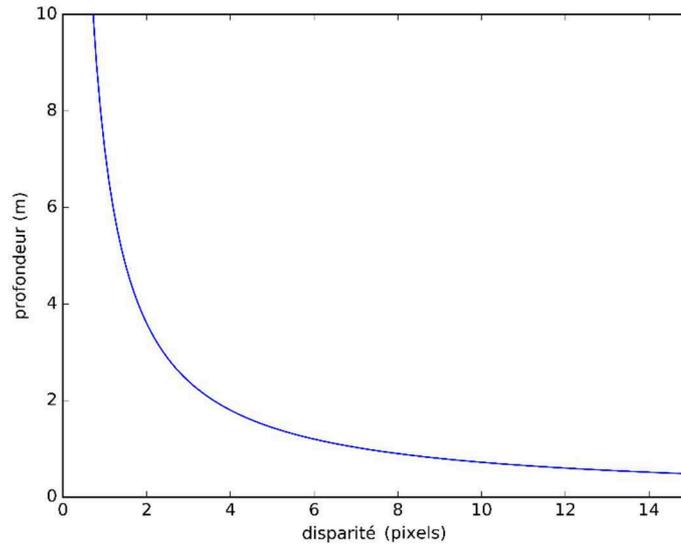


Figure 2.5 Profondeur Z en fonction de la disparité d suivant l'équation (2.1) avec des pixels de 7,4 μm .

La Figure 2.6 représente une vue éclatée du système multi-vues développé à partir des caractéristiques précédentes. Il est composé d'un capteur image et de quatre mini-lentilles. A droite, une vue complète du dispositif est présentée. Ce dispositif est constitué d'une caméra dont la face avant est modifiée pour pouvoir fixer les 4 mini-lentilles. La caméra utilisée est de type CCD, avec une résolution de 2048 x 2048 pixels, de dimensions 15,2 x 15,2 mm et de taille de pixels de 7,4 μm . Les mini-lentilles ont une focale de 8 mm et un diamètre de 8 mm. La distance entre les centres des mini-lentilles est de 9 mm.

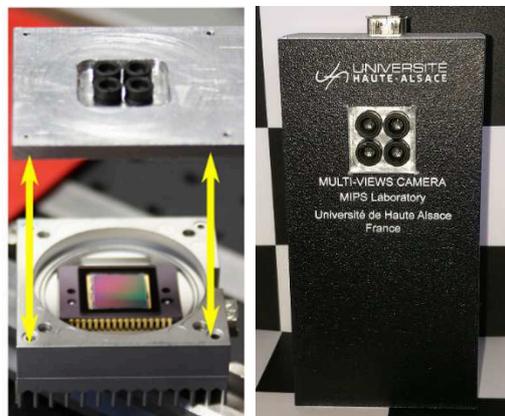


Figure 2.6 Vue éclatée du système multi-vues et vue complète de ce dispositif

Ce système est utilisé dans les chapitres 3 et 4 afin de prouver les intérêts de notre modélisation, de notre calibrage et des algorithmes proposés dans les domaines de la vision industrielle et de l'automobile.

2.3.2. Système multi-caméras

La perception visuelle est un élément important dans le domaine de la vision pour l'automobile. Dans ce travail, la distance de travail est de l'ordre de quelques dizaines de mètres. Ce dispositif a pour

objectif de faciliter la détection d'obstacles dans une scène routière et de pouvoir classifier ces obstacles en quatre zones de profondeur. Des simulations pour développer un système multi-vues adapté à ces grandeurs ont été effectuées. Elles ont permis de conclure qu'un tel dispositif est réalisable avec un capteur ayant les dimensions suivantes : 4008 x 2672 pixels. Cette solution est envisageable avec un capteur Full frame KODAK KAI-11002. Afin d'atteindre pour un coût raisonnable les performances d'un capteur Full frame nécessaire à ce type d'application, nous avons fait le choix de développer un système multi-caméras. Des simulations ont montré qu'il est possible d'obtenir des performances équivalentes entre un dispositif multi-vues et un système multi-caméras.

Ce prototype multi-caméras doit permettre de travailler sur une gamme de distance de quelques dizaines de mètres. Différentes solutions ont été étudiées. Seuls les résultats concernant la solution à quatre caméras sont finalement présentés. Les deux dispositifs (multi-vues et multi-caméras) sont étudiés au moyen de la mesure de la profondeur en fonction de la disparité donnée par l'équation (2.1) :

$$Z = \frac{b \cdot f}{d} \quad (2.1)$$

Avec Z : la profondeur calculée, b : la distance entre les centres optiques des deux objectifs, d : la disparité et f : la focale des objectifs.

Concernant les simulations, plusieurs configurations de montage multi-caméras ont été envisagées. Pour celles-ci, la disparité est comprise entre 0 et 15 pixels et la profondeur mesurable est comprise entre 5 m et 80 m. Les paramètres variables sont la distance b et la focale f . Les contraintes à prendre en compte au cours de cette conception sont les dimensions des caméras qui limitent la taille minimale de l'écart b entre les centres de deux caméras.

Cette configuration a été retenue puisqu'elle répond aux critères concernant la disparité et la profondeur. De plus, elle correspond aussi au système ayant l'encombrement le plus faible. La courbe de la Figure 2.7 représentant la profondeur en fonction de la disparité est calculée avec les valeurs suivantes : $b = 4,62$ cm ; $f = 12$ mm ; une taille de pixels $t_p = 5,3$ μ m. Sur cette courbe, la profondeur est comprise entre 2 m et 80 m.

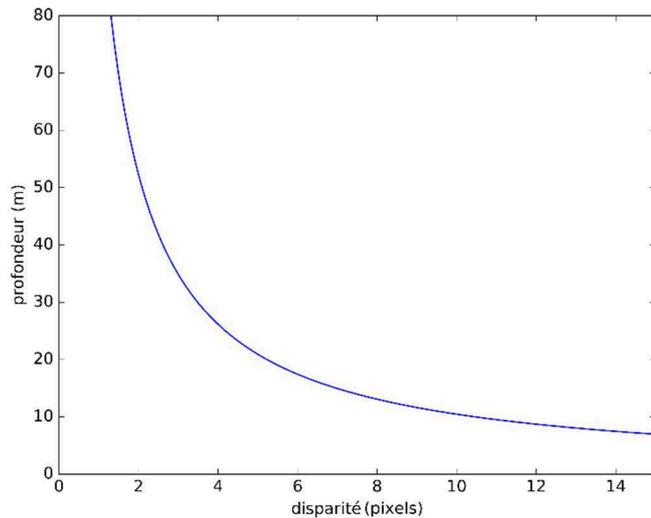


Figure 2.7 Profondeur en fonction de la disparité. Cette courbe est calculée avec un pixel de $5,3 \mu\text{m}$.

Le prototype a donc été réalisé suivant les caractéristiques précédentes. Les caméras choisies sont des caméras IDS de type CMOS avec un capteur de $1/1.8''$, de résolution de 1280×1024 pixels et de taille de pixels de $5,3 \mu\text{m}$. L'objectif présente les caractéristiques suivantes : une focale de 12 mm , une ouverture $f/1,4$. La distance entre les centres de chaque caméra est de $4,62 \text{ cm}$. La Figure 2.8 montre le système développé au laboratoire avec les caractéristiques précédemment citées.



Figure 2.8 Vue du dispositif à quatre caméras

Ce dispositif a été développé pour une application dans le domaine de l'automobile, présentée dans le Chapitre 4, partie 4.5.2.3.

2.4 Développement d'un système multi-vues plénoptique

Le concept de caméra multi-vues plénoptique découle des intérêts du système multi-vues. Celui-ci nous a interrogés sur les avantages possibles d'un dispositif permettant de transformer une caméra 2D en caméra 3D et de réduire la complexité et le nombre de vues par rapport une caméra plénoptique, comme expliqué section 2.2 et représenté par la Figure 2.4 à droite. Cette bague optique est composée de mini-lentilles, et doit être insérée entre l'objectif et le capteur d'une caméra conventionnelle.

Du point de vue du cahier des charges, la distance de travail désirée est de l'ordre d'un mètre. Ce dispositif doit être adaptable sur les caméras conventionnelles utilisées dans le domaine de la vision industrielle, limitant donc les différentes focales de l'objectif principal aux valeurs suivantes : 9 mm, 12 mm, 16 mm, 25 mm et 50 mm.

Sachant que le diamètre de l'arrière de l'objectif est de 25,4 mm, que le diamètre de la lentille arrière est de 12 mm et que l'on veut positionner au moins quatre mini-lentilles, le diamètre maximal des mini-lentilles doit être de 4 mm. De plus, comme la focale d'un système optique complexe est toujours inférieure à la plus petite des focales composant le dispositif, la focale des mini-lentilles doit être supérieure ou égale à 4,5 mm.

2.4.1. Modélisation géométrique

Le schéma sur lequel nous avons travaillé est décrit par la Figure 2.9. Des mini-lentilles sont insérées entre l'objectif et le capteur image, comme pour une caméra plénoptique de type 2.0.

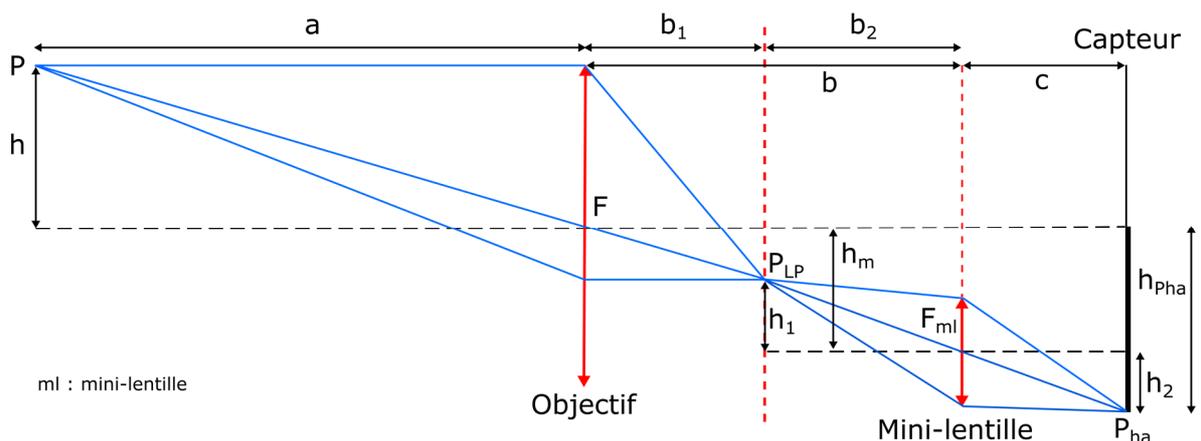


Figure 2.9 Schéma de notre dispositif multi-vues plénoptique et du tracé de rayons à travers une mini-lentille. Dans ce cas, seule une mini-lentille, de focale F_{ml} est représentée.

Sur la Figure 2.9, le point d'émission est défini par sa hauteur h et sa distance a par rapport à l'objectif de focale F . Le paramètre b équivaut à la distance entre l'objectif et les mini-lentilles, b_1 correspond à la focale de l'objectif, b_2 est la distance entre le plan focal image et les mini-lentilles, de focale F_{ml} . c est égale à la distance entre les mini-lentilles et le capteur. La hauteur entre le centre optique de l'objectif et celui de la mini-lentille est noté h_m . La distance b_1 étant dépendante de l'objectif, et le capteur image étant situé dans le plan image des mini-lentilles, les quatre principaux paramètres sont l'emplacement des mini-lentilles dans le chemin optique, leur hauteur et les distances focales de l'objectif et des mini-lentilles, c'est-à-dire b et h_m .

La relation de conjugaison pour l'objectif de focale F est égale à :

$$\frac{1}{F} = \frac{1}{a} + \frac{1}{b_1} \quad (2.2)$$

D'où

$$b_1 = \frac{a \cdot F}{a - F} \quad (2.3)$$

La relation de conjugaison pour la mini-lentille est égale à :

$$\frac{1}{F_{ml}} = \frac{1}{c} + \frac{1}{b_2} \quad (2.4)$$

D'où

$$c = \frac{b_2 \cdot F_{ml}}{b_2 - F_{ml}} \quad (2.5)$$

Sachant que $b_2 = b - b_1 = b - a \cdot F / (a - F)$, la distance mini-lentille – capteur c est alors égale à :

$$c = \frac{F_{ml} \cdot (b - \frac{a \cdot F}{a - F})}{b - \frac{a \cdot F}{a - F} - F_{ml}} \quad (2.6)$$

Avec F_{ml} : la focale de la mini-lentille, b : distance entre l'objectif et la mini-lentille, a : distance entre le point d'émission P et l'objectif de focale F .

Cette expression de la distance mini-lentilles – capteur c est exprimée en fonction des paramètres du système et de la distance de travail a . Cette dernière étant variable, cela engendrerait une distance mini-lentilles-capteur variables, la position idéale du capteur image est dépendante de la distance de travail. Dans la pratique, elle sera fixée pour une distance de travail de 1 m.

En utilisant les triangles semblables sur la Figure 2.9, la hauteur h_2 est égale à :

$$h_2 = \frac{c \cdot h_1}{b_2} \quad (2.7)$$

D'après la Figure 2.9, la hauteur h_{PL} est égale à :

$$h_{PL} = h_m - h_1 \quad (2.8)$$

De même, la hauteur h_{pha} du point P_{ha} sur le capteur est égale à :

$$h_{pha} = h_m + h_2 \quad (2.9)$$

En remplaçant dans l'équation (2.9), h_2 par son expression de l'équation (2.7) et h_1 par son expression de l'équation (2.8),

$$h_{pha} = h_m + c \frac{h_{PL} - h_m}{b_2} \quad (2.10)$$

En utilisant les triangles semblables dans la Figure 2.9, la hauteur h_{pha} est égale à :

$$h_{PL} = h \frac{b_1}{a} = \frac{h \cdot F}{a - F} \quad (2.11)$$

En remplaçant h_{PL} par son expression de l'équation (2.11), c par (2.6) et b_2 par $b - a \cdot F / (a - F)$ dans l'équation (2.10), la hauteur h_{pha} a alors pour expression :

$$h_{pha} = h_m + \frac{F_{ml} \cdot \left(\frac{h \cdot F}{a - F} - h_m \right)}{b - \frac{a \cdot F}{a - F} - F_{ml}} \quad (2.12)$$

Avec h_m : distance entre les axes optiques de l'objectif et de la mini-lentille, de focale F_{ml} ; h : hauteur du point d'émission P et a : sa distance à l'objectif, de focale F et b : distance entre l'objectif et la mini-lentille.

L'expression théorique de la distance c est définie en fonction des paramètres b, F, a, F_{ml} tandis que celle de la hauteur h_{pha} du point P_{ha} l'est en fonction des paramètres b, F, a, F_m et h_m . Tous ces paramètres sont fixés lors de la conception d'un prototype et seront précisés avec la présentation des premiers résultats.

2.4.1.1 Premières expérimentations

Un premier prototype avec des distances fixes, regroupées dans le Tableau 2.1 a été proposé (Figure 2.10).

Grandeurs	b	c	h_m
Valeurs (m)	$3,25 \cdot 10^{-3}$	$2,7 \cdot 10^{-3}$	$2,2 \cdot 10^{-3}$

Tableau 2.1 Distances b, c et h_m fixées pour la conception du premier prototype.

La Figure 2.11 représente un exemple d'image acquise avec le prototype et une image refocalisée. Comme on peut le voir sur l'image refocalisée, le taux de recouvrement est faible (une zone à peine

plus grande que le damier sur les vues d'origine). Dans ce cas, le damier se trouve environ à 50 cm de la caméra. Le damier mesure 9,4 cm par 7,6 cm et les dimensions de la plaque sont de 15 cm par 12 cm.



Figure 2.10 Premier prototype de système multi-vues plénoptique. A gauche, vue du dessus de la bague comprenant les mini-lentilles. A droite, vue complète du dispositif.

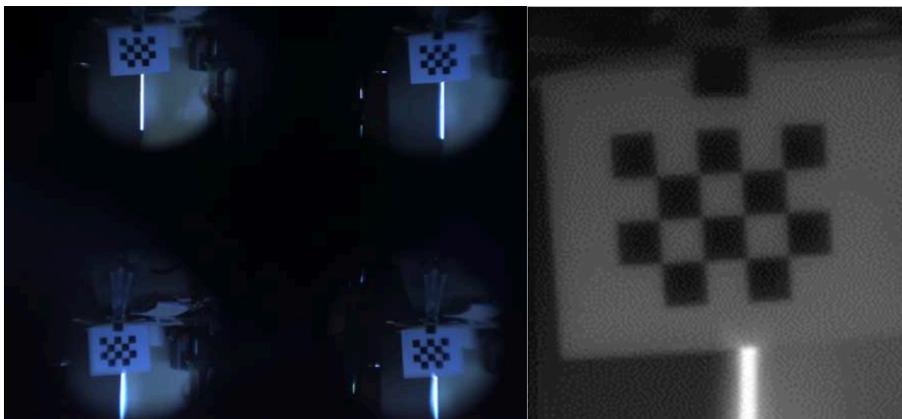


Figure 2.11 A gauche, exemple d'image acquise par le prototype. A droite, image refocalisée sur le damier.

Ce prototype nous a permis de montrer que le concept de système multi-vues plénoptique est envisageable. Cependant, le taux de recouvrement et le champ de vue sont trop faibles avec ce premier prototype. La nécessité de l'étude du lien entre les distances choisies et les caractéristiques telles que le grandissement, le taux de recouvrement et la disparité, a ainsi été mise en évidence. De plus, le paramètre le plus important est la distance b_2 puisqu'il est celui qui influence le plus les caractéristiques.

A partir de ces constats et des spécifications présentées dans cette partie, on cherche à concevoir un système multi-vues plénoptique pouvant fonctionner comme notre dispositif à quatre vues. Ainsi, le nombre des mini-lentilles est limité à quatre, dans la partie suivante. Pour les modéliser, un outil de simulation, basé sur l'optique de Gauss, est proposé dans ce chapitre afin d'étudier cette configuration.

2.4.2. Rappels d'optique de Gauss et d'optique matricielle

Dans le cadre de nos simulations, les rayons lumineux sont modélisés au moyen de l'optique géométrique. Ce modèle consiste à considérer la lumière comme un rayon lumineux défini comme suit : un

rayon lumineux est une ligne entre deux points de l'espace qui représente le chemin suivi par la lumière pour aller d'un point à un autre. Cette représentation de la lumière découle des lois de l'électromagnétisme en effectuant quelques simplifications, principalement liées aux dimensions des systèmes optiques. L'indépendance des chemins optiques des rayons lumineux traversant un instrument optique est un principe fondamental de l'optique géométrique. Grâce aux lois de Snell-Descartes, le changement d'angle du rayon à une interface peut être quantifié. Les éléments optiques sont donc considérés en fonction de leurs interfaces. Associer des interfaces séparant des matériaux d'indice optique différent permet de créer des systèmes optiques variés. Lorsque ces derniers possèdent une symétrie cylindrique, l'axe de révolution est appelé l'axe optique. Ceux sont des systèmes centrés.

Si les rayons se propagent au voisinage de l'axe optique dans le repère en coordonnées cylindriques et ont une faible inclinaison alors la simplification des lois de l'optique géométrique est possible. Ces limitations sont nommées conditions de Gauss, constituant l'optique paraxiale. Elle permet de linéariser les équations et de les modéliser au moyen de l'optique matricielle [Pére84].

Le principe de l'optique matricielle est le suivant : un système optique complet est composé de plusieurs sous-systèmes optiques plus ou moins complexes. Plusieurs matrices caractéristiques existent. Seules les matrices utiles à notre modèle sont présentées dans le Tableau 2.2. Ainsi, les éléments-fonctions utilisés dans notre système sont modélisés, soit par des lentilles minces, soit par des translations entre les lentilles qui ne sont pas collées.

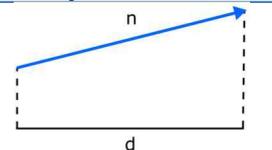
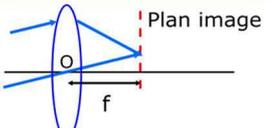
<i>Élément – fonction</i>	<i>Représentation</i>	<i>Matrice de transfert</i>
<i>Translation d'une distance d dans un milieu n</i>		$\begin{pmatrix} 1 & d \\ 0 & n \end{pmatrix}$
<i>Lentille mince</i>		$\begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix}$

Tableau 2.2 Matrices élémentaires utilisées pour notre simulation

Avec d la distance, n : l'indice du milieu, f : la focale de la lentille mince, O : le centre optique de la lentille mince. Dans notre cas, l'indice n est toujours égal à 1 puisque les translations s'effectuent dans l'air.

L'association de plusieurs éléments ou fonctions optiques s'effectue au moyen de multiplications matricielles. Par exemple, soit un dispositif optique composé de trois éléments et fonctions optiques dont les matrices de transfert sont nommées A, B et C, avec A étant le premier élément et C le dernier élément. La matrice résultat, représentant le dispositif complexe R, a pour équation :

$$R = C.B.A \tag{2.13}$$

Les matrices de transfert des éléments-fonctions optiques ainsi que les matrices équivalentes des systèmes optiques complexes sont des matrices (A,B,C,D). Un système optique peut être défini, entre deux plans conjugués (notamment ses plans principaux), par la matrice E :

$$E = \begin{pmatrix} \gamma & 0 \\ -V & \frac{n_s}{n_e} G \end{pmatrix} \quad (2.14)$$

Avec V : la vergence du système, γ : le grandissement transversal, n_s et n_e : les indices des milieux d'entrée et de sortie et G : le grandissement angulaire.

Il existe deux plans principaux dans un système optique centré (en pointillés gris sur la Figure 2.12) :

- le plan principal image : plan où chaque rayon incident parallèle à l'axe optique croise le rayon émergeant correspondant ;
- le plan principal objet : plan où chaque rayon émergeant parallèle à l'axe optique croise le rayon incident correspondant.

Ces deux plans sont perpendiculaires à l'axe optique. Leur intersection avec celui-ci est un point principal. Le point principal objet est noté H tandis que le point principal image est noté H'.

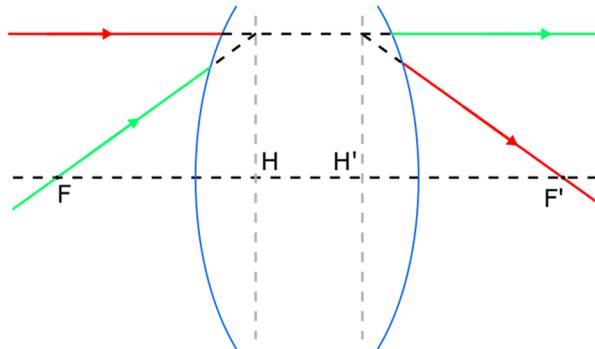


Figure 2.12 Plans principaux et points cardinaux d'un système optique centré. Tous les points cardinaux sont définis sur l'axe optique. F est le point focal objet, F' est le point focal image, H est le point principal objet et H' est le point principal image.

En utilisant la relation de Helmholtz-Lagrange et l'expression des paramètres en fonction des plans principaux, E a pour expression :

$$E = \begin{pmatrix} \gamma & 0 \\ -V & \frac{1}{\gamma} \end{pmatrix} = \begin{pmatrix} \frac{\overline{HA}}{\overline{H'A'}} & 0 \\ -\frac{1}{f} & \frac{\overline{H'A'}}{\overline{HA}} \end{pmatrix} \quad (2.15)$$

Avec $f = \overline{HF}$ est la focale du système, \overline{HA} est la distance entre le point objet A et le point principal objet H et $\overline{H'A'}$ est la distance entre le point principal image H' et le point objet A' (Figure 2.12).

En optique de Gauss, tous les éléments sont sur le même axe optique. On cherche un système d'écriture matricielle pour la prise en compte de la position en hauteur des mini-lentilles.

Hauteur et angle avec notation standard pour une lentille mince de focale f

A partir de la Figure 2.13 à gauche, les angles θ et θ' et les hauteurs h et h' sont reliés par les relations suivantes :

$$h' = h \tag{2.16}$$

$$\theta' = -\frac{h}{f} + \theta \tag{2.17}$$

En écrivant les équations (2.16) et (2.17) sous forme matricielle, on obtient :

$$\begin{pmatrix} h' \\ \theta' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \begin{pmatrix} h \\ \theta \end{pmatrix} \tag{2.18}$$

L'équation (2.18) est identique à celle représentant une lentille mince et décrite dans le Tableau 2.2.

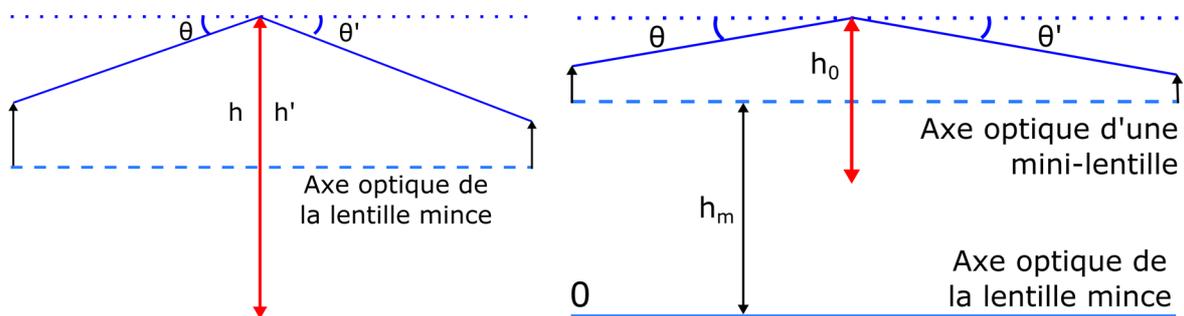


Figure 2.13 A gauche, représentation d'une lentille mince, de focale f et à droite, représentation d'une mini-lentille, de focale f_{ml} , décalée d'une distance h_m par rapport à l'axe optique d'une lentille mince.

Hauteur et angle d'une mini-lentille décalée de l'axe optique d'une distance h_m

L'optique de Gauss est définie pour travailler avec des composants optiques alignés sur le même axe. Pour modéliser les mini-lentilles, qui sont décalées en hauteur, nous avons trouvé que la position d'une mini-lentille, pouvait se formuler avec l'optique de Gauss (où les éléments optiques sont généralement alignés sur l'axe optique), en modifiant l'écriture habituelle en matrices 2x2, par une écriture matricielle 3x3.

Dans cette partie, la Figure 2.13 à droite est la figure de référence. La hauteur h_0 est la hauteur du rayon dans le repère de la mini-lentille. D'un point de vue géométrique, les hauteurs h et h' sont liés par l'équation suivante :

$$h' = h = h_m + h_0 \quad (2.19)$$

De même les angles θ et θ' le sont par l'équation suivante :

$$\theta' = -\frac{h_0}{f} + \theta \quad (2.20)$$

Or, d'après la Figure 2.13,

$$h_0 = h - h_m \quad (2.21)$$

En remplaçant h_0 dans l'équation (2.20) par son expression de l'équation (2.21), on obtient :

$$\theta' = -\frac{h}{f} + \frac{h_m}{f} + \theta \quad (2.22)$$

Il y a donc un terme en plus : h_m/f , par rapport à l'équation (2.17), représentant l'angle θ' en fonction de θ pour une lentille mince.

On peut l'intégrer dans le système matriciel en adoptant la notation suivante :

$$\begin{pmatrix} \theta' \\ h' \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{f} & 1 & \frac{h_m}{f} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \theta \\ h \\ 1 \end{pmatrix} \quad (2.23)$$

Les éléments entourés par le carré rouge correspondent à la matrice élémentaire d'une lentille mince comme indiqué dans le Tableau 2.2. L'élément entouré en bleu est le terme de translation en hauteur. Tous les éléments nécessaires à la simulation de la formation d'une image avec notre système optique complexe, au moyen de l'optique de Gauss et du tracé de rayons, ont été présentés. Les simulations sont effectuées avec la représentation, présentée ci-dessus, sous forme de matrice 3x3. Les autres matrices de transformation sont donc notées avec ce formalisme pour les simulations. Comme expliqué précédemment, il est utilisé afin de pouvoir prendre en compte le décalage du centre des mini-lentilles par rapport au centre optique de l'objectif principal.

2.4.3. Simulation et tracé de rayons du dispositif multi-vues plénoptique

Le montage optique est composé d'un objectif, de quatre mini-lentilles et d'un capteur. Deux configurations de placement des éléments les uns par rapport aux autres ont été étudiées : la première consistant à placer les mini-lentilles après le plan focal de l'objectif (configuration képlérienne (cf. Figure 2.14)) et la seconde pour laquelle les mini-lentilles sont situées avant le plan focal de l'objectif (configuration galiléenne (cf. Figure 2.15)). a équivaut à la distance entre le point d'émission et l'objectif.

Sur les Figure 2.14 et Figure 2.15, b correspond à la distance entre l'objectif et les mini-lentilles, b_1 est égale à la focale de l'objectif, b_2 est égale à la distance entre le plan focal et les mini-lentilles et c est la distance entre les mini-lentilles et le capteur. Afin de simuler le comportement des rayons lumineux au sein du système optique, le modèle a été développé en optique matricielle.

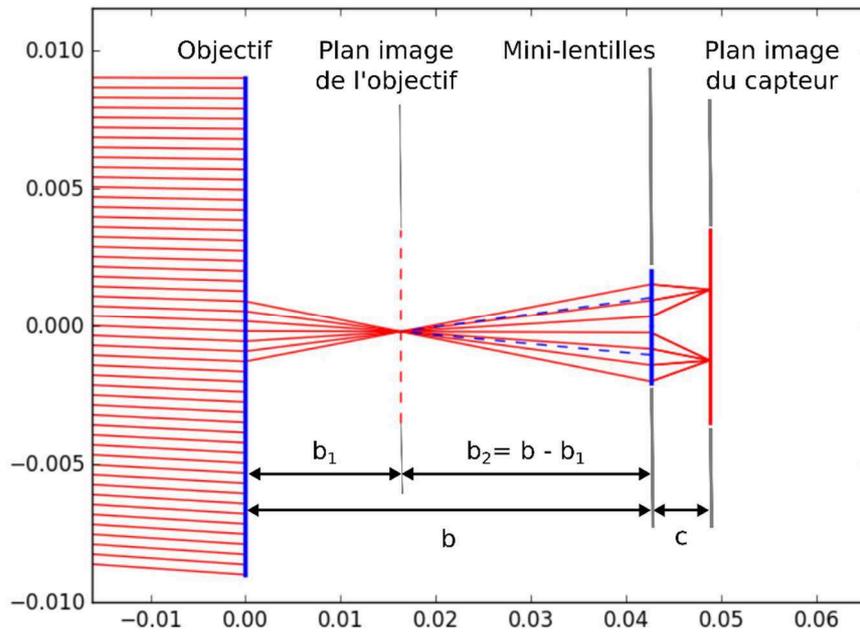


Figure 2.14 Schéma du montage optique de la caméra multi-vues plénoptique, composé d'un objectif représenté comme une lentille mince, de deux mini-lentilles placées après le plan focal de l'objectif et d'un capteur image. Cette configuration est aussi appelée configuration képlérienne

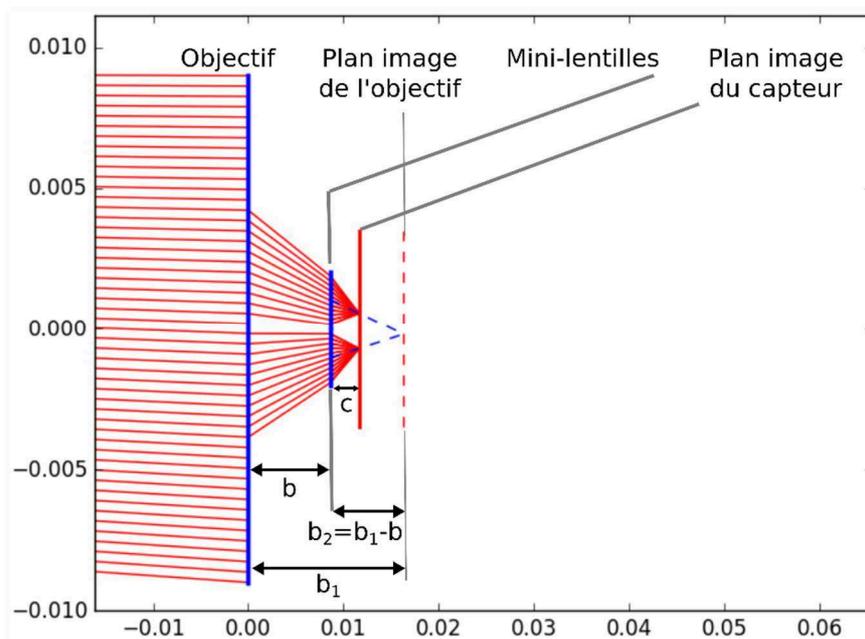


Figure 2.15 Schéma du montage optique de la caméra multi-vues plénoptique, composé d'un objectif représenté comme une lentille mince, de deux mini-lentilles placées avant le plan focal de l'objectif et d'un capteur image. Elle est aussi nommée configuration galiléenne.

A partir du point d'émission de la lumière, le rayon lumineux effectue une translation, d'une distance a , jusqu'à l'objectif modélisée par la matrice :

$$T_a = \begin{pmatrix} 1 & a & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.24)$$

Puis il traverse l'objectif, considéré comme une lentille mince, de focale F , et représenté par :

$$F = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{F} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.25)$$

Entre la sortie de l'objectif et l'entrée des microlentilles, le rayon est translaté d'une distance b :

$$T_b = \begin{pmatrix} 1 & b & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.26)$$

Le trajet au travers des microlentilles est modélisé au moyen des deux matrices suivantes, qui prennent en compte le décalage des mini-lentilles par rapport au centre optique de l'objectif principal :

$$F_{ha} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{F_{ml}} & 1 & \frac{h_m}{F_{ml}} \\ 0 & 0 & 1 \end{pmatrix} \text{ et } F_{ba} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{F_{ml}} & 1 & -\frac{h_m}{F_{ml}} \\ 0 & 0 & 1 \end{pmatrix} \quad (2.27) \text{ et } (2.28)$$

Avec F_{ml} : la focale des mini-lentilles et h_m : la distance entre les centres optiques de l'objectif et celui d'une des deux mini-lentilles.

Puis une dernière translation, d'une distance c , est effectuée par le rayon entre les sorties des microlentilles et le capteur :

$$T_c = \begin{pmatrix} 1 & c & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.29)$$

Comme présenté dans la partie précédente, la modélisation complète du dispositif équivaut à la multiplication des différentes matrices correspondant aux étapes précédemment décrites :

$$M_{ha} = T_c \cdot F_h \cdot T_b \cdot F \cdot T_a = \quad (2.30)$$

$$\begin{pmatrix} 1 - \frac{c}{F_{ml}} - \frac{b}{F} - \frac{c}{F} + \frac{b \cdot c}{F \cdot F_{ml}} & a + b + c - \frac{a \cdot c}{F_{ml}} - \frac{b \cdot c}{F_{ml}} - \frac{a \cdot b}{F} - \frac{a \cdot c}{F} + \frac{a \cdot b \cdot c}{F \cdot F_{ml}} & \frac{c \cdot h_m}{F_{ml}} \\ \frac{-F - F_{ml} + b}{F \cdot F_{ml}} & 1 - \frac{a}{F_{ml}} - \frac{b}{F_{ml}} - \frac{a}{F} + \frac{a \cdot b}{F \cdot F_{ml}} & \frac{h_m}{F_{ml}} \\ 0 & 0 & 1 \end{pmatrix} M_{ba} = T_c \cdot F_h \cdot T_{ba} \cdot F_{ba} \cdot T_a =$$

$$\begin{pmatrix} 1 - \frac{c}{F_{ml}} - \frac{b}{F} - \frac{c}{F} + \frac{b \cdot c}{F \cdot F_{ml}} & a + b + c - \frac{a \cdot c}{F_{ml}} - \frac{b \cdot c}{F_{ml}} - \frac{a \cdot b}{F} - \frac{a \cdot c}{F} + \frac{a \cdot b \cdot c}{F \cdot F_{ml}} & -\frac{c \cdot h_m}{F_{ml}} \\ \frac{-F - F_{ml} + b}{F \cdot F_{ml}} & 1 - \frac{a}{F_{ml}} - \frac{b}{F_{ml}} - \frac{a}{F} + \frac{a \cdot b}{F \cdot F_{ml}} & -\frac{h_m}{F_{ml}} \\ 0 & 0 & 1 \end{pmatrix} \quad (2.31)$$

Pour obtenir les coordonnées des points à partir du point d'émission, il suffit de multiplier le point P_e par l'une des matrices M_{ha} ou M_{ba} :

$$P_{ha} = M_{ha} \cdot P_e = M_{ha} \cdot \begin{pmatrix} h \\ \theta \end{pmatrix} \quad (2.32)$$

$$P_{ba} = M_{ba} \cdot P_e = M_{ba} \cdot \begin{pmatrix} h \\ \theta \end{pmatrix} \quad (2.33)$$

Avec h : hauteur du point P_e et θ : l'angle d'émission de P_e .

Cette simulation permet d'avoir un aperçu du comportement des rayons dans le dispositif et de prévoir la position des rayons sur le capteur. La Figure 2.16 illustre un exemple de simulation obtenu avec les paramètres résumés dans le Tableau 2.3. La représentation de la simulation est une vue en coupe où seules deux des quatre mini-lentilles sont visibles. Ces simulations permettent de visualiser le comportement des rayons lumineux à travers notre système.

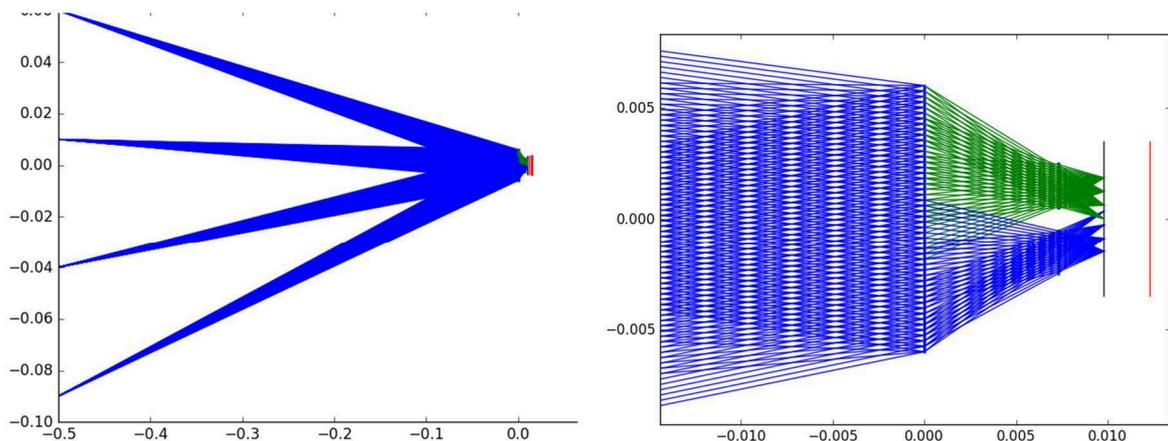


Figure 2.16 Simulation du système multi-vues plénoptique. La configuration utilisée est celle où les mini-lentilles sont placées avant le plan focal de l'objectif principal. A gauche, vue d'ensemble avec des rayons émis à une distance de 0,5 m du système. A droite, zoom sur le capteur (trait noir à droite), les rayons en vert représentent les rayons traversant la mini-lentille du haut tandis que les bleus passent par la mini-lentille du bas et le trait rouge représente le plan de focalisation de l'objectif.

Grandeurs	a	b	b_2	taille pixel t_p	F	F_{ml}	h_m	H
Valeurs (m)	0,5	$7,3 \cdot 10^{-3}$	$-5 \cdot 10^{-3}$	$3 \cdot 10^{-6}$	$12 \cdot 10^{-3}$	$5 \cdot 10^{-3}$	$1,5 \cdot 10^{-3}$	(0,06 0,01 -0,04 -0,09)

Tableau 2.3 Valeurs utilisées pour un exemple de simulation.

2.4.4. Caractérisation de notre système multi-vues plénoptique

A partir de ces simulations, il est nécessaire de pouvoir caractériser le dispositif aux moyens de critères conventionnels. Plusieurs caractéristiques ont été définies, en fonction de notre système :

- Taux de recouvrement, c'est-à-dire le pourcentage commun d'une même scène présent sur deux images
- Parallaxe ou disparité, c'est-à-dire le décalage relatif entre deux images, d'objets placés à différentes profondeurs
- Grandissement
- Ouverture numérique des mini-lentilles

Plusieurs raisons expliquent le choix de ces paramètres. Le taux de recouvrement est un paramètre important puisque les différentes vues doivent avoir une grande partie commune, utilisée par exemple pour calculer des images refocalisées. Comme notre dispositif peut être utilisé comme une caméra de profondeur, il est nécessaire de pouvoir prédire la disparité mesurable avec le système. Ensuite, la focale équivalente étant difficilement quantifiable, la caractérisation du dispositif avec le grandissement est un choix évident. Enfin, d'après Ren Ng [NLBD05], il est nécessaire que la valeur de l'ouverture numériques des mini-lentilles soit proche de celle de l'objectif principal pour minimiser le chevauchement entre les différentes vues sur le capteur.

Taux de recouvrement

Le taux de recouvrement est illustré par la Figure 2.17. Cette figure permet de représenter les différentes images par rapport aux éléments optiques composant le système. Dans ce cas, nous avons retenu la configuration où les mini-lentilles sont placées après le plan focal de l'objectif. Le principe de calcul demeure identique avec la seconde disposition des éléments optiques.

A partir des triangles semblables, la taille de l'image T_{i0} sur le plan focal de l'objectif et la taille de l'image t_{i0} sur le capteur sont reliés par la relation suivante :

$$\frac{T_{i0}}{t_{i0}} = \frac{b_2}{c} \quad (2.34)$$

Le taux de recouvrement, sur la Figure 2.17, est défini comme suit :

$$T_R = \frac{T_{i1}^+ - T_{i0}^-}{T_{i1}^+ - T_{i1}^-} \quad (2.35)$$

Il peut aussi être défini à partir des points T_{i0}^+ , T_{i0}^- et T_{i1}^+ par l'équation suivante :

$$T_R = \frac{T_{i1}^+ - T_{i0}^-}{T_{i0}^+ - T_{i0}^-} \quad (2.36)$$

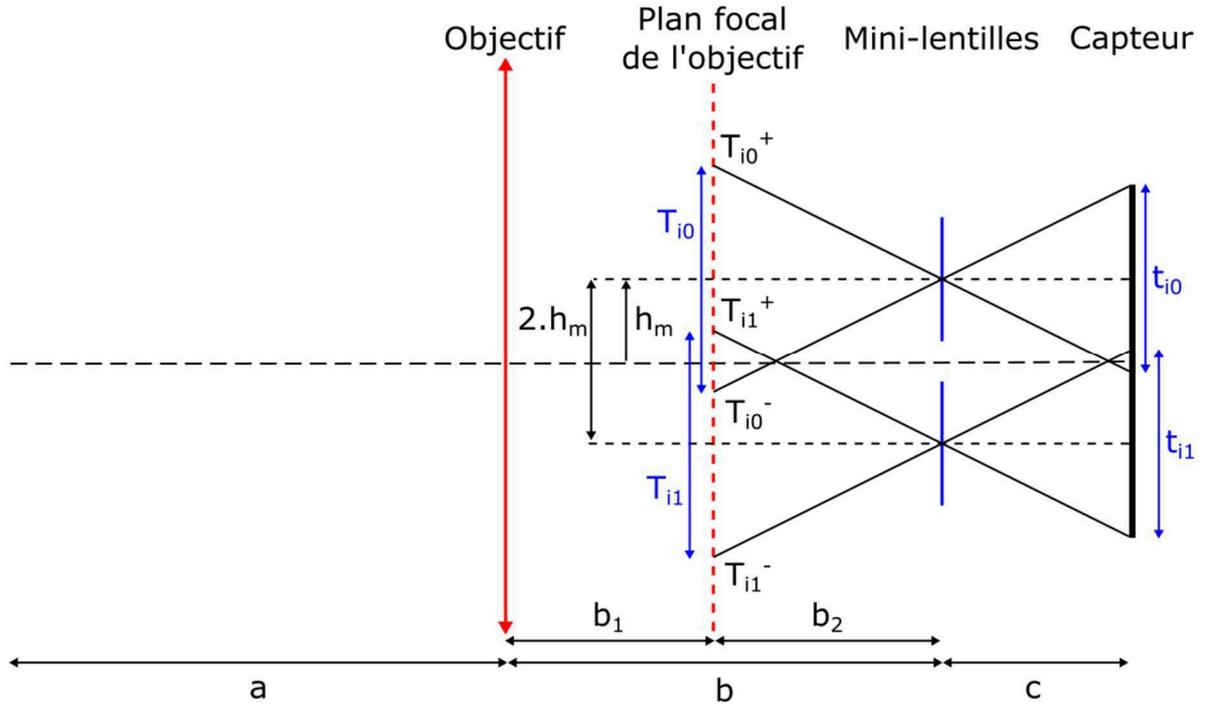


Figure 2.17 Illustration du taux de recouvrement avec notre dispositif multi-vues plénoptique. Les points T_{i0}^+ et T_{i0}^- sont les deux points extrêmes de l'image T_{i0} , objet imagé par la mini-lentille 0 et dont l'image sur le capteur est t_{i0} . De même, les points T_{i1}^+ et T_{i1}^- sont les deux points extrêmes de l'image T_{i1} , objet imagé par la mini-lentille 1 et dont l'image sur le capteur est t_{i1} .

D'après la Figure 2.17, les points T_{i1}^+ , T_{i1}^- et T_{i0}^- sont définis par les équations suivantes et à partir de l'équation (2.34):

$$T_{i1}^+ = -h_m + \frac{b_2}{c} \cdot \frac{t_{i1}}{2} \quad (2.37)$$

$$T_{i1}^- = -h_m - \frac{b_2}{c} \cdot \frac{t_{i1}}{2} \quad (2.38)$$

$$T_{i0}^- = h_m - \frac{b_2}{c} \cdot \frac{t_{i1}}{2} \quad (2.39)$$

En remplaçant T_{i1}^+ , T_{i1}^- et T_{i0}^- dans l'équation (2.35), par leurs expressions des équations (2.37), (2.38), (2.39), le taux de recouvrement T_R est égale à :

$$T_R = 1 - \frac{2 \cdot h_m \cdot c}{\left(b - \frac{a \cdot F}{a - F}\right) \cdot t_{i1}} \quad (2.40)$$

Avec h_m : la demi-distance entre les centres optiques de deux mini-lentilles, c : la distance entre les mini-lentilles et le plan image, b : la distance entre les mini-lentilles et l'objectif et t_{i1} : la taille d'une image à travers une mini-lentille.

Avec la géométrie de notre système, on montre que le taux de recouvrement dépend de la position à laquelle on place les mini-lentilles : en hauteur (h_m) et sur l'axe horizontal (b). La valeur de c étant considérée ici comme une constante de la figure (on ne change pas c au cours des acquisitions).

Parallaxe ou disparité

Dans cette partie, on définit la disparité ou la parallaxe d comme sur la Figure 2.18.

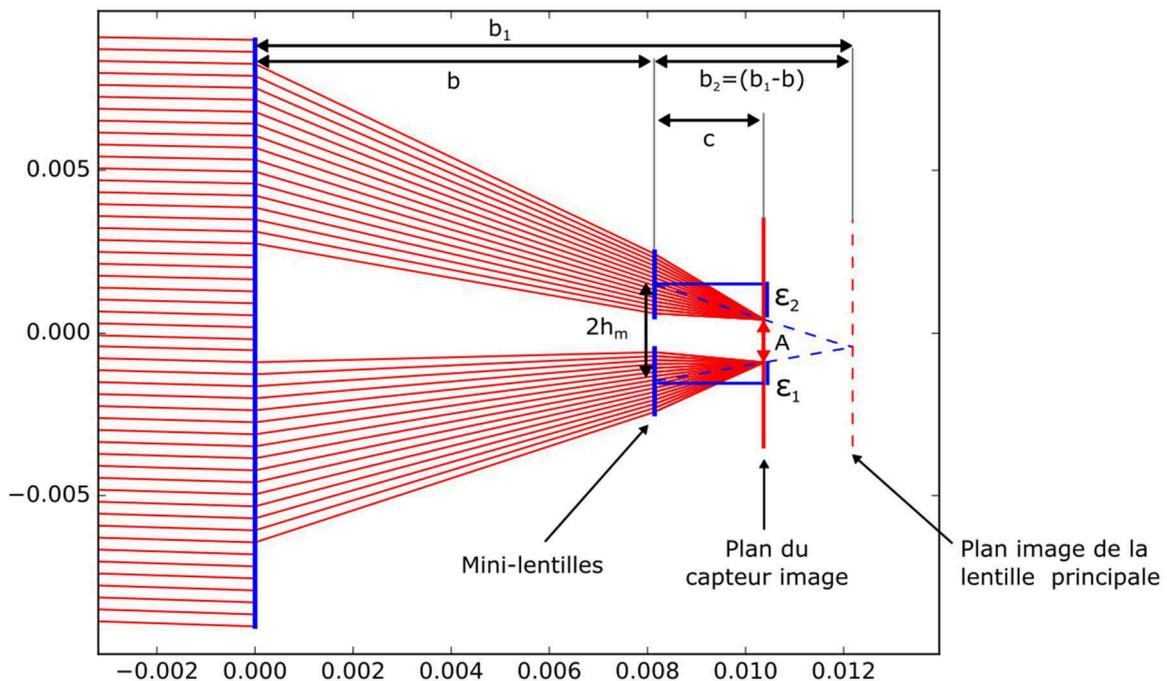


Figure 2.18 Illustration de la disparité en fonction des paramètres de notre dispositif.

A partir du théorème de Thalès,

$$\frac{A}{2 \cdot h_m} = \frac{b_2 - c}{b_2} \quad (2.41)$$

Or :

$$A + \varepsilon_1 + \varepsilon_2 = 2 \cdot h_m \quad (2.42)$$

D'où

$$A = 2 \cdot h_m - (\varepsilon_1 + \varepsilon_2) \quad (2.43)$$

Avec $d = (\varepsilon_1 + \varepsilon_2)$: disparité

En remplaçant c par $F_{ml} \cdot b_2 / (b_2 - F_{ml})$ et A par son expression de l'équation (2.43) dans l'équation (2.41), on obtient :

$$\frac{2 \cdot h_m - d}{2 \cdot h_m} = \frac{b_2 - c}{b_2} \quad (2.44)$$

La disparité d a donc pour expression :

$$d = \frac{2 \cdot h_m \cdot c}{b - \frac{a \cdot F}{a - F}} \quad (2.45)$$

Dans l'usage, il est plus habituel d'avoir une expression de la disparité qui tende vers 0, quand la distance des objets est assimilable à l'infini (c'est-à-dire quand elle tend vers l'infini). Dans ce cas de figure, nous exprimons la disparité par l'expression :

$$d = \frac{2 \cdot h_m \cdot c}{b - \frac{a \cdot F}{a - F}} - \frac{2 \cdot h_m \cdot c}{b - F} \quad (2.46)$$

Avec h_m : la demi-distance entre les centres optiques de deux mini-lentilles, c : la distance entre les mini-lentilles et le plan image, b : la distance entre l'objectif et le capteur, a : la distance entre l'objet et l'objectif et F : la distance focale de l'objectif.

Cela revient simplement à aligner les repères des deux vues projetées. La disparité d dépend donc du positionnement des mini-lentilles : suivant l'axe horizontal (b) et l'axe vertical (h_m), de la position des objets par rapport au dispositif (a), de la focale de l'objectif principal (F) et de la position du capteur sur l'axe horizontal (c).

Grandissement

Le grandissement G correspond au ratio entre la hauteur h_h du point sur le capteur et la hauteur h du point d'émission P_e . Pour l'obtenir, on utilise les triangles semblables, décrits par la Figure 2.9 et la Figure 2.19 sachant que $h_{ml} = h_m - h_1$:

$$\frac{h}{a} = \frac{h_{ml}}{b_1} \quad (2.47)$$

$$\frac{h_m - h_{ml}}{b_2} = \frac{h_h}{c} \quad (2.48)$$

A partir de ces deux équations, on obtient les expressions de h_h et de h_{ml} :

$$h_{ml} = \frac{b_1 \cdot h}{a} \text{ et } h_h = \frac{c \cdot (h_m - h_{ml})}{b_2} \quad (2.49)$$

$$h_h = \frac{c \cdot (h_m - h_{ml})}{b_2} \quad (2.50)$$

En remplaçant h_{ml} dans l'équation (2.50) par son expression de l'équation (2.49), on obtient :

$$h_h = \frac{c \cdot h_m}{b_2} - \frac{c \cdot b_1}{a \cdot b_2} h \quad (2.51)$$

En considérant que h_m est égal à 0, et sachant que le grandissement équivaut au ratio entre la hauteur h_h et la hauteur h , alors on a :

$$\frac{h_h}{h} = - \frac{c \cdot b_1}{a \cdot b_2} \quad (2.52)$$

En remplaçant dans l'équation (2.52), c par $F_{ml} \cdot b_2 / b_2 \cdot F_{ml}$ et b_1 par $F \cdot a / (a - F)$, le grandissement G est égal à :

$$G = - \frac{F \cdot F_{ml}}{(a - F) \cdot (b - F_{ml}) - a \cdot F} \quad (2.53)$$

Avec F : la focale de l'objectif, F_{ml} : la focale des mini-lentilles, a : la distance entre l'objet et l'objectif et b : la distance entre l'objectif et les mini-lentilles.

Du point de vue géométrique, le grandissement G dépend des focales des mini-lentilles (F_{ml}) et de l'objectif (F), de la position des mini-lentilles suivant l'axe optique (b) et de la position de l'objet à imager par rapport à l'objectif (a).

Ouverture numérique des mini-lentilles

L'ouverture numérique d'une mini-lentille ON_{ml} est définie en utilisant la formulation proposée par Perwass et al. [PeWi12] et ayant pour expression :

$$ON_{ml} = ON_L - \frac{c}{D_L} \quad (2.54)$$

Avec ON_L : ouverture numérique de l'objectif, c : distance mini-lentilles – capteur et D_L : diamètre de l'objectif.

Cette caractéristique permet de prévoir si les différentes sous-images vont se chevaucher. Elle dépend de la position des mini-lentilles par rapport au capteur suivant l'axe optique (c), de deux caractéristiques de l'objectif : son ouverture numérique (ON_L) et son diamètre (D_L).

A partir de l'ouverture numérique des mini-lentilles ON_{ml} , il est possible de définir la limite de diffraction de ce dispositif, ce qui permet de déterminer la taille de pixels minimale que le dispositif peut avoir sans présenter de diffractions. La limite de diffraction est caractérisée par la taille d de la tâche d'Airy, exprimé par :

$$d = 2,44 * \lambda * ON_{ml} \quad (2.55)$$

Avec λ : longueur d'onde (ici 550 nm) et ON_{ml} : ouverture numérique des mini-lentilles.

Dans cette partie, nous avons fourni quatre caractéristiques du système et un outil de simulation du comportement des rayons au travers de notre dispositif. Les éléments proposés dans cette partie aident à fixer les distances entre les éléments composant le dispositif en fonction de critères comme le taux de recouvrement, la disparité et le grandissement.

2.4.5. Exemples d'acquisition du système multi-vues plénoptique

Chacune des caractéristiques, présentées dans la section précédente, est vérifiée de manière expérimentale au moyen du montage optique (Figure 2.19) suivant :

- Un capteur CCD ayant pour dimension 1/2,3" (6,140 mm x 4,605 mm), dont la taille de pixel est de 1,25 μm et avec une résolution de 4912 x 3684 pixels. Ce capteur est utilisé en mode binning, la taille des pixels est de 2,5 μm et la résolution est de 2456 x 1842 pixels ;
- Un réseau de 4 mini-lentilles de 5 mm de focale et de 2 mm de diamètre ;
- Un objectif avec monture C de focale 16 mm.

Cette section est divisée en deux parties : la première permettant de mettre en évidence l'influence du paramètre b_2 sur le taux de recouvrement. Cette étape aboutit à la sélection d'une configuration utilisée pour tester les différentes caractéristiques. Pour chacun des critères, la valeur théorique est comparée à celle calculée à partir des données expérimentales acquises avec le système de la Figure 2.19.

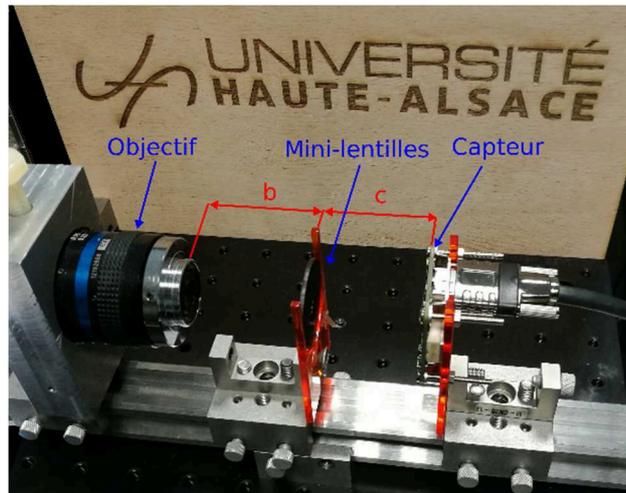


Figure 2.19 Vue du montage expérimental composé de l'objectif à gauche, des mini-lentilles au centre et du capteur image à droite.

2.4.5.1 Influence du paramètre b sur le taux de recouvrement

En utilisant ce montage, une première expérimentation pour mettre en évidence l'influence de la distance b sur le taux de recouvrement est effectuée. Pour chacun des cas, une image illustrant ce cas et la valeur de b sont indiqués. Au cours de cette expérimentation, il a été mis en évidence qu'il existe une plage de distance autour de la focale de l'objectif pour laquelle aucune image ne peut être acquise.

Quatre configurations du système sont étudiées. La distance b entre les mini-lentilles et l'objectif est fixée pour chacun des cas et le capteur est ensuite réglé de manière à acquérir une image nette de la scène. Pour la première disposition, l'image est inversée (Figure 2.20 (1)) car la configuration utilisée est képlérienne et est effectuée avec une distance b égale à 27,6 mm. Dans ce cas, la taille des sous-images est trop importante pour que les quatre vues soient entièrement imagées par le capteur. On choisit de se placer dans des configurations galiléennes pour les trois autres cas. L'image (2) de la Figure 2.20 correspond à un cas où le taux de recouvrement entre les images est faible, et est obtenue avec une distance b égale à 10,3 mm. L'image (3) de la Figure 2.20 est obtenue avec une distance b égale à 4,6 mm et le taux de recouvrement est intéressant. Enfin, l'image (4) correspond au cas où le support des mini-lentilles est collé à l'objectif, ce qui équivaut à une distance b de 3,6 mm. Le taux de recouvrement est similaire à celui obtenu avec la solution choisie pour l'image (3).

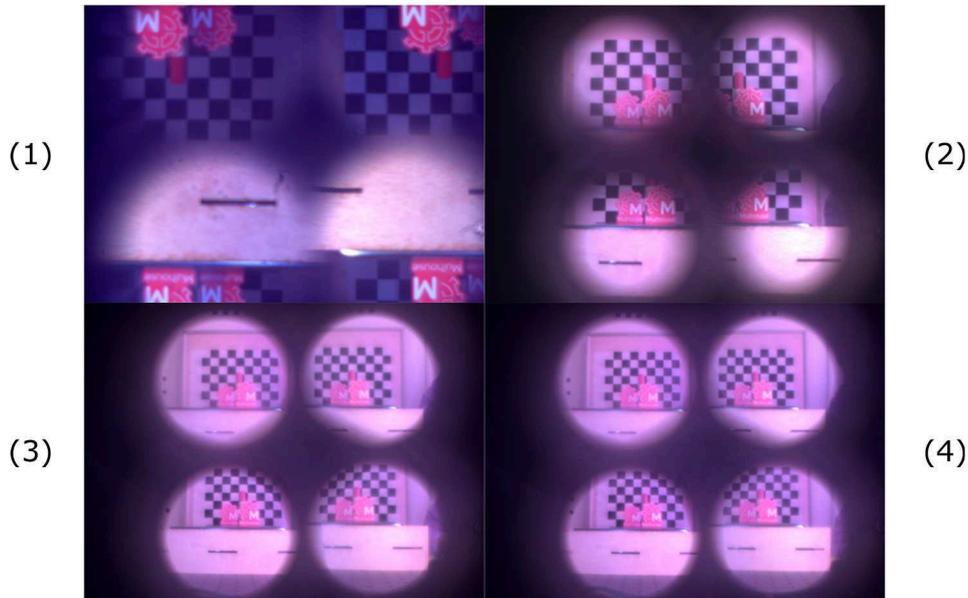


Figure 2.20 Quatre acquisitions d'une même scène effectuées chacune avec une valeur différente de la distance b entre l'objectif et les mini-lentilles : (1) $b = 27,6$ mm, image inversée car configuration képlérienne ; (2), (3) et (4) : configuration galiléenne ; (2) $b = 10,3$ mm ; (3) $b = 4,6$ mm et (4) $b = 3,6$ mm.

Dans la suite de cette partie, toutes les comparaisons entre les mesures théoriques et expérimentales sont effectuées avec une unique configuration, celle utilisée pour l'image (3) de la Figure 2.20. Tous les paramètres expérimentaux sont résumés dans le Tableau 2.4. La scène est composée de quatre éléments : deux objets, intitulés roue de Mulhouse, respectivement placés à une distance $a_{r1} = 56$ cm et $a_{r2} = 66$ cm de l'objectif, un feutre placé à une distance $a_f = 76$ cm de l'objectif et un damier situé à une distance a_d de 1,54 m.

Grandeurs	b	b_1	b_2	C	taille pixel t_p	F	F_{ml}	h_m
Valeurs (m)	$4,6 \cdot 10^{-3}$	$16 \cdot 10^{-3}$	$11,4 \cdot 10^{-3}$	$2,5 \cdot 10^{-3}$	$2,5 \cdot 10^{-6}$	$16 \cdot 10^{-3}$	$5 \cdot 10^{-3}$	$1,05 \cdot 10^{-3}$

Tableau 2.4 Paramètres expérimentaux utilisés pour la comparaison des données théoriques et expérimentales.

2.4.5.2 Comparaison des valeurs théoriques et expérimentales des caractéristiques

Toutes les comparaisons sont effectuées avec la configuration sélectionnée ci-dessus. Les caractéristiques testées sont celles définies dans la section 2.4.4.

Taux de recouvrement

Pour évaluer le bien fondé de notre définition, le taux de recouvrement est mesuré de manière expérimentale en relevant les positions des objets présents dans la scène, sur les quatre sous-images. A partir des résultats présentés dans le Tableau 2.5, on constate que les valeurs théoriques et expérimentales du taux de recouvrement sont proches. Notre définition du taux de recouvrement semble bien prévoir le comportement de notre dispositif.

Méthodes		Taux de recouvrement T_R
Expression littérale (équation (2.40))	$T_R = 1 - \frac{2 \cdot h_m \cdot c}{\left(b - \frac{a \cdot F}{a - F}\right) \cdot t_{i1}}$	0,79
Expérimentation	$\frac{\text{Zone de recouvrement}}{\text{Taille image}}$	0,78

Tableau 2.5 Comparaison entre les résultats théoriques et expérimentaux du taux de recouvrement. Les deux valeurs obtenues sont similaires.

Parallaxe ou disparité

Sur la Figure 2.21, les mesures expérimentales de la disparité sont représentées en rouge pour les quatre objets définis précédemment ($a_{r1} = 56$ cm, $a_{r2} = 66$ cm, $a_f = 76$ cm et $a_d = 154$ cm) tandis que les valeurs théoriques sont représentées en bleu. Les valeurs théoriques et expérimentales ne sont pas égales mais présentent la même allure. Pour la profondeur a_d , les deux valeurs sont très proches. Malgré ces différences, la définition théorique de la disparité permet de prévoir les valeurs de la disparité à un pixel près.

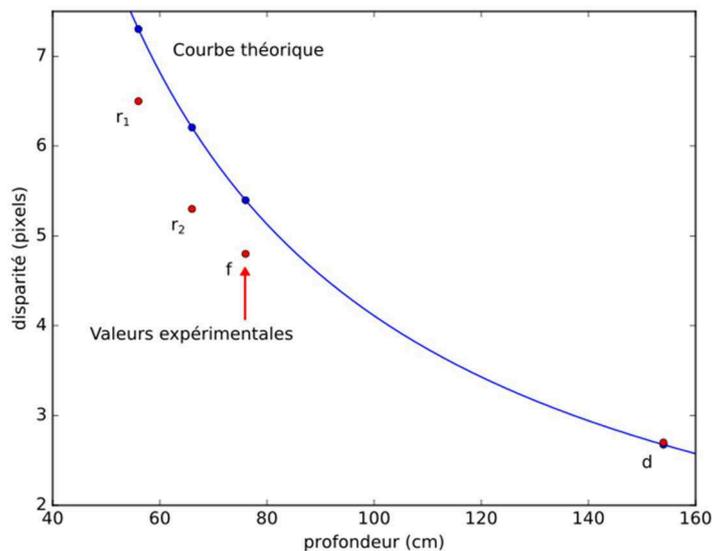


Figure 2.21 Courbes expérimentales et théoriques de la mesure de la disparité pour les quatre objets de la Figure 2.20 (3). En bleu, les valeurs théoriques et en rouge, les valeurs expérimentales. L'allure des courbes est globalement similaire.

Grandissement

A partir de l'expression de l'équation (2.53), nous allons comparer les valeurs que nous obtenons par calcul et celles obtenues de manière expérimentale.

Grandissement G	Objets			
	Roue Mulhouse 1	Roue Mulhouse 2	Capuchon du feutre	Carreau du damier
Expression littérale (éq. (2.53)) $G = -\frac{F \cdot F_{ml}}{(a - F) \cdot (b - F_{ml}) - a \cdot F}$	$8,7 \cdot 10^{-3}$	$7,4 \cdot 10^{-3}$	$6,4 \cdot 10^{-3}$	$3,2 \cdot 10^{-3}$
Expérimentation $\frac{\text{taille objet sur l'image} * t_p}{\text{taille réelle objet}}$	$7,4 \cdot 10^{-3}$	$6,5 \cdot 10^{-3}$	$5,3 \cdot 10^{-3}$	$2,7 \cdot 10^{-3}$

Tableau 2.6 Récapitulatif des résultats comparatifs entre les différentes méthodes d'obtention du grandissement G .

Comme expliqué, au début de la partie, quatre objets sont présents dans la scène : deux roues Mulhouse, un feutre et un échiquier. Afin de mesurer le grandissement pour ces quatre distances connues, la taille des deux roues de Mulhouse, la taille du bouchon du feutre est mesurée de même que celle d'un des carreaux du damier. Les mesures et les résultats sont regroupés dans le Tableau 2.6.

Les résultats expérimentaux confirment la tendance des valeurs théoriques pour les quatre distances. Les différences observées peuvent s'expliquer par la difficulté d'effectuer une mesure précise du paramètre b puisque les éléments optiques ont une épaisseur. Il est possible de conclure que la formulation théorique met en évidence le grandissement observé expérimentalement pour un système multi-vues plénoptique.

Ouverture numérique des mini-lentilles

A partir des données fournies par le constructeur et celles expérimentales, l'ouverture numérique d'une mini-lentille est égale à :

$$ON_{ml} = 1,6 - \frac{2,5 \cdot 10^{-3}}{42 \cdot 10^{-3}} = 1,54 \quad (2.56)$$

Comme l'explique Ren Ng [NLBD05] : lorsque l'ouverture numérique de l'objectif principal et celle des mini-lentilles sont égales, la solution optimale est obtenue pour ne pas avoir de chevauchement entre les différentes vues. Comme les deux ouvertures sont voisines dans notre configuration, on est proche de la solution optimale.

La limite de diffraction est calculée en utilisant l'équation (2.55) et est égale à

$$d = 2,44 * 550 \cdot 10^{-9} * 1,54 = 2,07 \cdot 10^{-6} \text{ m} \quad (2.57)$$

La taille des pixels t_p du capteur doit donc être supérieure à $2,07 \mu\text{m}$ (ici $t_p = 2,5 \mu\text{m}$) pour que le système ne présente pas de diffraction.

Dans cette partie, nous avons fourni quatre caractéristiques du système. Les comparaisons entre la théorie et l'expérimentation montrent que les définitions proposées sont correctes.

2.4.5.3 Images refocalisées

Même si le concept de refocalisation n'a pas encore été détaillé, il nous semble intéressant de confirmer que notre dispositif se classe bien parmi les dispositifs *light-field*, capables d'être utilisés pour générer des images refocalisées (cf. 1.5.1). Avec ce type de dispositif, il est possible de générer des images refocalisées dans différents plans de la scène, comme expliqué dans le chapitre 1 (cf. section 1.5.1). Dans le cas présenté, l'image d'origine correspond à l'image (3) de la Figure 2.20. Une pile d'images refocalisées sur différents objets est générée. Cependant, trois images refocalisées sont présentées sur la Figure 2.22 : la première sur la roue de Mulhouse la plus proche, la seconde sur la deuxième roue de Mulhouse et la troisième sur le damier au fond de l'image. La méthode de refocalisation utilisée diffère de celles présentées dans le chapitre 1 et est présentée dans le chapitre 4.

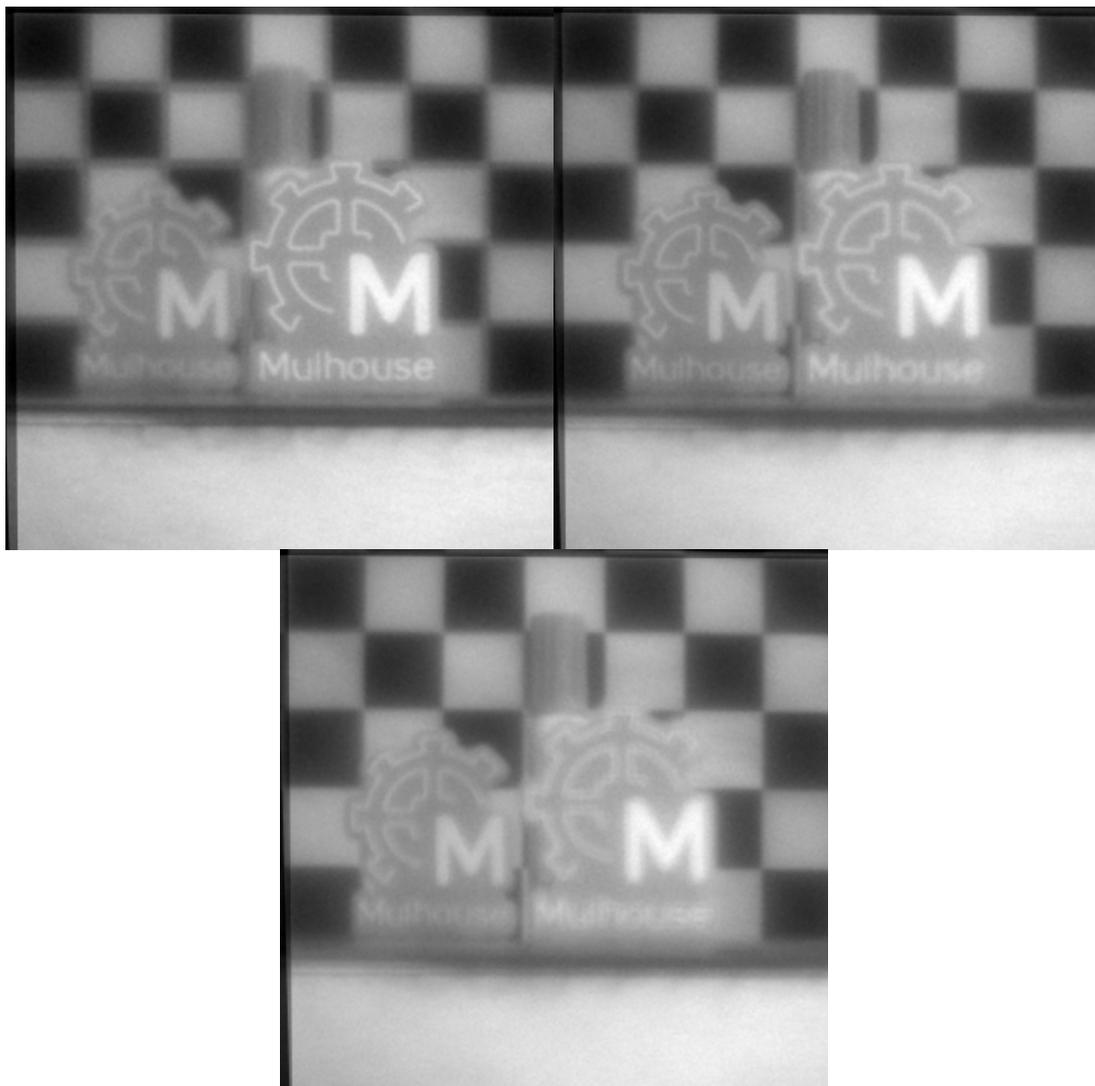


Figure 2.22 Illustration de la refocalisation. En haut à gauche, refocalisation sur le premier logo Mulhouse ; en haut à droite : refocalisation sur le deuxième logo Mulhouse et en bas au centre : image refocalisée sur le damier.

Nous avons donc caractérisé notre dispositif expérimental de caméra multi-vues plénoptique à l'aide de quatre critères, et nous avons généré des images refocalisées.

2.5 Conclusion

Dans ce chapitre, l'équivalence d'un point de vue optique des systèmes multi-caméras et plénoptiques est présentée. Puis, la conception de deux systèmes *light-field* développés au laboratoire est abordée : un système multi-vues et un système multi-caméras. Ils sont utilisés dans les chapitres 3 et 4 pour montrer les intérêts de notre modélisation et des applications mises en œuvre. Le dispositif multi-vues est utilisé dans le domaine de la vision industrielle tandis que le dispositif multi-caméras l'est dans le domaine de la perception automobile. Dans la dernière partie, afin de transformer une caméra 2D en une caméra 3D, un système multi-vues plénoptique a été envisagée. Il est composé d'un objectif principal, d'un capteur et de mini-lentilles, positionnées entre les deux précédents éléments. Un outil de simulation de tracé de rayons a été développé pour prédire le comportement des rayons lumineux à travers les éléments optiques de ce type de dispositif. De plus, des caractéristiques ont été définies en fonction de la géométrie de ce type de système : le taux de recouvrement, la parallaxe, le grandissement et l'ouverture numérique des mini-lentilles. Les résultats théoriques ont été confirmés par les résultats expérimentaux obtenus.

3 Modélisation par l'homographie variable et calibrage

3.1 Introduction

Dans le chapitre précédent, nous avons abordé l'étude géométrique des systèmes *light-field*. Afin de transformer ces systèmes en caméras de profondeur, il est nécessaire de les calibrer. Cependant, comme cela a été montré dans le chapitre 1 et plus particulièrement dans la section 1.4, la plupart des méthodes proposées sont complexes. Dans le cadre du prototype multi-vues développé au laboratoire, ou du système à quatre caméras que nous utilisons pour la vision automobile, nous proposons dans ce chapitre une méthode de calibrage originale et rapide. Cette méthode repose sur le concept de l'homographie variable, et s'avère être précise et facile à mettre en œuvre.

Dans [XCKF11, XCKF12], le MIPS a proposé l'extension de l'homographie variable, définie par Zhang et Greenspan [ZhGr07], à la mesure de hauteur de fibres émergentes sur des tissus en fibre de verre non tissés. Cette méthode fonctionne avec des échantillons progressant sur un tapis roulant. L'acquisition déclenchée de deux images successives était nécessaire pour effectuer la mesure 3D.

Dans ce travail, nous avons retenu les avantages de l'homographie variable pour des mesures le long de l'axe optique. En développant un système d'acquisition composé de quatre lentilles placées devant un unique capteur image, présenté dans le chapitre 2, section 2.3.1, le nombre d'acquisition nécessaire est réduit à une seule. On obtient ainsi quatre sous-images projetées sur un seul capteur CCD. Le système est considéré comme une caméra *light-field*. La formulation de l'homographie variable a été adaptée à cet appareil. Nous proposons une nouvelle démarche pour déterminer la profondeur à l'aide de caméras plénoptiques et en particulier avec notre dispositif multi-vues.

Afin de montrer que l'homographie variable est un principe extensible aux caméras plénoptiques, nous utilisons principalement le prototype multi-vues, présenté dans le Chapitre 3 partie 4.1. Cependant, elle peut être appliquée à d'autres systèmes *light-field* tels que les équipements multi-caméras, comme celui présentée dans le Chapitre 2 section 2.3.2.

Dans ce chapitre, le modèle sténopé est présenté ainsi que les principales méthodes de calibrage qui en découlent. Ensuite, les principes de l'homographie seront abordés, afin d'effectuer des projections entre un plan de la scène et une image, puis entre deux plans images de deux caméras différentes. Nous montrons par la suite comment l'homographie peut être étendue à la géométrie de notre caméra et comment l'homographie variable s'adapte à la mesure 3D. Cette approche permet de proposer une nouvelle méthode de calibrage, pour laquelle il ne faut connaître que l'homographie de référence contrairement aux autres méthodes nécessitant la connaissance des paramètres intrinsèques. Un critère de calibrage est proposé pour valider l'incertitude de mesures des homographies calculées. Les valeurs de ce critère seront notamment comparées à celles obtenues par la caméra multi-vues, développée pour la vision industrielle, et le système multi-caméras, dédié aux mesures supérieures à 2m. Ensuite,

dans le but de valider nos résultats, une comparaison entre notre méthode de calibrage et une approche consistant à calculer individuellement tous les paramètres intrinsèques [Zhan00], est proposée. Enfin, une brève analyse des résultats obtenus est effectuée et les perspectives concernant ces méthodes seront abordées.

3.2 Modèle sténopé

Sur une caméra, l'image est formée au moyen d'un objectif. Il s'agit d'un système optique complexe, composé de plusieurs lentilles, focalisant les rayons lumineux d'une source ponctuelle sur le plan image (capteur). Les objectifs servent à collecter les rayons entrants et à les focaliser vers le plan image. La scène peut être considérée comme une combinaison de sources ponctuelles. L'image formée sur le capteur correspond à une projection partielle de la scène faisant face à la caméra. La modélisation, la plus couramment employée, de cette projection est le modèle sténopé. En considérant que la scène est correctement focalisée sur le capteur, la formation des images peut être simplifiée en représentant uniquement le rayon passant par le centre optique de l'objectif. Cette modélisation correspond à une projection perspective. Elle permet de transformer un point 3D de l'espace en un point m de l'image. Sa décomposition peut être effectuée en 3 transformations simples consécutives : une première transformation entre le repère monde et celui de la caméra, une seconde entre le repère caméra et le repère capteur, et une troisième entre le repère capteur (en coordonnées métriques) et le repère image (en coordonnées pixels). Ces 3 transformations sont résumées par la Figure 3.1 et la Figure 3.2.

Dans cette partie, toutes les coordonnées seront exprimées en coordonnées homogènes, permettant d'exprimer les transformations avec une relation linéaire.

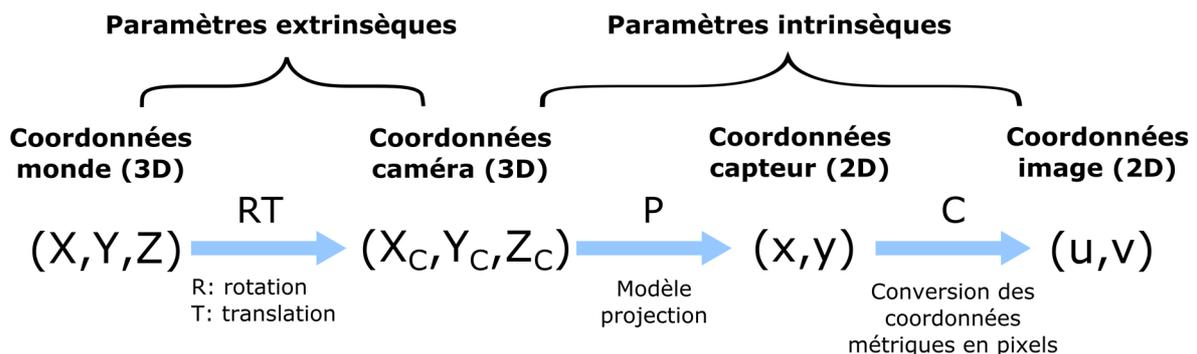


Figure 3.1 Les différentes coordonnées et leurs relations avec les paramètres intrinsèques et extrinsèques de la caméra.

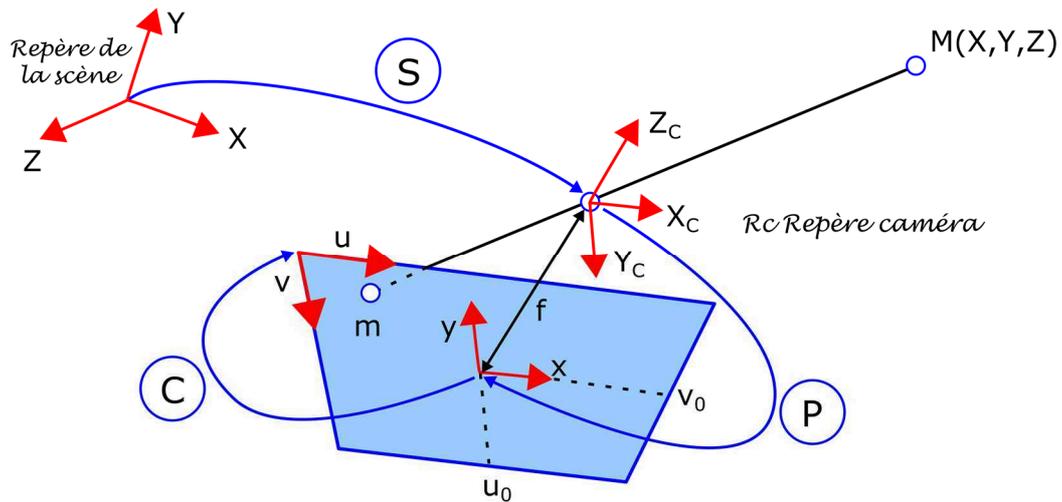


Figure 3.2 Projection d'un point M de l'espace dans le plan image de la caméra avec le modèle sténopé.

3.2.1 Transformation repère monde – repère caméra

Le repère monde est choisi arbitrairement et l'origine du repère caméra est située au centre optique de la caméra. La transformation P peut être décomposée en une rotation R et une translation T , ayant pour expressions :

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \text{ et } T = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.1)$$

La matrice de changement de repère S a pour expression :

$$S = \begin{pmatrix} R & T \\ 0_3^T & 1 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{12} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.2)$$

Les paramètres de cette transformation correspondent aux paramètres extrinsèques de la caméra.

3.2.2 Transformation repère caméra – repère capteur

La seconde transformation permet de passer du repère caméra au repère capteur. Il s'agit d'une projection perspective (représentée par une matrice 3x4 notée P) transformant un point 3D de coordonnées (X_c, Y_c, Z_c) en un point-image de coordonnées (x, y) en coordonnées métriques. Cette transformation a pour expression :

$$s \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = P \begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix} \quad (3.3)$$

Où f est la focale de l'objectif et s est le facteur d'homogénéisation ($s = Z$) pour pouvoir utiliser des coordonnées homogènes.

3.2.3 Transformation repère capteur – repère image

La troisième transformation permet de passer du repère capteur (unité métrique) au repère image (unité pixel). La relation entre un point image (u, v) (coordonnées en pixel) et un point image (x, y) (coordonnées métriques) est la suivante :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = A \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} k_u f & 0 & u_0 \\ 0 & k_v f & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (3.4)$$

où $\alpha_u = k_u f$, $\alpha_v = k_v f$. α_u et α_v représentent la distance focale exprimée en pixels (une fois horizontalement et une fois verticalement). u_0 et v_0 sont les coordonnées du point principal exprimées dans le repère image (en pixels).

Dans certains cas, un autre paramètre intrinsèque s (ou *skew factor*) est introduit pour prendre en compte la non orthogonalité des pixels. Donc la matrice A a pour expression :

$$A = \begin{pmatrix} \alpha_u & s & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.5)$$

3.2.4 Modèle sténopé complet

Le modèle sténopé complet [HaZi03] permet de passer d'un point de l'espace, de coordonnées (X, Y, Z) à un point image de coordonnées (u, v) (pixels). Il résulte d'une composition des trois transformations précédentes (comme on peut le voir sur la Figure 3.2) et a pour expression :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = APT \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (3.6)$$

Où A est la matrice des paramètres intrinsèques de la caméra, T la matrice des paramètres extrinsèques.

3.2.5 Prise en compte de la distorsion

Dans le modèle sténopé, on suppose que les lignes sont imagées comme des lignes. Cependant, dans la réalité, cette supposition n'est pas toujours exacte. Ainsi des lignes droites dans la réalité peuvent apparaître courbes sur l'image. Cet effet est dû à des phénomènes de distorsion. La plus importante est souvent la distorsion radiale. Elle est dépendante de la distance du point image au centre optique. Elle est de plus en plus importante à mesure que la distance focale (et la qualité) des lentilles décroît. Deux approches différentes existent pour modéliser cette distorsion : une approche paramétrique et une approche non paramétrique.

La première approche consiste à rajouter des termes au modèle sténopé. Ces ajouts rendent le modèle non-linéaire. Elle est inspirée des modèles utilisés pour corriger les aberrations géométriques des systèmes centrés et correspondent à différents types de distorsion : radiale, prismatique et de décentrage [Garc01, WCHO92, ZeOt96]. Elle se caractérise par l'ajout d'une transformation entre le passage du repère caméra au repère image. Elle correspond au calcul des coordonnées corrigées dans le repère caméra. Elle peut être modélisée comme suit :

$$\begin{aligned}\bar{m}_r &= \mathcal{D}(m_r) = m_r + \delta(m_r) \\ &= m_r + \overbrace{\delta_r(m_r) + \delta_d(m_r) + \delta_p(m_r)}\end{aligned}\quad (3.7)$$

Un modèle, désigné sous le nom de R3D1P1, a été proposé par Ricolfe-Viala et al [RiSá10]. Il est suffisant pour la majorité des objectifs, dont la focale est supérieure à 5 mm. Cependant, dans la plupart des cas, seul un modèle radial (d'ordre 1 à 3), le modèle R3, est utilisé. Le modèle de la caméra est non linéaire et peut s'écrire sous la forme d'une fonction vectorielle.

La correction de la distorsion peut s'effectuer de plusieurs manières, dont certaines sont présentées dans la section « The perspective camera inverse model » dans la thèse de J. Sola [Sola07]. Un autre modèle de la distorsion radiale et de sa correction sont détaillées dans le livre de Hartley [HaZi03].

L'approche non paramétrique [CGSM04] consiste à modéliser la distorsion de manière non-paramétrique au moyen de fonctions splines. Il s'agit par conséquent d'une modélisation purement mathématique. Dans ce cas, la fonction de correction de la distorsion correspond à l'approximation des composantes horizontales et verticales du champ de correction de distorsion par deux surfaces splines S_x et S_y .

3.2.6 Calibrage d'une caméra

Ce sujet a été très étudié dans la littérature. Dans ce chapitre, une seule méthode est testée en comparaison de celle que nous avons proposée. Dans la majorité des cas, le calibrage sert à déterminer les paramètres intrinsèques et/ou extrinsèques de la caméra (ou des caméras). L'un des difficultés réside

dans la dissociation des paramètres extrinsèques et intrinsèques d'un point de vue calculatoire. Ainsi, dans les processus standards, ces deux types de paramètres sont déterminés simultanément.

Les méthodes peuvent être classées suivant plusieurs critères [SaAB02] :

- Calibration linéaire ou non linéaire (prise en compte ou non de la distorsion),
- Recherche des paramètres intrinsèques et/ou extrinsèques,
- Calibration implicite ou explicite (explicite : des paramètres physiques sont recherchés),
- Méthodes utilisant des mires 2D ou 3D,
- Méthodes d'auto-calibration s'appuyant sur des propriétés, comme la perspective et les points de fuite ou d'autres types de propriétés géométriques des scènes.

Pour calibrer avec la méthode de Zhang [Zhan00], il est nécessaire d'acquérir une ou plusieurs images d'une mire de calibrage (sa structure et ses dimensions sont parfaitement connues). Une mire de calibrage est constituée d'éléments caractéristiques tels que des points, des cercles ou bien des carrés. La position de ces cibles est connue précisément dans un repère associé à la mire. L'une des étapes du calibrage correspond à l'extraction des points caractéristiques de la mire par traitement d'images. Ensuite, une phase de mise en correspondance pour chaque élément dans l'image est nécessaire pour déterminer l'élément 3D dont elle est la projection dans le plan image. Il s'agit d'un processus semi-automatique. Enfin, en utilisant des relations connues entre les paramètres intrinsèques et extrinsèques, il est possible de calculer ces paramètres. Des méthodes ont notamment été proposées par Tsai et Zhang [Zhan00].

Dans la partie 3.5.1, la méthode de Zhang, l'une des méthodes de référence est présentée en détails. Elle permet de déterminer les paramètres intrinsèques et extrinsèques d'une caméra 2D. Elle est utilisée dans la partie 3.6 de ce chapitre afin d'effectuer une comparaison avec notre méthode de calibrage.

3.3 Homographies

3.3.1 Projection d'un plan de l'espace sur le plan image d'une caméra : homographie

Dans cette partie, on passe de la projection d'un point à la projection d'un plan π de l'espace sur un plan image. Le modèle sténopé complet (équation (3.6)) peut s'écrire sans la dimension Z , si ce plan est associé au repère de la scène (alors $Z=0$). Avec cette formulation, on met en évidence que tout point du plan π peut être projeté sur le plan image à l'aide d'une matrice 3×3 , appelée matrice homographique, notée H_π .

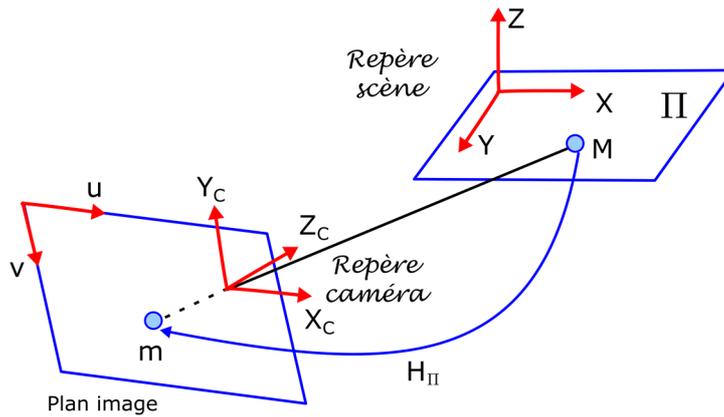


Figure 3.3 Projection de tout point du plan π sur le plan image avec l'homographie H_π

Tout point du plan π , de coordonnées $(X, Y, Z=0)$ a pour projection le point-image de coordonnées (u, v) , et est calculé suivant l'expression suivante :

$$s. \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} = H_\pi \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \quad (3.8)$$

3.3.2 Homographies entre deux plans image de deux caméras

Si on considère deux caméras et les projections m_1 et m_2 d'un point M du plan π sur leurs deux plans image, il est aisé de montrer que ces projections sur les deux plans images sont aussi reliées au moyen d'une homographie :

$$\begin{cases} m_1 = H_{\pi 1} \cdot M \\ m_2 = H_{\pi 2} \cdot M \end{cases} \Leftrightarrow M = H_{\pi 2}^{-1} \cdot m_2 \Leftrightarrow m_1 = H_{\pi 1} \cdot H_{\pi 2}^{-1} \cdot m_2 \quad (3.9)$$

Cette relation a de nombreuses applications en vision telles que la correction de perspective, la rectification d'image, le mosaïquage d'image ou la réalité virtuelle.

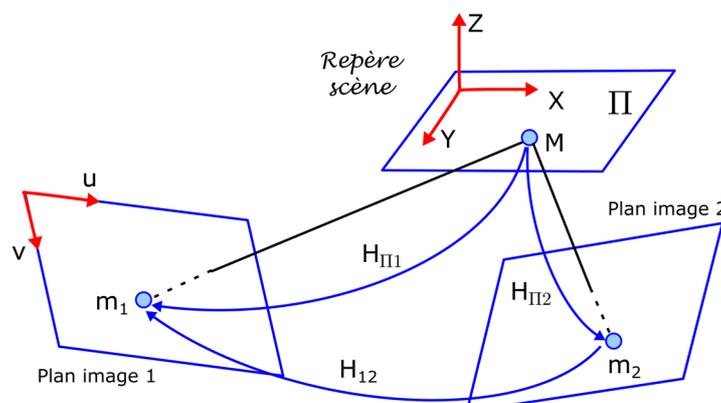


Figure 3.4 Relation homographique entre 2 points image sur deux plans image distincts d'un unique point appartenant au plan π de la scène.

3.4 Homographie variable et mesures de profondeur à partir de la disparité

Dans cette partie, nous allons aborder le principe de l'homographie variable ainsi que son extension aux caméras *light-field* : son utilisation pour la modélisation d'une caméra plénoptique et pour les mesures de profondeur.

3.4.1. Principe de l'homographie variable entre images

On considère les points $P_c (X_c, Y_c, Z_c)$ et $P (X, Y, Z)$ appartenant respectivement aux plans parallèles C et Z. Ils ont les mêmes coordonnées en X et en Y, soit $X_c=X$ et $Y_c=Y$.

La projection du point P_c sur le plan image peut s'écrire à l'aide de la matrice intrinsèque :

$$Z_c \begin{pmatrix} u_{1c} \\ v_{1c} \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_{u1} & 0 & u_{01} \\ 0 & \alpha_{v1} & v_{01} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} \quad (3.10)$$

Les coordonnées du point p_{1c} s'écrivent donc :

$$\begin{cases} u_{1c} = \alpha_{u1} \frac{X_c}{Z_c} + u_{01} \\ v_{1c} = \alpha_{v1} \frac{Y_c}{Z_c} + v_{01} \end{cases} \quad (3.11)$$

De même, les coordonnées du point p_1 ont pour expression :

$$\begin{cases} u_1 = \alpha_{u1} \frac{X}{Z} + u_{01} \\ v_1 = \alpha_{v1} \frac{Y}{Z} + v_{01} \end{cases} \quad (3.12)$$

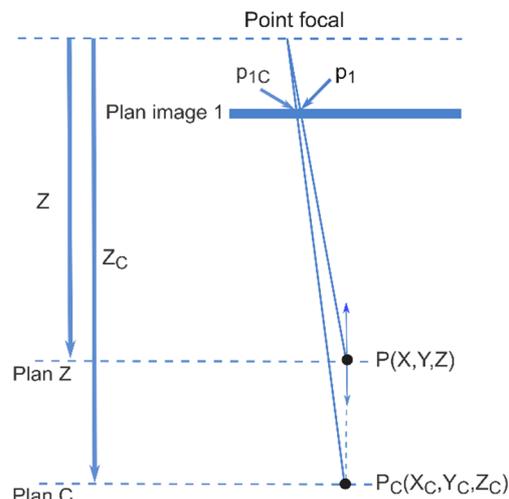


Figure 3.5 Homographie variable : projection des points P_c et P sur le plan image

Sachant que $X_c=X$ et $Y_c=Y$, les deux équations précédentes permettent de relier les coordonnées des projections p_{1c} et p_1 . La relation, ainsi mise en évidence, fait intervenir un facteur k , correspondant au rapport des distances Z_c et Z . Elle relie ces deux projections au moyen d'une matrice K_1 (indice 1 pour image 1) :

$$\begin{cases} u_1 = (u_{c1} - u_{01})k + u_{01} \\ v_1 = (v_{c1} - v_{01})k + v_{01} \\ k = \frac{Z_c}{Z} \end{cases} \Leftrightarrow p_1 = \begin{pmatrix} k & 0 & (1-k)u_{01} \\ 0 & k & (1-k)v_{01} \\ 0 & 0 & 1 \end{pmatrix} p_{1c} = K_1(k) \cdot p_{1c} \quad (3.13)$$

Ainsi, la projection p_1 d'un point P aligné en X et en Y avec le point P_c peut être calculée à partir de la projection p_{1c} du point P_c .

Si on considère une seconde image, notée 2, on a les mêmes relations pour la seconde caméra :

$$p_2 = \begin{pmatrix} k & 0 & (1-k)u_{02} \\ 0 & k & (1-k)v_{02} \\ 0 & 0 & 1 \end{pmatrix} \cdot p_{2c} = K_2(k) \cdot p_{2c} \quad (3.14)$$

Pour tout point appartenant au plan C , l'homographie H_c relie les points p_{2c} et p_{1c} :

$$p_{1c} = H_c \cdot p_{2c} \quad (3.15)$$

On peut donc écrire la relation entre les projections p_1 et p_2 comme suit :

$$p_1 = K_1(k) \cdot p_{1c} = K_1(k) \cdot H_c \cdot K_2(k)^{-1} \cdot p_2 = H(k) \cdot p_2 \quad (3.16)$$

Cette dernière expression permet de déduire l'homographie H , se formulant en fonction de l'homographie H_c et des matrices K_1 et K_2 , et du facteur k :

$$H(k) = K_1(k) \cdot H_c \cdot K_2(k)^{-1} \quad (3.17)$$

3.4.2. Modélisation d'un système multi-vues

La Figure 3.6 montre notre caméra, formée de quatre mini-lentilles et d'un unique plan image, et à droite de manière schématique, seuls deux vues sont représentées.

La Figure 3.6 à droite correspond à une section 2D représentant seulement deux projections voisines : donnant les sous-images l_i et l_{i+1} (dans ce cas $i=1$). La valeur est comprise entre 0 et 3, elle correspond au nombre de sous-images acquises par la caméra. Avec cette configuration, chaque sous-image a 3 voisines possibles. Par exemple, l'index i représente la sous-image l_2 alors que l'index $i+1$ peut correspondre à l'image l_3 , l_1 ou l_0 . Ce schéma permet de voir notre caméra comme une caméra basée sur deux systèmes de stéréovision parallèles. Ainsi, il est possible d'utiliser la triangulation pour calculer la

profondeur d'un point P dans la scène, quand les paramètres intrinsèques et la distorsion sont connus. Si ces paramètres sont identiques pour chaque caméra, nous sommes dans la situation idéale. Cependant, c'est rarement le cas dans la réalité et il est nécessaire que le calibrage soit précis afin de pouvoir déterminer les valeurs numériques des centres optiques et des distances focales. Ces paramètres sont utiles à la reprojection géométrique, permettant d'obtenir la profondeur par intersection de deux rayons provenant chacun des deux pixels associés (chacun se trouvant sur une vue différente). Par exemple, sur la Figure 3.6, les paramètres f_1, f_2, O_1 et O_2 doivent être connus pour retrouver la profondeur du point P à partir des rayons $(O_1 p_1)$ et $(O_2 p_2)$. En conséquence, nous avons décidé de tirer profit de l'homographie variable pour modéliser notre système en développant une nouvelle approche basée sur ses principes.

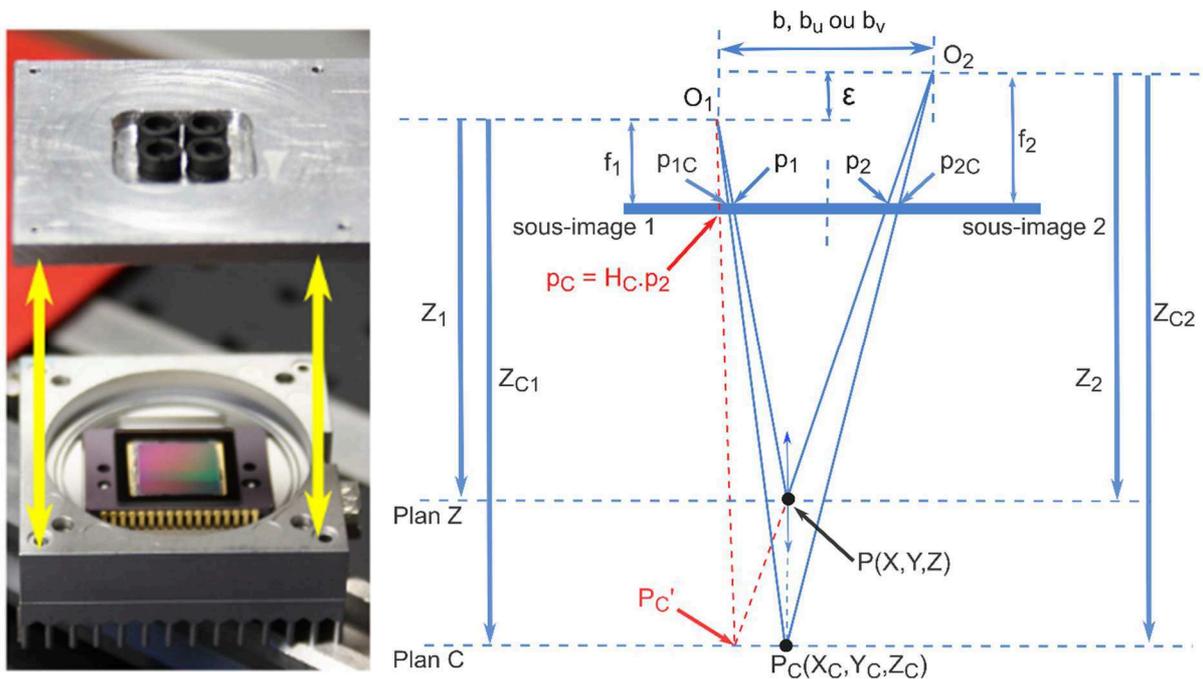


Figure 3.6 Illustrations du montage à 4 vues. A gauche, vue éclatée de ce montage. A droite, représentations schématique où seuls deux sous-images sont représentées.

3.4.3. Généralisation de l'homographie variable au dispositif multi-vues

L'homographie variable présente des propriétés intéressantes pour sa généralisation à un dispositif multi-vues. Considérons des homographies H_c et H définies pour les plans parallèles C et Z et vérifiant $p_{1C} = H_c \cdot p_{2C}$ et $p_1 = H \cdot p_2$:

$$H(k) = K_i H_c K_{i+1}^{-1} \quad (3.18)$$

$$\text{où } K_n = \begin{pmatrix} k_n & 0 & (1 - k_n)u_{0n} \\ 0 & k_n & (1 - k_n)v_{0n} \\ 0 & 0 & 1 \end{pmatrix} \quad (3.19)$$

avec $n = i$ ou $n = i+1$ et où u_{0n} et v_{0n} désignent les coordonnées des centres optiques.

Le principal paramètre de la matrice K_n est le ratio de distances k_n défini par $k_n = Z_{C_n}/Z_n$, donnant la possibilité de calculer H à partir de H_c . En faisant varier k autour de valeurs réalistes et en utilisant une méthode comparant les mises en correspondance, il est possible de trouver des correspondances fiables pour lesquelles $k = \text{argmax}(\text{Matching}(p_1(k), p_2))$. Nous proposons l'extension de cette méthode, développant ainsi les expressions des homographies H_c et H pour chacune des quatre mini-lentilles de la caméra. Le but étant de simplifier l'expression de la reprojection utilisée pour calculer la profondeur et d'inclure facilement dans notre modèle tous les paramètres intrinsèques décrits dans la Figure 3.6.

3.4.4. Expressions des homographies H_c et H

Le plan C est considéré comme le plan de référence et sert à calibrer le système. Notre caméra étant formée de quatre mini-lentilles, le cas où les lentilles n'ont pas la même distance focale a été étudié. L'homographie H_c est déterminée en supposant que l'homographie est une combinaison de paramètres et de matrices de translation, avec comme définition [HaZi03] :

$$H_c = \begin{pmatrix} \frac{f_1 Z_{C2}}{f_2 Z_{C1}} & 0 & u_{O1} + \frac{b_u f_1}{Z_{C1}} - \frac{f_1 Z_{C2}}{f_2 Z_{C1}} u_{O2} \\ 0 & \frac{f_1 Z_{C2}}{f_2 Z_{C1}} & v_{O1} + \frac{b_v f_1}{Z_{C1}} - \frac{f_1 Z_{C2}}{f_2 Z_{C1}} v_{O2} \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \alpha & 0 & u_{O1} + \frac{b_u f_1}{Z_{C1}} - \alpha u_{O2} \\ 0 & \alpha & v_{O1} + \frac{b_v f_1}{Z_{C1}} - \alpha v_{O2} \\ 0 & 0 & 1 \end{pmatrix} \quad (3.20)$$

Où b_u et b_v sont respectivement les distances horizontales et verticales entre les deux centres optiques.

En considérant que $Z_{C2} = Z_{C1} + \varepsilon \approx Z_{C1}$ à la distance de travail, H_c a été simplifiée dans l'équation (3.20) par l'introduction du ratio de distances focales $\alpha = f_1/f_2$. Les équations (3.18) et (3.20) permettent d'obtenir une expression simplifiée de l'homographie H , représentant la transformation géométrique de n'importe quel point P, localisé dans le plan virtuel Z, parallèle au plan de référence C :

$$H = \begin{pmatrix} \alpha & 0 & u_{O1} + k_1 \frac{b_u f_1}{Z_{C1}} - \alpha u_{O2} \\ 0 & \alpha & v_{O1} + k_1 \frac{b_v f_1}{Z_{C1}} - \alpha v_{O2} \\ 0 & 0 & 1 \end{pmatrix} \quad (3.21)$$

3.4.5. Calcul de la profondeur Z

A partir des homographies H_c et H , la profondeur de n'importe quel point P peut être calculée, en utilisant la définition de la disparité d comme étant la différence suivante : $d = H.p_2 - H_c.p_2$. En considérant que $p_1 = H.p_2$ et à partir des expressions (3.20) et (3.21), d peut être exprimé comme suit :

$$d = H \cdot p_2 - H_C \cdot p_2 = \text{dist}(p_1, p_C) = \left(\left(\underbrace{\left((k_1 - 1) \frac{\overbrace{b_u f_1}^{C_u}}{Z_{C1}} \right)^2}_{d_u} \right) + \left(\underbrace{\left((k_1 - 1) \frac{\overbrace{b_v f_1}^{C_v}}{Z_{C1}} \right)^2}_{d_v} \right) \right)^{1/2} \quad (3.22)$$

En supposant que $k_i = Z_{Ci}/Z_i$ et $b = \sqrt{b_u^2 + b_v^2}$, les relations entre la profondeur Z_1 et la disparité d sont finalement données par :

$$Z_1 = Z_{C1} \frac{\sqrt{C_u^2 + C_v^2}}{d \cdot Z_{C1} + \sqrt{C_u^2 + C_v^2}} = \frac{b f_1}{d + \frac{b f_1}{Z_{C1}}} \quad \text{ou} \quad d = \frac{b f_1}{Z_1} - \frac{b f_1}{Z_{C1}} \quad (3.23)$$

La définition de la disparité est légèrement différente de celle utilisée avec la triangulation directe entre deux points de vue. Dans la partie suivante, nous étudierons les différences et les avantages de cette définition pour les mesures 3D, spécifiquement dans un contexte de vision industrielle.

3.4.6. Comparaison avec la triangulation

Quand deux caméras sont séparées par une translation latérale sans rotation, la méthode de triangulation permet d'obtenir la profondeur relativement aux coordonnées des points p_1 et p_2 . Ainsi à partir de la Figure 3.6, la profondeur Z_1 peut être formulée par :

$$Z_1 = \frac{b + \frac{\varepsilon}{f_2} (p_{O_2} - p_2)}{\frac{(p_2 - p_{O_2})}{f_2} - \frac{(p_1 - p_{O_1})}{f_1}} \quad (3.24)$$

Cette relation est habituellement simplifiée en supposant que $f_1 = f_2$ et que les centres optiques O_1 et O_2 ont des coordonnées identiques dans leur système respectif de coordonnées correspondant à chaque sous-image. Ainsi, la profondeur peut être exprimée comme suit :

$$Z_1 = \frac{b f_1}{d} \quad \text{avec} \quad d = p_2 - p_1 \quad (3.25)$$

En comparant cette expression avec celle obtenue à partir de l'homographie variable, on constate qu'il y a une similarité évidente, et que l'homographie variable génère une translation de la disparité d suivant le terme $b f_1 / Z_{C1}$.

Sur la Figure 3.7, nous comparons les expressions (3.23) et (3.25). Les courbes sont tracées en utilisant les dimensions réelles de notre caméra : une taille de pixels de 7,4 μm , des centres optiques espacés de 9 mm et les distances focales f_1 et f_2 égales à 8 mm. La courbe correspondant à la triangulation est tracée en rouge tandis les autres obtenues avec l'équation (3.23) sont tracées en bleu avec plusieurs valeurs de distance de référence Z_{C1} . En fonction de la valeur de Z_{C1} , les courbes bleues sont décalées

vers la gauche. Cette translation présente des propriétés intéressantes quand la plage de mesure est prédictible : il devient possible de travailler dans la même plage de disparité quel que soit la profondeur de travail, tout en choisissant correctement la distance de référence Z_{C1} .

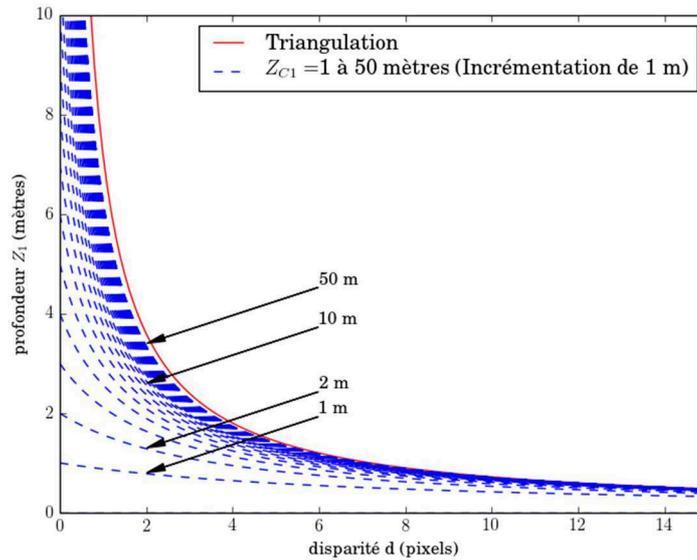


Figure 3.7 Profondeurs Z_1 en fonction de la disparité d . Comparaison du calcul de la disparité à partir de la triangulation et de l'homographie variable.

Il est assez aisé de démontrer que l'erreur de profondeur ∂Z_1 reste identique pour les deux approches, signifiant que la translation n'influence pas la précision de mesure. En reprenant les équations (3.23) et (3.25),

$$d = \frac{bf_1}{Z_1} - \frac{bf_1}{Z_{C1}} \quad (3.23) \quad \text{et} \quad d = \frac{bf_1}{Z_1} \quad (3.25)$$

Sachant que bf_1/Z_{C1} est une constante, alors les dérivées de la disparité en fonction de Z_1 sont donc égales et ont pour expression :

$$\frac{\partial d}{\partial Z_1} = -\frac{bf_1}{Z_1^2} \rightarrow \partial Z_1 = -\frac{Z_1^2}{bf_1} \partial d \quad (3.26)$$

Il y a des avantages à utiliser l'homographie variable pour travailler dans une plage de disparité déterminée. Le même système et le même algorithme peuvent être utilisés pour un large nombre de distances de travail, allant d'applications macroscopiques à des échelles plus grandes. Par exemple, si Z_{C1} est égale à 7 cm, la profondeur Z_1 est comprise entre 7 cm et 5,8 cm pour un intervalle de disparité d de [0 ; 20] pixels. Pour une inspection dans un contexte de vision industrielle 3D, quand la distance Z_{C1} est fixée à 90 cm, les profondeurs mesurées sont comprises entre 90 cm et 26 cm, tout en conservant le même intervalle de disparité. On peut noter que pour ces différentes échelles de mesure, on a utilisé la même caméra. Intervertir les différentes échelles de mesure nécessite uniquement l'ajustement de la distance focale des lentilles. Ce dernier point est important, parce que dans ce cas la précision de la mesure va nécessiter de recalculer tous les paramètres intrinsèques utilisés dans l'équation (3.24)

quand la méthode de triangulation est appliquée. La partie suivante va permettre d'expliquer comment l'homographie variable facilite le calibrage.

3.5. Calibrage des systèmes multi-vues, modélisés par des homographies variables

L'intérêt de l'homographie variable, est d'avoir une approche implicite du calibrage. Tous les paramètres intrinsèques n'ont besoin d'être déterminés individuellement comme pour la méthode de Zhang, présentée dans la partie suivante. En utilisant l'homographie variable, le calibrage revient à déterminer les paramètres nécessaires au calcul de H (éq. (3.21)). Il suffit donc de calculer Z_{C1} , bf_1 et H_C , afin de disposer de l'ensemble des termes permettant de calculer H , pour n'importe quelle distance de travail $Z=k.Z_{C1}$. Puis, une méthode perfectionnée, prenant en compte la distorsion, est proposée.

3.5.1 Calibrage avec la méthode de Zhang

Zhang [Zhan00] a proposé une méthode permettant de calibrer des caméras en se basant sur un plan simple plutôt que sur une plateforme de calibrage tridimensionnelle. Ainsi, les exigences de calibrage sont réduites au minimum. Elle a mené à des résultats précis. Elle requiert plusieurs images d'un motif plan (le plus souvent un échiquier) avec des orientations différentes. Ces dernières ne doivent jamais correspondre à de simples translations les unes par rapport aux autres. Cette technique est simple à mettre en œuvre en sachant qu'elle nécessite de fixer le motif sur une surface plane.

Cette méthode est composée de plusieurs étapes :

1. Estimation de l'homographie pour chacune des vues,
2. Détermination des paramètres intrinsèques de la caméra,
3. Détermination des paramètres extrinsèques,
4. Estimation de la distorsion radiale de la lentille,
5. Raffinement des paramètres.

Seules les étapes 2 et 3 sont présentées en détails. Plusieurs méthodes peuvent être utilisées pour l'estimation de l'homographie telle que la Transformation Linéaire Directe [Dubr09] ou en effectuant une minimisation non-linéaire réalisée avec l'algorithme de Levenberg-Marquardt [Zhan00].

Deux contraintes concernant les paramètres intrinsèques ont été définies en utilisant la relation homographique entre l'homographie H et la matrice intrinsèque A . La méthode proposée permet aussi de calculer les paramètres extrinsèques et les paramètres de distorsion. Dans le cadre de notre comparaison, seuls les calculs des paramètres intrinsèques et extrinsèques nous intéressent.

L'homographie dépend de la rotation, exprimée par les deux premières colonnes r_1 et r_2 , d'une matrice de rotation 3x3 et de la translation T entre le repère caméra et le repère monde et est définie comme suit :

$$H = \lambda \cdot A(r_1 \ r_2 \ T) \quad (3.27)$$

A partir de cette équation, on obtient :

$$H = A(r_1 \ r_2 \ T) = \lambda(h_1 \ h_2 \ h_3) \quad (3.28)$$

où les vecteurs h_x et r_x représentent les colonnes de chaque matrice.

Ensuite, on obtient,

$$r_1 = \lambda^{-1}A^{-1}h_1 \quad (3.29)$$

$$r_2 = \lambda^{-1}A^{-1}h_2 \quad (3.30)$$

Comme r_1 est orthogonal à r_2 , on obtient :

$$r_1^T r_2 = 0 \Leftrightarrow (\lambda^{-1}A^{-1}h_1)^T (\lambda^{-1}A^{-1}h_2) = 0 \Leftrightarrow h_1^T (A^{-1})^T A^{-1} h_2 = 0 \quad (3.31)$$

Cette équation permet de définir B , composé des paramètres intrinsèques et donne la première contrainte sur B qui peut être formulée avec la propriété suivante des vecteurs orthogonaux :

$$\begin{aligned} r_1^T \cdot r_1 = r_2^T \cdot r_2 = 1 &\Rightarrow (A^{-1}h_1)^T \cdot (A^{-1}h_1) = (A^{-1}h_2)^T \cdot (A^{-1}h_2) \\ &\Rightarrow (h_1)^T (A^{-1})^T A^{-1} h_1 = (h_2)^T (A^{-1})^T A^{-1} h_2 \end{aligned} \quad (3.32)$$

Ainsi, une deuxième contrainte sur B peut être définie :

$$h_1^T B h_1 = h_2^T B h_2 \quad (3.33)$$

B est donc défini comme le produit de $(A^{-1})^T A^{-1}$. Ces deux conditions sont suffisantes pour déterminer B à partir d'un jeu d'homographies. Les paramètres intrinsèques peuvent être déduits à partir des relations suivantes (les éléments des matrices sont notés b_{ij}) :

$$u_0 = -\frac{b_{12}f_x^2 f_y}{\lambda} \quad (3.34)$$

$$v_0 = \frac{b_{12}b_{13} - b_{11}b_{23}}{b_{11}b_{22} - b_{12}^2} \quad (3.35)$$

$$\alpha_x = \sqrt{\lambda/b_{11}} \quad (3.36)$$

$$\alpha_y = \sqrt{\lambda b_{11} / (b_{11} b_{22} - b_{12}^2)} \quad (3.37)$$

Après avoir calculé les paramètres intrinsèques, les paramètres extrinsèques sont calculés. Comme la matrice intrinsèque A est connue, ces paramètres peuvent être aisément calculés. A partir de l'équation (3.27), on a donc :

$$r_1 = \lambda A^{-1} h_1 \quad (3.38)$$

$$r_2 = \lambda A^{-1} h_2 \quad (3.39)$$

$$r_3 = r_1 \times r_2 \quad (3.40)$$

$$t = \lambda A^{-1} h_3 \quad (3.41)$$

Ainsi la matrice 3x3 de rotation R est définie comme suit :

$$R = [r_1, r_2, r_3] \quad (3.42)$$

Après avoir présenté la méthode de calibrage proposé par Zhang, qui est utilisé dans la partie 3.6 pour effectuer une comparaison, le principe des homographies est expliqué.

3.5.2. Calibrage simple

En pratique, toutes les lentilles ne sont jamais parfaitement alignées dans les caméras multi-lentilles ou plénoptiques ; les distances focales ne sont pas exactement égales [DaPW13]. Ce qui est spécifiquement vrai pour notre prototype et génère quelques distorsions indésirables affectant les mesures de profondeurs. En pratique, les imperfections sont rarement prises en compte. La littérature donne généralement l'expression (3.25) comme méthode de référence pour calculer la profondeur au lieu de fournir l'expression complète correspondant à l'équation (3.24). Considérant un système multi-vues, la détermination complète de ces paramètres intrinsèques peut devenir une tâche difficile, requérant de multiples acquisitions [Zhan00], mais indispensables pour recueillir des mesures précises. Un autre des principaux avantages de l'homographie variable est que la matrice homographique H_c inclut tous les paramètres intrinsèques entre les vues de deux sous-images. En projetant les points de mesures de la sous-image 2 vers la sous-image 1 avec H_c , le nombre de paramètres requis par la mesure 3D est ainsi réduit et est facile à déterminer via un simple processus de calibrage.

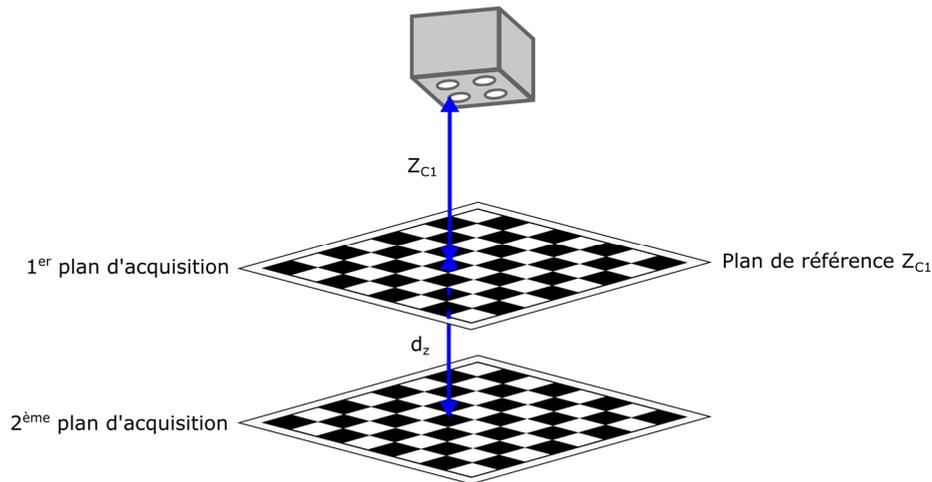


Figure 3.8 Calibrage avec deux acquisitions : la première dans le plan de référence et la seconde dans un autre plan situé à une distance d_z connue du plan de référence.

Les paramètres à retrouver durant le calibrage sont uniquement : Z_{C1} , H_c et la constante bf_1 . La méthode de calibrage proposée est aisée et robuste. Elle est réalisée en utilisant un motif de damier de référence afin de détecter les coins en effectuant deux acquisitions successives. La première acquisition est effectuée à la distance de référence Z_{C1} , la seconde l'est à une profondeur $Z_1 = Z_{C1} + d_z$. Les coins de la première acquisition sont utilisés pour calculer toutes les combinaisons existantes d'homographies H_c entre les sous-images. Pour chaque acquisition, la distance entre le premier et le dernier coin détecté sur l'échiquier est calculée (Figure 3.8). On note d_c la distance pour la première acquisition et d_1 la distance pour la seconde. Z_{C1} est ainsi obtenue par triangulation et est formulée comme suit :

$$Z_{C1} = \frac{d_z \cdot d_1}{d_1 - d_{C1}} \quad (3.43)$$

Finalement, durant la seconde acquisition, Z_{C1} , H_c , d_z et les distances d calculées pour les coins correspondants p_1 et p_2 , et en utilisant l'équation (3.23), permettent de déterminer la meilleure valeur de bf_1 . Pour conclure, l'intérêt de ce calibrage réside dans sa simplicité, parce qu'il est facile de réaliser deux acquisitions plutôt que de déterminer tous les paramètres intrinsèques pour plusieurs images, comme c'est fait habituellement dans les méthodes standards connues [Zhan00].

3.5.3. Calibrage amélioré

Pour notre caméra multi-vues, des lentilles à bas coût ont été utilisées. Elles présentent un coefficient de distorsion proche de 1,5 % (valeur extraite de la documentation des lentilles). Pour améliorer la précision des mesures, on propose de tenir compte de la distorsion dans notre modèle de calibrage. L'utilisation d'une approche standard, comme les corrections basées sur des modèles de distorsions radiales, permet de déterminer la position précise de chaque centre optique des sous-images. Comme les paramètres intrinsèques sont combinés dans l'homographie H_c , ils ne sont pas immédiatement accessibles pendant cette étape de notre processus de calibrage. Nous proposons par conséquent une

approche alternative. On considère que les effets de la distorsion sont comparables à un grandissement non uniforme de l'image au niveau du plan image. Cette approche permet d'ajuster le calcul de la profondeur Z_1 en fonction de la localisation spatiale des correspondants dans chaque sous-image, ce qui est possible en utilisant une relation proportionnelle existante entre le grandissement, la distance focale et le paramètre bf_1 . Nous proposons une modélisation alternative assez similaire aux travaux menés au laboratoire [BSCB11] ou à des approches polynomiales présentées par Bugarin et al. [BBHL15], pour la détermination de matrices fondamentales. La correction est effectuée en utilisant l'opérateur Γ , qui transforme le vecteur représentant le point P de coordonnées (u, v, 1) en un vecteur de coordonnées modifiées \tilde{P} [BaDa06]:

$$P(u, v, 1)^T \rightarrow \tilde{P} = \Gamma(P, P) = (u^2, v^2, uv, u, v, 1)^T \quad (3.44)$$

En utilisant \tilde{p}_2 comme emplacement de référence, bf_1 est calculé grâce à des fonctions quadratiques définies par :

$$bf_1 = (a \ b \ c \ d \ e \ f) \cdot \tilde{p}_2 \quad (3.45)$$

Les coins collectés, pendant ce calibrage, constituent une base de données pour calculer les coefficients polynomiaux en utilisant une méthode des moindres carrés. Cette étape ne requiert pas d'autre manipulation.

3.5.4. Précision des mesures

En utilisant un banc de mesures, des résultats expérimentaux ont pu être obtenus en reculant un échiquier par pas de 1 cm entre deux acquisitions. Le contexte expérimental et les mesures sont illustrés par la Figure 3.9. La profondeur ainsi mesurée est calculée pour chaque acquisition et sur le premier coin détecté. Elle est utilisée pour tracer la courbe représentant la profondeur en fonction de la disparité.

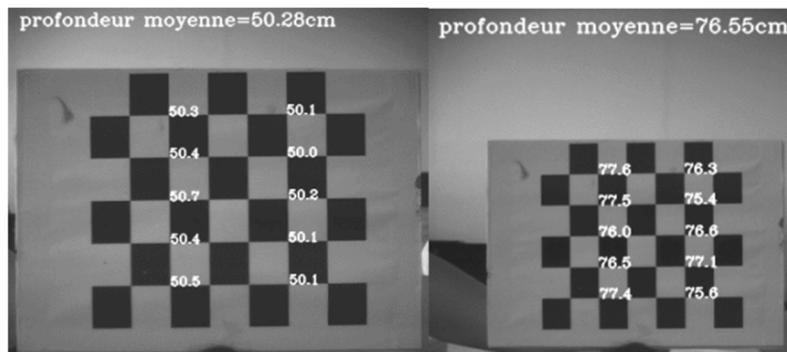


Figure 3.9 Illustration des mesures expérimentales afin de tracer $Z_1=f(d)$

La courbe, représentée par la Figure 3.10, est conforme à la profondeur Z_1 fournie par l'équation (3.23) et elle est très régulière. Cependant, la profondeur devient difficilement calculable lorsque la disparité

d est proche de 1 ou 2 pixels. De plus, la qualité des mesures est fortement dépendante des méthodes de mise en correspondance utilisées pour calculer la disparité.

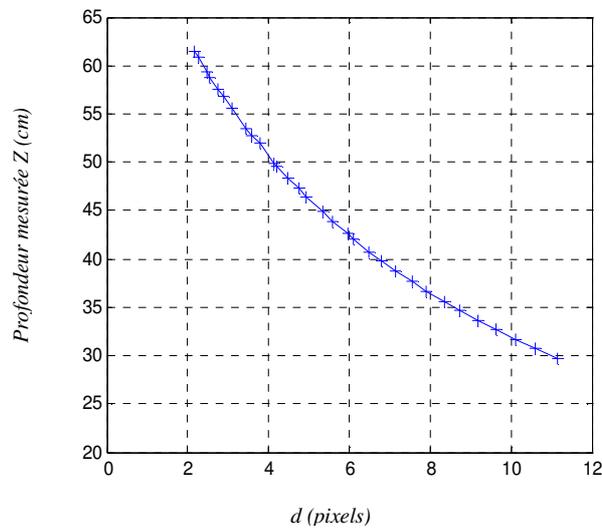


Figure 3.10 Résultats expérimentaux de $Z_1=f(d)$. La courbe suit l'expression de l'équation (3.23), tracée avec $Z_c=40$ cm.

Nous présentons des mesures expérimentales pour trois rangs de profondeur, afin d'estimer la précision de la mesure en termes d'erreur de mesure de profondeur. Dans ce cas, seules les mesures effectuées sur les sous-images I_3 et I_2 sont présentées pour cette expérimentation. Comme il l'a été montré avec l'expression (3.26), l'erreur de profondeur dépend de deux facteurs : l'estimation des paramètres géométriques et la mesure de la disparité.

Dans ce cas, nous sommes intéressés par la mesure de la qualité des paramètres estimés durant le calibrage. Afin de minimiser le rôle de la mesure de la disparité dans notre estimation, nous avons utilisé un échiquier comme cible. La détection des coins est effectuée avec la fonction correspondante de la librairie OpenCV. Pour chaque niveau de profondeur, de nombreuses acquisitions ont été réalisées, permettant d'obtenir un millier de mesures. Nous avons positionné l'échiquier à 4 cm, 33 cm et 88,5 cm de la caméra et la profondeur mesurée correspond à Z_1 (la distance séparant l'échiquier et le centre optique).

La Figure 3.11 donne la distribution de la profondeur mesurée pour chacun des coins. En travaillant à de courtes distances, les images présentent des distorsions optiques importantes. Dans ce cas, la seconde méthode offre une amélioration significative, fournissant un écart-type inférieur à 1 %. Pour évaluer l'erreur de profondeur, le pourcentage n'est pas significatif ici, parce que l'incertitude de mesures est directement liée à la plage de profondeur (voir équation (3.26)), ce qui signifie que l'erreur de profondeur croît avec celle-ci. Pour cette raison, notre évaluation est présentée en comparaison de l'erreur prédite théoriquement avec la mesure de l'écart-type. Dans ce cas, la principale difficulté concerne l'évaluation de la variation de la disparité ∂d , ce qui est difficilement réalisable. Même si nous utilisons un détecteur sous pixellique, ∂d dépend fortement de l'échelle de l'échiquier dans les sous-images ainsi que de la distance de travail. Pour fournir la meilleure estimation, nous avons bornés ces

valeurs par deux valeurs réalistes, fixées à 0,1 et 0,5 pixel. Ces valeurs sont ensuite utilisées pour la comparaison des résultats présentés dans le Tableau 3.1. Ce tableau montre que l'écart-type observé au cours de nos mesures est de l'ordre de l'erreur de profondeur estimée. Cela confirme que le calibrage proposé est bien adapté à notre approche de mesure par homographie variable. Ces valeurs de l'erreur de profondeur permettent de conserver une bonne incertitude de mesures pour des applications dans le domaine de la métrologie.

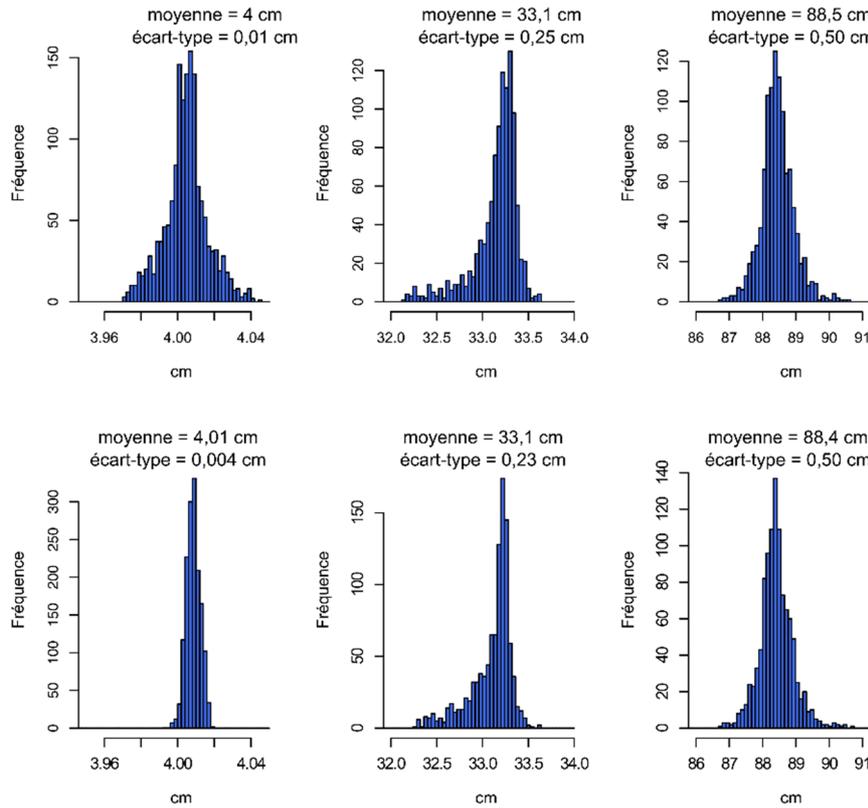


Figure 3.11 Distributions de la mesure de profondeur pour trois rangs de mesures : 4 cm, 33 cm et 88.5 cm. En haut : résultats obtenus avec la première méthode de calibrage (lentilles idéales). En bas : ceux avec la seconde méthode (modélisation des distorsions).

Profondeur Z_1 (cm)	Erreur de profondeur prédite $\partial Z = \left(\frac{z^2}{bf_1}\right) \cdot \partial d$	Ordre de grandeur de l'erreur théorique de profondeur estimée		Ecart-type observée expérimentale CS / CP
		$\partial d_{\min} = 0,1 \text{ pixel}$	$\partial d_{\max} = 0,5 \text{ pixel}$	
4	$(0,22 \cdot 10^{-3}) \cdot \partial d$	0,022 mm	0,111 mm	0,12 mm / 0,04 mm
33	$(15,1 \cdot 10^{-3}) \cdot \partial d$	1,51 mm	7,56 mm	2,47 mm / 2,29 mm
88,5	$(107 \cdot 10^{-3}) \cdot \partial d$	10,7 mm	53,8 mm	49,6 mm / 50,1 mm

Tableau 3.1 Comparaison entre les mesures de l'erreur théorique et expérimentale ($bf_1=7,2$ pixels est obtenue pendant l'étape de calibrage). CS : calibrage simplifié ; CP : calibrage perfectionné.

Dans cette partie, nous avons présenté les méthodes de calibrage, basées sur l'homographie variable, que nous avons développées. De plus, l'erreur de mesure pour ces deux méthodes est identique à celle de la méthode de triangulation. Enfin, les erreurs de mesures expérimentales et théoriques ont été comparées, et elles sont du même ordre de grandeur. Ainsi, nous avons démontré que nos calibrages

basés sur l'homographie variable permettent d'effectuer un calibrage simplifié d'un dispositif multi-vues tout en conservant la précision de mesure de la méthode par triangulation.

3.5.5. Critère de calibrage

Nous avons déterminé un critère facilitant la vérification de la qualité du calibrage du point de vue du calcul des homographies. Il ne sert pas à prévenir les erreurs de mesures et n'est pas utilisé pour affiner les calculs. Il facilite uniquement l'indication selon laquelle le calibrage n'a pas été effectué dans de bonnes conditions, notamment au niveau des réglages de la focale des mini lentilles.

Le critère est basé sur le constat que le produit des homographies $H_{I_0-I_1}$, $H_{I_1-I_2}$, $H_{I_2-I_3}$ et $H_{I_3-I_0}$ est égale à un critère identité :

$$H_R = H_{I_0-I_1} \cdot H_{I_1-I_2} \cdot H_{I_2-I_3} \cdot H_{I_3-I_0} = I \quad (3.46)$$

Après avoir effectué ce calcul à partir des matrices déterminées lors du calibrage, la matrice R des résidus est calculée comme suit :

$$R = H_R - I \quad (3.47)$$

Ensuite, à partir de la matrice R, il est possible de calculer le critère C de calibrage, ayant pour expression la somme des coefficients de cette matrice résidu :

$$C = \sum_{i=0}^n \sum_{j=0}^m R(i, j) \quad (3.48)$$

Si le calibrage des matrices est parfait alors le critère C est égal à 0. En prenant en compte les courbes théoriques, faisant le lien entre la profondeur et la disparité et les calculs de précision, le seuil du critère est fixé à 0,5 suivant les données du Tableau 3.1.

Valeur du critère avec la caméra multi-vues

Du point de l'expérimentation, le calibrage est réalisé à l'aide de deux acquisitions, comme expliqué dans la partie 3.5. Les dimensions de la scène étudiée sont de 40 cm par 40 cm. Le calibrage a été réitéré une vingtaine de fois. Puis, la moyenne du critère de calibrage a été calculée sur l'ensemble des calibrages. Les résultats obtenus sont en moyenne inférieurs à 0,5 à partir des données obtenues avec une vingtaine de calibrage.

Valeur du critère avec le système multi-caméras

Le calibrage a aussi été effectué avec un système multi-caméras, développé pour les applications grande distance et présenté dans le chapitre 3. La procédure, appliquée à la caméra multi-vues est aussi utilisable par ce système. Cependant, comme elle nécessite de calibrer à des distances comprises entre 2,5 m et 5 m, la taille de la mire nécessaire au calibrage est de 55x50 cm. De plus, l'une des difficultés rencontrées est de s'assurer que les deux plans de calibrage sont bien parallèles. Cependant, les résultats obtenus du point de vue du critère sont toujours inférieurs à 1 : la moyenne obtenue, pour un échantillon d'homographies obtenues à partir de 10 mesures, est de l'ordre de 0,4. La précision est moins bonne qu'avec l'autre système, mais les résultats restent fiables pour de telles distances.

3.6. Comparaison avec le calibrage à partir de plusieurs vues [Zhan00]

Dans cette partie, notre calibrage est comparé à la méthode de Zhang (avec la Camera Calibration Toolbox for Matlab implémentée par J-Y Bouguet). L'échiquier utilisé pour les calibrages est composé de 5 par 6 carreaux. Notre calibrage est effectué avec deux plans parallèles distants d'une valeur d_z de 20 cm. Pour la méthode de Zhang, une vingtaine d'acquisitions sont effectuées dont seize sont recommandées pour le calcul des paramètres intrinsèques (cf. Figure 3.12) afin de réaliser une détection robuste. Les valeurs des paramètres intrinsèques obtenues pour chacune des vues sont détaillées dans le Tableau 3.2.

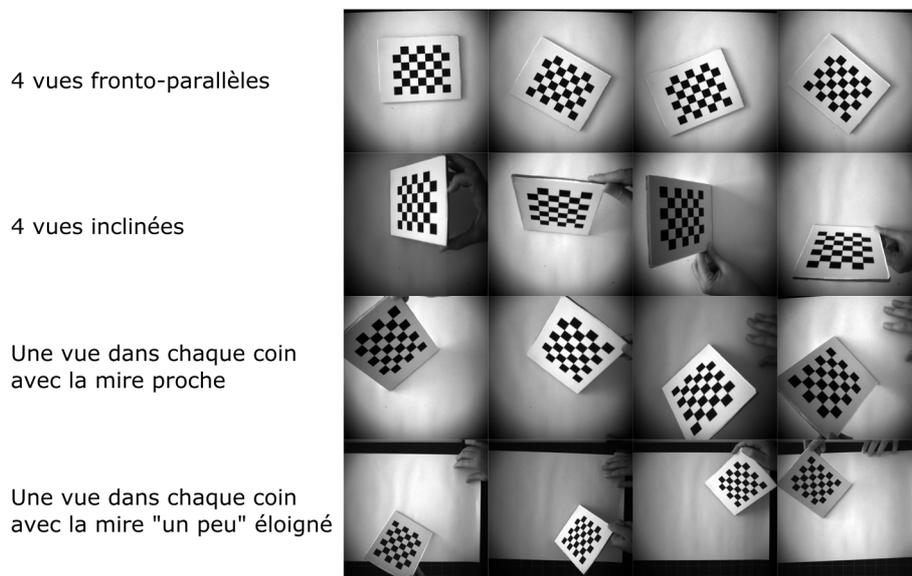


Figure 3.12 Les 16 vues recommandées pour le calibrage des paramètres intrinsèques.

Vues	Paramètres intrinsèques			
	u_0 (pixels)	v_0 (pixels)	f_x (pixels)	f_y (pixels)
l_0	289,7	236,0	798,0	790,5
l_1	303,5	241,5	800,0	794,3
l_2	284,9	224,8	796,2	791,3
l_3	292,2	229,9	801,8	795,1

Tableau 3.2 Valeurs des paramètres intrinsèques pour chacune des vues

Couple de vues	Paramètres extrinsèques					
	r_x (rad)	r_y (rad)	r_z (rad)	t_x (mm)	t_y (mm)	t_z (mm)
l_0-l_1	$6,0 \cdot 10^{-3}$	$-1,3 \cdot 10^{-2}$	$8,7 \cdot 10^{-4}$	$-1,5 \cdot 10^{-2}$	8,8	3,7
l_1-l_2	$-4,7 \cdot 10^{-3}$	$2,5 \cdot 10^{-2}$	$-1,0 \cdot 10^{-3}$	8,2	-0,2	-1,7
l_2-l_3	$-9,7 \cdot 10^{-3}$	$-2,4 \cdot 10^{-2}$	$6,8 \cdot 10^{-4}$	$-1,6 \cdot 10^{-2}$	-8,3	2,0
l_3-l_0	$8,4 \cdot 10^{-3}$	$1,3 \cdot 10^{-2}$	$3,3 \cdot 10^{-4}$	-8,3	-0,3	-4,0

Tableau 3.3 Valeurs des paramètres extrinsèques pour chacun des quatre couples : l_0-l_1 , l_1-l_2 , l_2-l_3 et l_3-l_0 . r_x , r_y et r_z sont les angles d'Euler utilisés pour construire la matrice rotation.

A partir de ces paramètres intrinsèques et des termes de translation des paramètres extrinsèques (où $b_u=t_x$ et $b_v=t_y$), obtenus avec le calibrage de Zhang, les matrices homographiques H_{Cij} définies par l'équation (3.20) sont calculées pour les comparer avec celles de notre calibrage.

Afin de comparer les deux méthodes, le critère présenté dans la partie 3.5.5 est calculé avec quatre matrices sur une plage de profondeur Z , comprise entre 30 cm et 1 m, dans les deux cas. Les résultats ainsi obtenus sont représentés sur la Figure 3.13. Ils montrent que notre méthode de calibrage s'avère plus précise que celle, utilisant les paramètres intrinsèques et extrinsèques, obtenus avec la méthode de Zhang. Nous l'expliquons par le fait que notre méthode ne nécessite pas de calculer individuellement l'ensemble des paramètres intrinsèques et extrinsèques.

De plus, l'effet de ces différences est mis en évidence en superposant l'image l_0 et l'image l_R , résultat du produit suivant : $I_R = H_R \cdot I_0$, définie par l'équation (3.46). La matrice H_R est calculée deux manières, soit en utilisant les matrices homographiques définies au cours de notre calibrage, soit avec les matrices homographiques obtenues à partir des paramètres intrinsèques et extrinsèques calculées avec le calibrage de Zhang. Pour chacune des méthodes, deux cas de figure sont traités : le premier correspondant à une image dans laquelle l'échiquier est situé à 40,0 cm et le second où l'échiquier est placé à 80 cm. Les résultats sont présentés sur la Figure 3.14. Pour les deux profondeurs, notre méthode procure un meilleur alignement des deux images superposées. A 80 cm, la différence d'alignement est inférieure au pixel ce qui la rend difficilement perceptible sur la Figure 3.14.

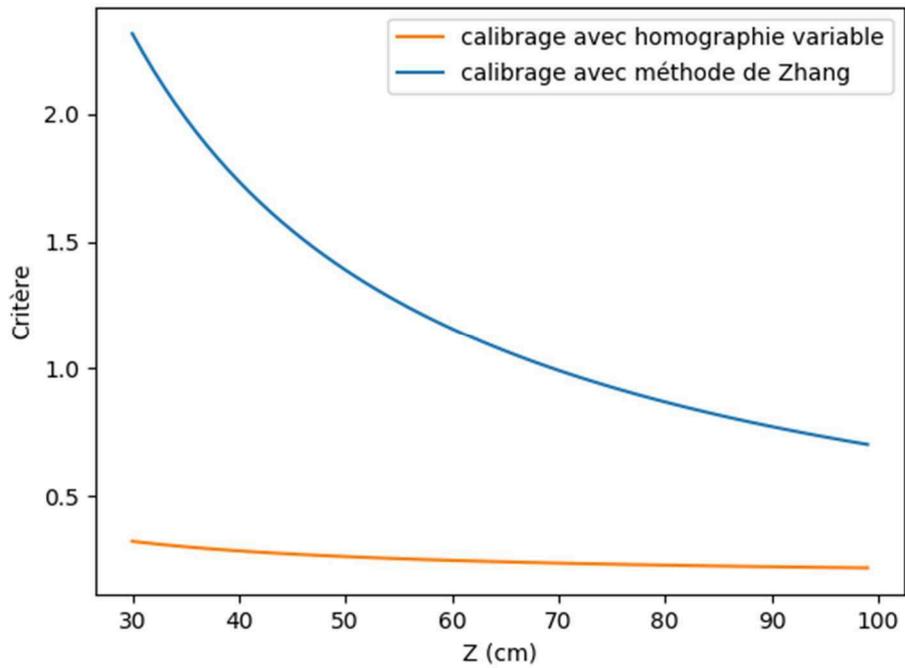


Figure 3.13 Comparaison des critères obtenus pour un calibrage avec l'homographie variable et avec la méthode de Zhang, pour Z variant de 30 cm à 1 m.

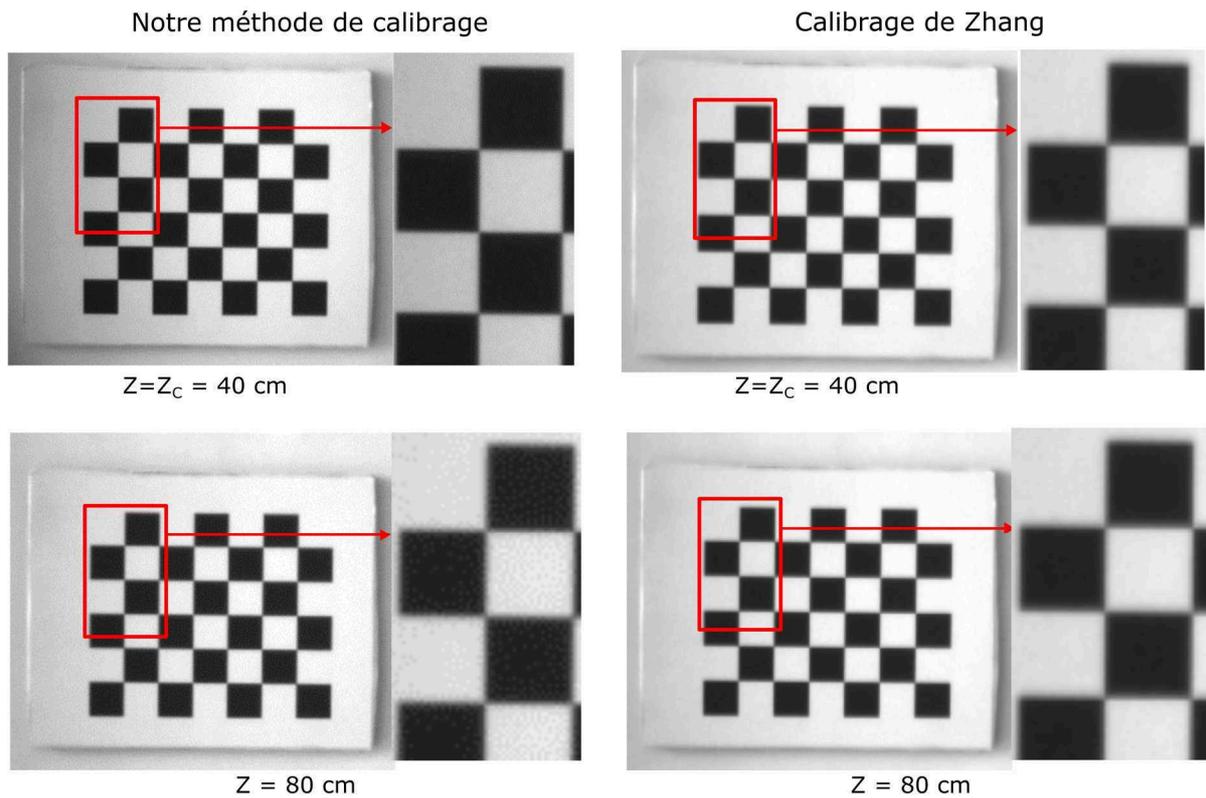


Figure 3.14 Comparaison visuelle des deux méthodes de calibrage, obtenues par $I_0 + I_R$ pour les distances $Z = 40$ cm (distance de calibrage) et $Z = 80$ cm.

Notre méthode de calibrage requiert le calcul de trois grandeurs par couple d'images ($b_{uf_{u1}}$, $b_{vf_{v1}}$ et Z_c) et d'une homographie H_c tandis que celle de Zhang nécessite le calcul de quatre paramètres intrinsèques par caméra et de six paramètres extrinsèques par couple de caméras. D'un point de vue numérique, les solutions seront plus stables et surtout le résultat est moins sujet à l'addition des écart-types

des grandeurs déterminées. Notre calibrage est donc plus facile à mettre en œuvre que celle de Zhang. De plus, la définition des matrices homographiques est indispensable aux calculs et aux traitements des images dans le chapitre suivant. Elle est donc la plus adaptée aux algorithmes définis au cours de cette thèse.

3.7. Conclusion

Ce chapitre a décrit notre système à multiples vues, considéré comme une caméra plénoptique et utilisé comme une caméra de profondeur. Ce système peut fonctionner pour détecter des objets à une profondeur allant jusqu'à deux mètres et donc présente un intérêt pour le contrôle par vision industrielle.

L'homographie variable est une approche alternative de la modélisation des systèmes à vues multiples. Le principal intérêt de cette méthode est de proposer une procédure simple et rapide de calibrage nécessitant seulement deux acquisitions, sans avoir besoin de déterminer au préalable les paramètres intrinsèques de la caméra. Une comparaison avec la méthode bien connue de triangulation a été effectuée. Nous avons démontré que les incertitudes de mesures sont comparables entre les deux méthodes. Pour notre modèle, le calibrage est plus simple, surtout si les paramètres intrinsèques ne sont pas considérés comme identiques pour chaque sous-système. Si les distorsions optiques sont significatives, comme à l'échelle macroscopique, nous avons proposé une méthode améliorée de calibrage, réduisant significativement l'incertitude de mesures. Le processus de calibrage reste identique entre le calibrage normal et perfectionné. Les deux méthodes ne requièrent que deux acquisitions. Les principes développés avec la méthode de l'homographie variable aboutissent à l'utilisation d'une homographie de référence nommée H_c . Les mesures sont basées sur une disparité d dépendante des homographies H_c et H . Un critère de calibrage a été mis en place pour vérifier la fiabilité du calcul des homographies. De plus, notre méthode de calibrage est comparée avec celle de Zhang. On constate que nos résultats sont meilleurs pour le calcul des matrices homographiques puisque les matrices sont directement définies au cours de notre calibrage tandis que celles obtenues à partir des paramètres définis par la méthode de Zhang sont calculées.

Cette modélisation et ce calibrage sont extensibles aux systèmes multi-caméras. Seuls les résultats concernant le critère de calibrage ont été présentés pour ces derniers. Cependant, la méthode de calibrage reste applicable de manière identique à celle présentée pour les dispositifs multi-vues. Notre objectif est de simplifier cette approche pour ces systèmes car sa mise en œuvre est contraignante. Ce modèle est extensible à des dispositifs ayant plus de 4 vues, tels que les systèmes *light-field*.

A partir de cette modélisation et de ce calibrage, nous avons envisagés différentes applications. Ces dernières sont les cartes de profondeur et les images refocalisées calibrées en profondeur, utilisées dans les domaines de la vision industrielle et de l'automobile.

4 Imagerie multi-vues : intérêts et algorithmes

4.1. Introduction

A partir des systèmes présentés dans le Chapitre 2 et du calibrage proposé dans le Chapitre 3, une propriété liée aux projections de différentes images dans une vue de référence nous a permis d'envisager plusieurs applications associées à ces dispositifs.

Ainsi, dans une première partie, nous présenterons le principe de rectification des différentes vues dans une image de référence utilisant les homographies. Dans la seconde partie, nous aborderons le principe de la refocalisation numérique calibrée en profondeur, et le concept de plan de travail ainsi que leurs applications possibles dans les domaines de la vision industrielle et de l'automobile. En vision industrielle, les images ainsi refocalisées permettent par exemple de faciliter certains traitements classiques tels que la détection de code-barres, et d'ajouter une information concernant la hauteur des objets que d'autres systèmes d'imagerie conventionnels ne fournissent pas ou plus difficilement. Dans l'automobile, l'information de profondeur issue du dispositif multi-caméras est exploitée afin de faciliter la détection d'obstacles en découpant la scène perçue en différentes zones selon la distance caméras-objet. Dans un troisième temps, nous présenterons notre méthode générant des cartes de disparité pour les objets présentant de larges zones uniformes. Elle est basée sur la propagation des contours. Son principal avantage est de ne pas nécessiter une étape de mise en correspondance.

En résumé, pour ces systèmes *light-field*, deux domaines d'application avec deux applications principales ont été définis tels que décrit dans le Tableau 4.1.

Application	Commentaires sur l'état de l'art	Objectifs pour la thèse
Refocalisation numérique	<ul style="list-style-type: none">• Utilisation pour la photographie• Refocalisation rarement calibrée en profondeur	<ul style="list-style-type: none">• Utilisation en vision industrielle notamment, et aussi dans le domaine de l'automobile• Refocalisation numérique calibrée en profondeur
Carte de profondeur	<ul style="list-style-type: none">• Méthodes complexes ne fonctionnant pas le plus souvent dans les zones uniformes• Mise en correspondance le plus souvent nécessaire	<ul style="list-style-type: none">• Utilisation privilégiée dans les zones uniformes• Pas de mise en correspondance

Tableau 4.1 Résumé des objectifs du chapitre 4

4.2 Rectification des images

L'extraction d'informations de profondeur, de même que la refocalisation numérique nécessite un recalage des images. Cette étape peut être assimilée à une rectification d'images, identique à celle utilisée en stéréovision. Contrairement à la stéréovision, où les caméras ne sont pas coplanaires, avec notre système, les droites épipolaires sont parallèles entre deux sous-images voisines. Cependant, les quatre mini-objectifs placés devant le capteur image, ne sont pas alignés au pixel près, ce qui n'est pas surprenant, avec des pixels mesurant $7,4 \mu\text{m} \times 7,4 \mu\text{m}$. La conséquence est que les droites épipolaires ne sont pas strictement horizontales ou verticales, comme le montre l'illustration de la Figure 4.1 avec une image brute.

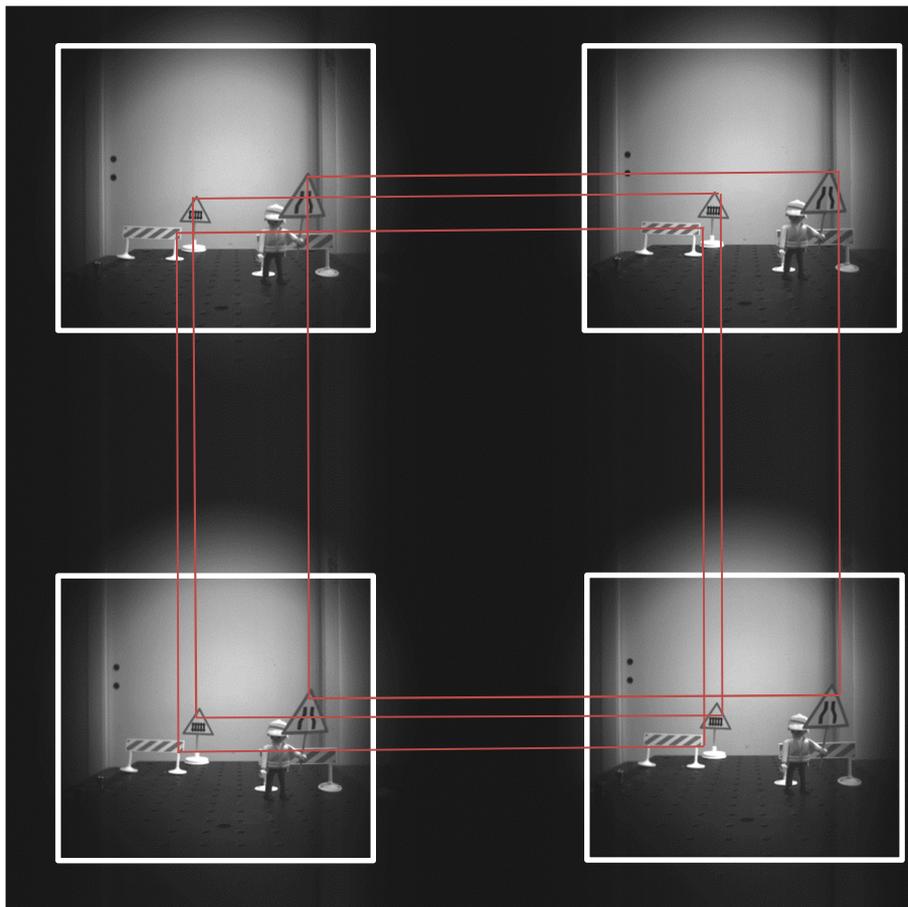


Figure 4.1 Sur l'image brute, où les mini-objectifs ne sont pas exactement alignés sur le capteur image, les droites épipolaires ne sont pas strictement horizontales ou verticales.

La rectification revient alors à aligner les images sur un réseau de droites épipolaires horizontales et verticales. La rectification des réseaux de caméras n'a jamais été réellement abordée dans la littérature pour les systèmes *light-field*. On peut néanmoins citer les travaux de Kang [KaLH08], cherchant à aligner la *baseline* de n caméras alignées horizontalement.

Nous avons abordé la rectification au moyen de l'homographie variable et de sa définition (3.18), où k est remplacé par Z/Z_c . Les sous-images brutes sont projetées dans le repère de l'image de référence.

La généralisation à un système à quatre vues, conduit pour les trois vues I_n voisines de l'image de référence I_{ref} , à calculer les rectifications (ou projections), notées I_n^R :

$$I_n^R(Z) = H_{i,c}(Z) \cdot I_i = [K_i(Z) \cdot H_{i,c}(Z_c) \cdot K_i^{-1}(Z)] \cdot I_i \quad (4.1)$$

La rectification ne connaît pas de solution unique, puisque les images sont rectifiées pour une profondeur Z , qu'il convient de choisir suivant le contexte applicatif. La Figure 4.2 montre deux exemples de rectification pour $Z = 60$ cm et $Z = 35$ cm. Les réseaux de droites épipolaires en trait plein montrent que l'alignement peut se réaliser sur le fond ou l'avant de la scène. Les deux faisceaux en traits pointillés (de dimensions identiques aux réseaux en trait plein), montrent que les droites épipolaires restent bien alignées, mais mettent en évidence la parallaxe, liée à la différence de profondeur des deux objets dans la scène.

Pour la suite, la rectification sera le préalable pour appliquer l'algorithme de calcul de cartes de profondeur que nous avons développé. La rectification ajustée en profondeur est également l'élément essentiel pour créer des images artificiellement refocalisées.

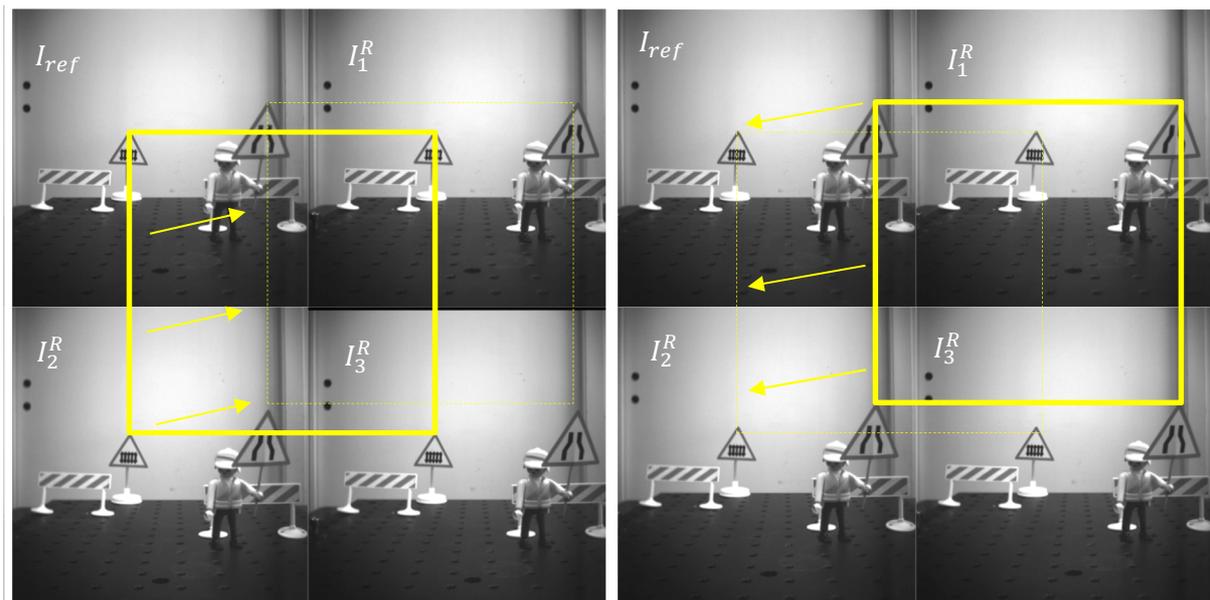


Figure 4.2 Exemples de rectification à $Z=60$ cm (droite) et $Z=35$ cm (image référence prise en haut à gauche). Le réseau carré de segments épipolaires en trait plein est translaté sur chacune des vues (traits pointillés), pour mettre en évidence la différence de parallaxe entre ces deux exemples de rectification.

4.3 De la rectification à la refocalisation numérique calibrée

D'un point de vue optique, la focalisation résulte de l'intersection des rayons optiques récoltés par un objectif, sur la surface inférieure à un pixel du capteur image. Quand le capteur est trop distant du plan image, on observe un effet de flou. Cet effet résulte donc de la superposition d'images de sources ponctuelles voisines mal focalisées sur le capteur image.

L'une des propriétés des systèmes *light-field* est de capturer l'orientation de rayons lumineux, et permet ainsi de déplacer artificiellement le plan image le long de l'axe optique. Dans ce cas de figure, il devient possible de générer un jeu d'images, où le degré de focalisation est fonction de la position du plan image virtuel. Dans les travaux de Vaish, avec un réseau de caméras, les images refocalisées sont calculées au moyen de projections homographiques [VWJL04]. Notre approche est similaire à celle-ci, avec la nuance que nos images sont refocalisées avec l'homographie variable pour les trois dispositifs présentés dans le Chapitre 2. Ce qui présente l'avantage d'être en mesure de générer une image focalisée à n'importe quelle distance.

Dans ce travail, nous recréons avec les dispositifs *light-field* du laboratoire, des images à focalisation variable, en additionnant les images rectifiées à l'image de référence et en jouant sur la profondeur Z utilisée comme paramètre de rectification :

$$S(Z) = \frac{1}{N} \sum_{i=0}^{N-1} H_{i,c}(Z) \cdot I_i = \frac{1}{N} \sum_{i=0}^{N-1} [K_i(Z) \cdot H_{i,c}(Z_c) \cdot K_i^{-1}(Z)] \cdot I_i \quad (4.2)$$

Pour une image $S(Z)$, tous les éléments de la scène situés à une distance Z de la caméra coïncident. Ce n'est pas le cas des autres éléments se trouvant à des distances différentes de Z , qui se superposent dans l'image avec une erreur de parallaxe, créant ainsi un effet de focalisation. Ce principe est illustré sur la Figure 4.3. Pour les dispositifs multi-vues, la somme d'images peut être généralisée à l'ensemble des N sous-images disponibles, à condition d'avoir préalablement réalisé un calibrage entre chacune des sous-images. L'utilisation de plusieurs images va alors fournir un effet visuel assez proche d'une focalisation, pour les objets de la scène situés à une distance Z .

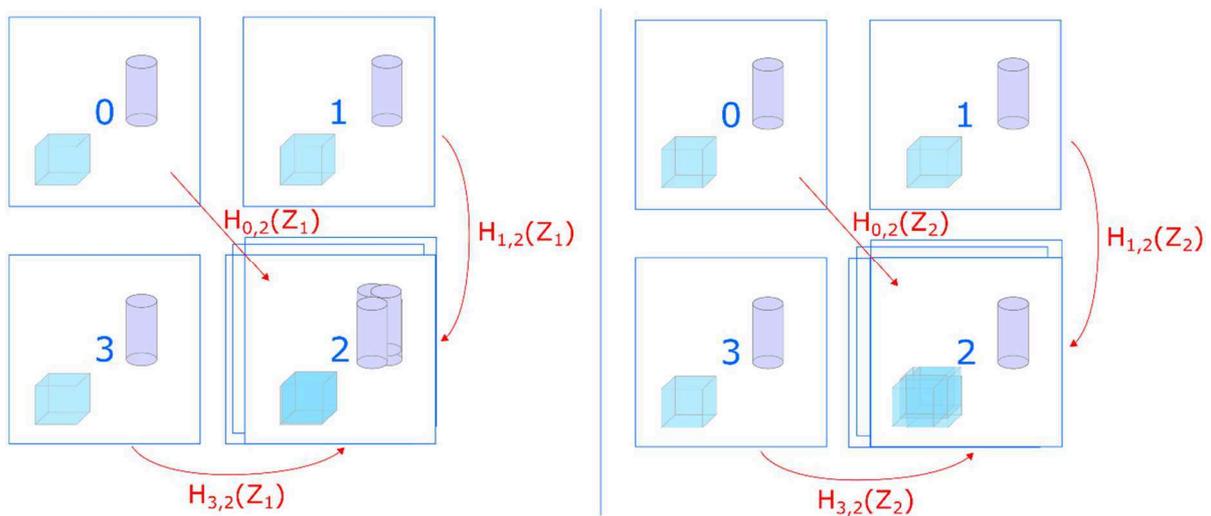


Figure 4.3 Etape de projection (l'image 2 est l'image de référence) ; à gauche, les images 0, 1 et 3 sont projetées dans l'image de référence (image 2), pour la profondeur Z_1 , le cube apparaît focalisé tandis que le cylindre apparaît dupliqué ; à droite, les images sont toujours projetées dans l'image 2, pour la profondeur Z_2 , le cube apparaît dupliqué tandis que le cylindre semble focalisé.

Le principe de refocalisation est présenté sur la Figure 4.3 d'un point de vue théorique. En considérant un cube placé à la profondeur Z_1 et un cylindre à Z_2 , quand la refocalisation est effectuée à la distance

Z_1 , les projections sont alignées sur le cube mais désalignées sur le cylindre, entraînant un effet de défocalisation. Cet effet est inversé quand on refocalise à la profondeur Z_2 .

Avec cette approche, nous sommes capables de générer un effet de refocalisation des profondeurs définies précisément. Comme la refocalisation est basée sur l'alignement ou le désalignement de la série d'images, il y a des limites à l'obtention d'une bonne perception visuelle des effets de focalisation ou de défocalisation. Par exemple, pour un désalignement de plus de 1 pixel, l'effet de flou peut disparaître et être remplacé par un effet de duplication, comme illustré sur la Figure 4.4, pour le bord de la table au fond de l'image. Inversement, un désalignement inférieur à 1 pixel n'est pas clairement visible sur l'image finale.

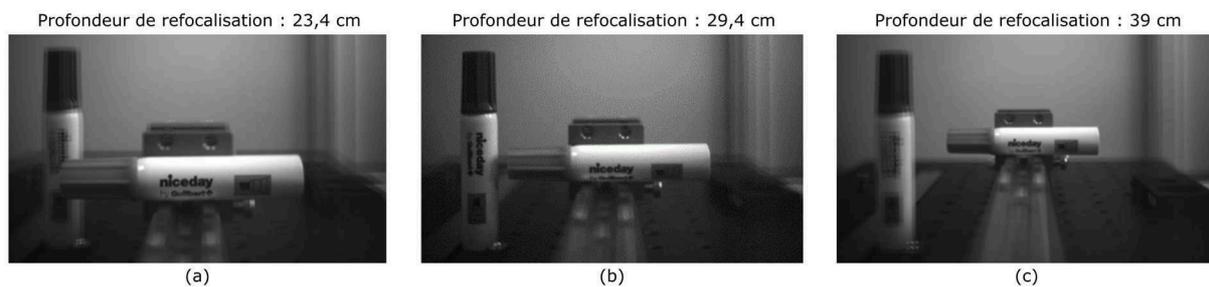


Figure 4.4 Exemple composé de 3 images refocalisées ; à gauche le plan refocalisé est à 23,4 cm ; au centre, il est à 29,4 cm ; à droite, le plan de refocalisation est à 39 cm.

Ces calculs étant assez faciles à mettre en œuvre, une image refocalisée peut être calculée en temps réel. Avec un ordinateur standard (Intel® core i5), ce calcul nécessite seulement 3 ms. Les projections par homographies sont assez aisées à mettre en œuvre avec la librairie OpenCV. Toutefois, parmi les méthodes d'interpolations, listées dans OpenCV, et qui sont mises en œuvre dans tout algorithme de projection d'images, il convient d'être vigilant et de ne pas choisir une méthode d'interpolation du type du plus proche voisin, mais une méthode d'interpolation non-linéaire pour avoir un rendu sub-pixellique.

Intérêts de la refocalisation numérique

C'est une approche 3D alternative : pas de calculs de cartes de disparité, mais mise en évidence de zones saillantes à la distance de focalisation, que nous traduisons comme un plan de travail sélectionné suivant une distance caméra-scène.

Nous verrons au paragraphe 4.5.1, que cette notion peut être mise à profit pour faciliter le contrôle et l'analyse 2D d'images. Nous montrerons également que la création d'une pile d'images permet d'aborder la vision 3D de façon différente et alternative aux approches connues. Dans ce dernier cas, nous relirons la mesure du degré de défocalisation à une mesure 3D.

4.4 Etude de la mesure du flou sur les images refocalisées

4.4.1 Critère de mesure du flou

Dans cette partie, nous nous intéressons plus particulièrement à la mesure de la focalisation pour montrer qu'une approche 3D par mesure de focalisation est envisageable. Nous avons décidé d'utiliser des méthodes couramment employées pour la mesure du focus, développées notamment pour le *shape-from-focus* [PePG13]. Les méthodes proposées vont être comparées afin de déterminer la plus adaptée à nos images, sachant que le flou présent dans ce type d'image n'équivaut pas à un flou optique classique généré par la mise au point d'un objectif. Dans notre cas, lorsque la disparité est inférieure à 3 pixels, on peut considérer que le flou dans les images refocalisées produit un effet visuel équivalent à celui dû à un objectif. Cependant lorsque la disparité est trop importante, l'effet de flou est remplacé par des duplications de l'objet. Ces multiples apparitions de l'objet ont pour conséquence de fausser la mesure de flou dans certains cas. Il devient donc nécessaire de bien choisir la distance de calibrage en fonction de la plage de travail limitant la valeur de la disparité et réduisant l'effet de duplications sur l'image finale.

Les méthodes exploitées ici sont décrites dans l'analyse comparative de différents critères de mesure de flou dans le domaine du *shape-from-focus* [PePG13]. Les différents opérateurs sont regroupés au sein de six familles définies selon leur principe de travail :

- Opérateurs basés sur le Gradient : ils sont basés sur la mesure du gradient ou de la dérivée première de l'image. Ils s'appuient sur la supposition selon laquelle, plus l'image est focalisée, plus l'image présente de contours nets.
- Opérateurs basés sur le Laplacien : cette famille cherche à mesurer la quantité de contours présents dans l'image, comme pour la catégorie précédente, mais en utilisant soit la dérivée seconde soit le Laplacien.
- Opérateurs basés sur les ondelettes : cette famille tire profit de la capacité des coefficients de la transformée discrète en ondelettes à décrire le contenu spatial et fréquentiel des images. Ces coefficients peuvent être utilisés pour la mesure du degré de flou.
- Opérateurs basés sur les statistiques : ces opérateurs utilisent les statistiques de plusieurs images comme descripteurs de textures pour calculer le degré de flou.
- Opérateurs basés sur la transformée en cosinus discrète : comme pour la famille des ondelettes, celle-ci tire profit des coefficients de la transformée en cosinus discrète pour calculer le degré de focalisation d'une image à partir du contenu fréquentiel.
- Opérateurs « micellaires » : cette famille regroupe les opérateurs n'appartenant pas aux autres.

La comparaison de la majorité des opérateurs proposés dans cet article est effectuée en utilisant les outils fournis par Pertuz et développés sous Matlab (disponible sur Mathworks avec la référence Said Pertuz). Cependant, comme nous ne disposons pas de la bibliothèque Ondelettes de Matlab, aucun des opérateurs de la famille Ondelettes n'est étudié. Les images testées dans l'article [PePG13] n'étant pas refocalisées et calibrées en profondeur, nos propres essais comparatifs sont effectués sur les critères présentés. Seules les méthodes présentant les meilleurs résultats sont décrites page suivante.

Pour comparer les différents critères de mesure du degré de flou, une expérience de mesure 3D est mise en œuvre à partir d'une pièce que nous considérons comme un étalon, et qui possède deux plans parallèles, distants de 4,5 cm (Figure 4.5). La mesure consiste à générer une pile d'images refocalisées et à mesurer le degré de focalisation sur deux régions d'intérêt (ROI), choisies sur les deux plans parallèles de cette pièce. Les images refocalisées et calibrées en profondeur sont calculées entre 24,3 cm et 38,2 cm par pas de 1 mm. La première ROI définie correspond à la zone la plus haute de la pièce, où est inscrite Université Haute-Alsace et la seconde située au niveau d'une des pattes de l'objet, où le logo MIPS est visible, (cf. Figure 4.5).



Figure 4.5 Pièce étalon utilisée afin de démontrer l'intérêt de l'utilisation d'un critère pour la mesure de flou dans les images refocalisées et calibrées en profondeur.

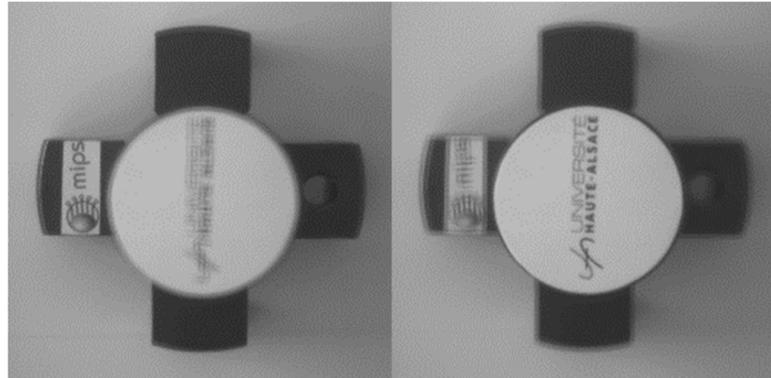


Figure 4.6 Images refocalisées sur la pièce étalon. A gauche : image refocalisée sur le logo MIPS ; à droite : image refocalisée sur le logo Université Haute-Alsace.

Dans le Tableau 4.2, les résultats de l'expérimentation sont regroupés pour chacune des méthodes testées. Les résultats les plus importants sont regroupés dans les colonnes 3 et 4 : la première fournissant la hauteur entre les deux plans et la seconde montrant l'erreur entre la mesure effectuée avec un pied à coulisse et celle obtenue avec l'opérateur testé. Ainsi, quatre méthodes, surlignées en bleu, présentent une erreur nulle : l'opérateur Brenner's focus measure, l'Image contrast, le DCT energy ratio et le critère de Tenengrad.

L'opérateur *Brenner's focus measure* (BREN) appartient à la famille « Micellaire ». Il est basé sur une mesure de différence entre pixels voisins horizontaux dans une fenêtre :

$$BREN = \sum_{(i,j)} I(i,j) \cdot I(i,j+2) \quad (4.3)$$

Le critère Image contrast (CONT) est aussi de la famille « Micellaire » et est basé sur la mesure de contraste dans une fenêtre, suivant la relation :

$$C(x,y) = \sum_{i=x-1}^{x+1} \sum_{j=y-1}^{y+1} |I(x,y) - I(i,j)| \quad (4.4)$$

Il a été adapté par Pertuz pour le *shape-from-focus* en accumulant le contraste sur le voisinage du pixel suivant l'équation :

$$CONT = \sum_{(i,j) \in \Omega(x,y)} C(i,j) \quad (4.5)$$

L'opérateur DCT energy ratio (DCTE) appartient à la famille « Transformée en cosinus discrète ». Soit $F_{u,v}$, la DCT d'un sous-ensemble $M \times N$ de l'image (typiquement, $M=N=8$). La mesure du degré de flou associée avec ce sous-ensemble Φ_S , peut être calculée comme suit :

$$\Phi_S = \frac{\sum_{\substack{u=0 \\ (u,v) \neq (0,0)}}^{M-1} \sum_{v=0}^{N-1} F(u,v)^2}{F_{0,0}^2} \quad (4.6)$$

Le critère de Tenengrad (TENG) est une mesure du degré de focalisation basée sur l'amplitude du gradient image et est défini comme suit :

$$\Phi_{x,y} = \sum_{(i,j) \in \Omega(x,y)} (G_x(i,j)^2 + G_y(i,j)^2) \quad (4.7)$$

Avec G_x et G_y , gradients image en X et Y calculés par convolution de l'image I avec les opérateurs de Sobel.

Sachant que ces quatre méthodes sont celles ayant obtenu les meilleurs résultats pour notre expérimentation, ceux-ci ont été représentés sur les deux courbes de la Figure 4.7. Pour les deux ROI, les courbes montrent bien que, pour ces quatre méthodes, le maximum est bien identifiable. Ainsi, du point de vue de la première ROI, les courbes TENG et BREN présentent une plus grande amplitude que pour les deux autres critères. Pour la ROI 2, on constate que, comme pour la ROI 1, l'amplitude est la plus importante pour les opérateurs TENG et BREN. Cependant, les deux courbes présentent un maximum local proche de leur maximum. Les deux maximums de la courbe TENG sont plus éloignés l'un de

l'autre, facilitant l'identification du maximum recherché. Pour la suite, nous avons choisi de conserver l'opérateur de Tenengrad pour le reste de nos expérimentations.

Méthode	d_1 =distance caméra - ROI 1 (cm)	d_2 =distance caméra - ROI 2 (cm)	$\epsilon=d_2-d_1$ = hauteur de la pièce (cm)	Erreur (%)
GRAT : Thresholded absolute gradient	29,88	35,68	4,8	24,4
ACMO : Absolute central moment	29,68	34,38	4,7	4,4
BREN : Brenner's focus measure	29,98	34,48	4,5	0
CONT : Image contrast	29,98	34,48	4,5	0
CURV : Image curvature measure	29,98	34,08	4,1	8,9
DCTE :DCT energy ratio	29,98	34,48	4,5	0
DCTR : DCT reduced energy ratio	29,98	31,48	1,5	66,6
GDER : Gaussien derivate	32,68	37,68	5,0	11,1
GLVA : Gray-level variance	29,98	34,48	4,5	0
GLLV : Gray-level local variance	29,88	38,18	8,3	84,4
GLVN : Normalised gray-level variance	29,98	38,18	8,2	82,2
GRAS : Squared gradient	30,08	35,58	5,5	22,2
HELM : Helml and Scherer's mean method	29,98	35,18	5,2	15,6
HISE : Histogram entropy	34,18	28,68	-5,5	22,2
LAPE : Energy of Laplacian	29,38	33,78	4,4	2,2
LAPM : Modified Laplacian	29,98	34,38	4,4	2,2
LAPV : Variance of Laplacian	29,98	34,38	4,4	2,2
LAPD : Diagonal Laplacian	29,98	34,38	4,4	2,2
SFRQ : Special frequency measure	30,58	36,38	5,8	28,9
VOLA : Vollath's autocorrelation	29,88	34,88	5,0	11,1
TENV : Tenengrad variance	29,88	34,48	4,6	2,2
SFIL : Steerable filters-based measure	32,68	37,58	4,9	8,9
HISR : Histogram range	29,98	36,38	6,4	42,2
GRAE : Gradient energy	30,08	35,68	5,6	24,4
TENG : Tenengrad	29,98	34,48	4,5	0

Tableau 4.2 Comparaison des différentes méthodes de calculs de degré de flou. Les distances d_1 et d_2 correspondent au maximum du critère.

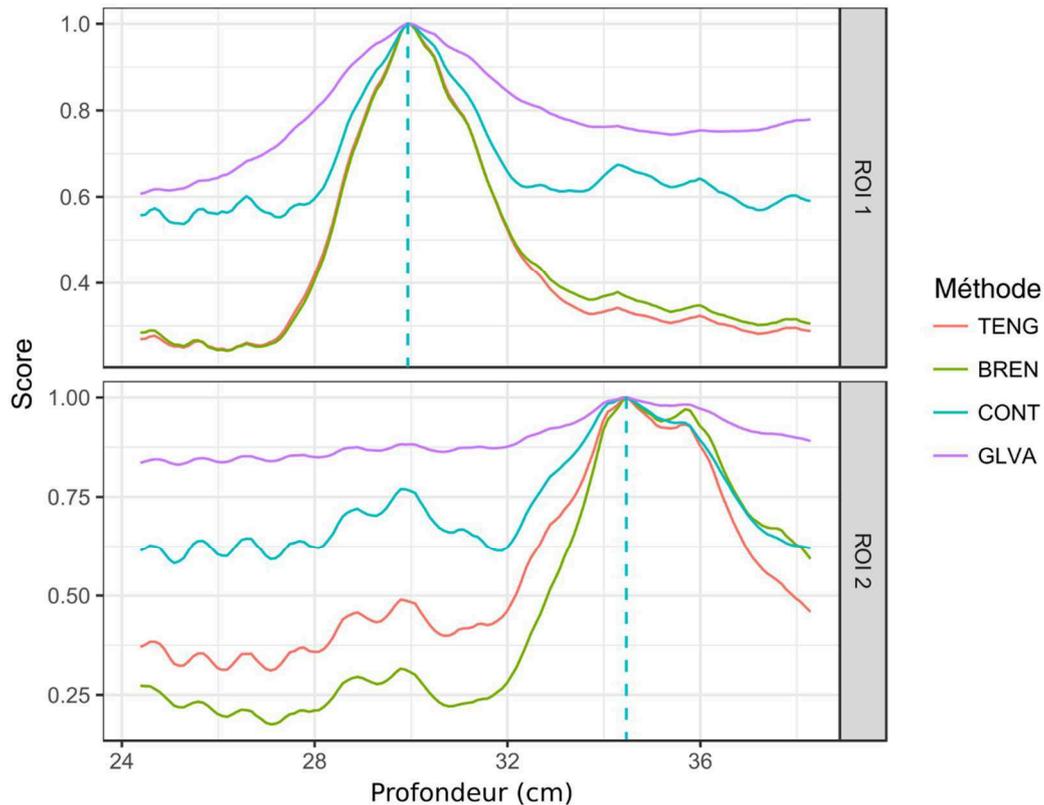


Figure 4.7 Courbes comparatives entre les quatre critères de calcul du degré de flou. Chaque image refocalisée est séparée par un pas de 1 mm.

4.4.2 Incertitudes de mesures et influence des paramètres

Cette partie traite des incertitudes des mesures effectuées en travaillant avec une pile d'images refocalisées $S(Z)$, pour extraire une information 3D d'une prise de vues. Au Chapitre 3, nous avons reporté une étude consacrée à la qualité du calibrage. L'étude menée ici est différente, d'une part, car le procédé de mesure n'est pas exactement le même et d'autre part, on étudie plus particulièrement les influences des paramètres mis en jeu dans la mesure 3D par degré de focalisation.

Les deux paramètres clés de ce processus de mesure sont :

- le pas d'échantillonnage en Z pour créer la pile d'images refocalisées ;
- la mesure de Tenengrad, dont le principal paramètre est la taille de la fenêtre d'analyse.

Ces deux paramètres sont de natures très différentes, et comme souvent en imagerie, il est difficile d'établir un calcul exact des incertitudes de mesures liées à chaque paramètre. Nous avons donc choisi une approche statistique, en utilisant une pièce mécanique comme étalon de mesure (Figure 4.5).

Nous aurions pu ajouter aux deux items précédents, un paramètre lié à la création de la pile d'images refocalisées. Les images de la pile sont la somme de projections homographiques (voir eq. (4.2)), qui

reposent notamment sur un algorithme d'interpolation. Cette étape importante est réalisée par une fonction de la librairie OPENCV et ne possède pas de paramètres ajustables. Les qualités des interpolations seront donc évaluées par le biais des mesures qui seront présentées.

Notons que toutes les valeurs numériques présentées ici, sont celles des paramètres de la caméra multi-vues du MIPS.

4.4.2.1 Conditions expérimentales

Les mesures présentées dans ce paragraphe sont effectuées sur un étalon de mesure de 45 mm de hauteur et placé à une trentaine de centimètre de la caméra. Nous utilisons un nuage de points de référence, composé de plus de 4500 points de mesure. Chaque point de mesure est le centre d'une zone d'intérêt, qui balaye suivant l'axe Z la pile d'images refocalisées, afin de calculer le critère de Tenengrad C_i pour chaque point P_i du nuage de points. La position du maximum de C_i sur l'ensemble de la pile, traduit une zone à fort gradient dans une des images de la pile, et donc (a priori) une mesure de focalisation optimum à cette position.

Avec l'ensemble des points du nuage de points, il est facile de construire un histogramme de mesures, qui mettra en évidence les mesures cohérentes (*inliers*) et aberrantes (*outliers*). A partir de la connaissance des distances caméra-objet, nous avons placé des seuils fixes de mesures, permettant de distinguer les *inliers* et les *outliers*, et de déduire les distances moyennes et les écart-types pour l'ensemble des mesures de la classe de *inliers*. La Figure 4.8 illustre l'étalon de mesure et un nuage de points après le classement en *inliers* (vert) et *outliers* (rouge) apparaissant dans les zones uniformes.

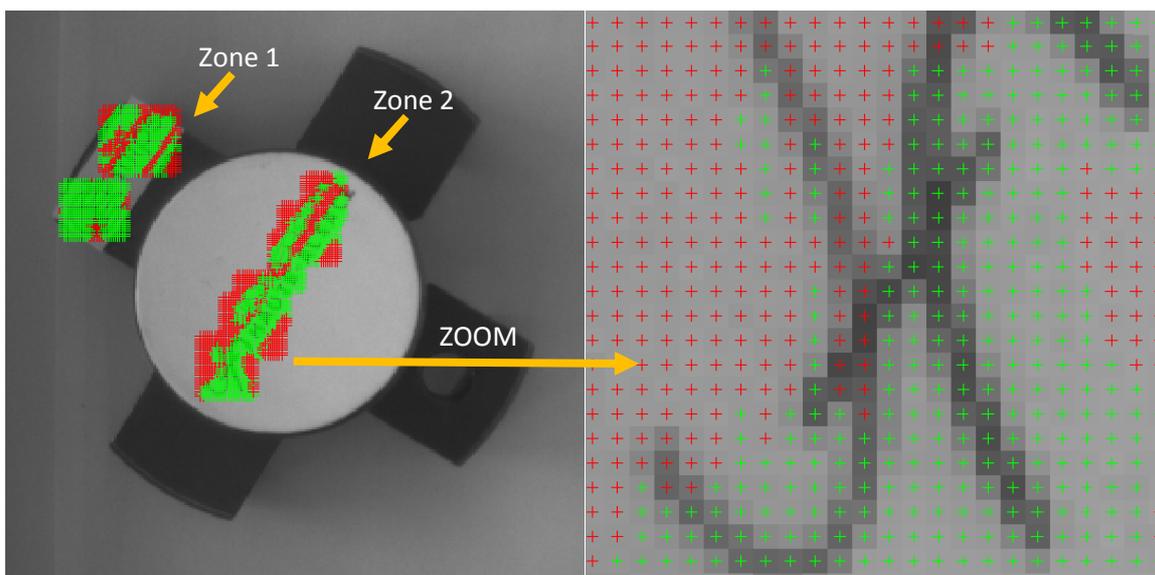


Figure 4.8 Maillage des points de mesure et exemples d'inliers (vert) et outliers (rouge).

4.4.2.2 Echantillonnage et limite de résolution

La création d'une pile d'images $S(Z)$ nécessite de définir une plage de profondeur comprise entre Z_{min} et Z_{max} et un pas d'échantillonnage dZ . Ce pas peut être fixe, mais il peut aussi être variable. Si on considère que la mesure du degré de refocalisation revient à mesurer l'alignement local (c'est-à-dire la disparité) des quatre sous-images additionnées ensemble (éq. (4.2)), il est préférable d'évaluer la méthode en fonction d'un échantillonnage régulier de la disparité.

Comme cela avait été vu dans le chapitre consacré au calibrage (éq. (3.23)), la profondeur est liée à la disparité entre les images par l'expression :

$$d = \frac{bf}{Z} - \frac{bf}{Z_c} \quad (4.8)$$

et le pas d'échantillonnage dZ est donné en fonction du pas d'échantillonnage dp (et réciproquement) :

$$dZ = -\frac{Z^2 dp}{bf} \text{ et réciproquement } dp = -\frac{bf dZ}{Z^2} \quad (4.9) \text{ et } (4.10)$$

dZ peut alors être vu comme une incertitude de mesure, qui est systématique et est fonction du pas d'échantillonnage dp .

La Figure 4.9 illustre un exemple, où le calibrage est réalisé en choisissant un plan de référence situé à $Z_c = 50$ cm de la caméra, ayant pour caractéristiques : $b = 9$ mm et $f = 8$ mm. Sur cet exemple, la disparité s'étend entre -10 et 10 pixels avec un pas dp de 1 pixel. La pile de 20 images correspondantes débiterait par une image refocalisée à 30 cm, tandis que la dernière serait refocalisée à 1,2 m et l'échantillonnage dZ entre deux images serait irrégulier. Toutes mesures 3D issues de l'analyse de cette pile d'images, aurait une incertitude de mesure systématique égale à dZ .

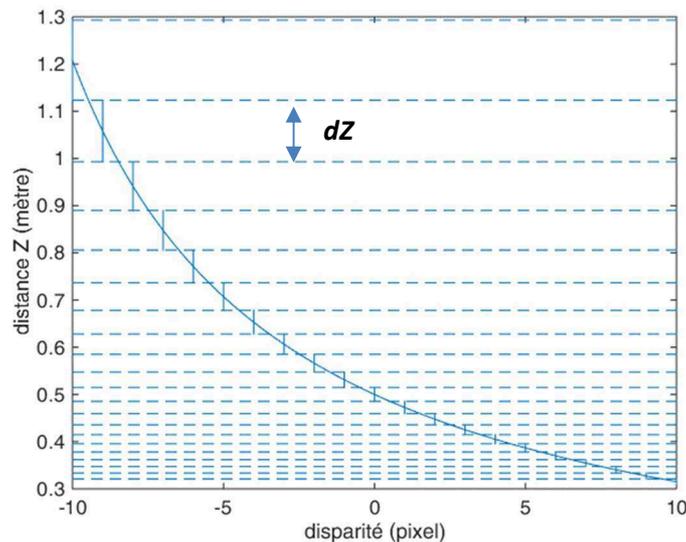


Figure 4.9 Incertitude de mesure dZ (barres verticales), si la disparité est échantillonnée par pas de 1 pixel

L'influence de dp est donc déterminante pour la définition de l'incertitude de mesure systématique. La Figure 4.10 montre cette influence pour trois valeurs de dp , allant de 1 à 0,01 pixel. Ces courbes doivent être interprétées avec précaution. Par exemple, expérimentalement, on peut douter du fait qu'une résolution inférieure au mm soit atteinte à 1,20 m en ayant fixé dp à 0,01 pixel (voir courbe (c)). Cela nécessiterait de disposer d'un opérateur capable d'analyser des images refocalisées, constituées de sous-images recalées au centième de pixel.

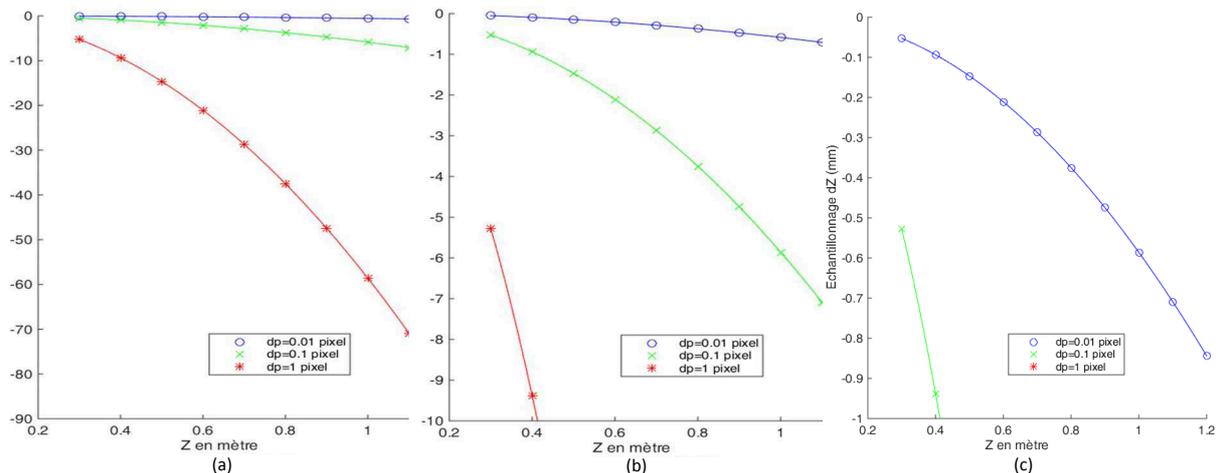


Figure 4.10 Evolution de l'incertitude de mesure systématique $dZ=f(Z)$, suivant le pas d'échantillonnage dp . Ce sont les trois mêmes courbes, vues successivement à pleine échelle.

Il faut donc retenir que les paramètres de création d'une pile d'images vont définir une valeur d'erreur systématique, pour les mesures qui seraient réalisées avec cette pile. D'autre part, il y a deux raisons pour lesquelles il est inutile de choisir un pas d'échantillonnage dZ trop faible :

- Un pas faible induit un nombre d'images important à traiter, ce qui peut augmenter significativement le temps de calcul,
- Une faible valeur de dZ n'est pas l'assurance d'avoir une incertitude de mesure faible. En effet, l'incertitude de mesure est composée d'une part de l'erreur systématique dZ , qui vient d'être définie, et d'autre part des valeurs des écart-types de mesures obtenus avec l'opérateur de Tenengrad, qui sont étudiés au paragraphe suivant. Ce ne serait pas cohérent d'avoir une valeur très faible de dZ , si les mesures du degré de focalisation présentent des écart-types significatifs.

4.4.2.3 Influence de la fenêtre de Tenengrad

Le critère de Tenengrad a été retenu comme critère de la mesure du degré de focalisation autour d'un pixel. Ce critère est simplement défini comme étant la moyenne du module de gradient de Sobel, dans une fenêtre carrée. Le choix de la taille de la fenêtre est à définir. Cela peut être une fenêtre étroite, comme par exemple pour une mesure sur une zone précise, ou bien peut être fixée plus librement, sur des zones dégagées de l'objet à mesurer.

Cette série expérimentale est menée avec une pile d'images échantillonnée avec un pas de disparité de 1/100 de pixel, correspondant à une erreur systématique dZ de l'ordre de 0,05 mm. Cette erreur est suffisamment faible, pour ne pas influencer significativement cette série de mesures.

La taille de fenêtre varie de 3x3 à 45x45 pixels, et nous calculons le pourcentage d'inliers pour les 4689 points de mesure, ainsi que les hauteurs moyennes et les écart-types pour les zones 1 et 2. Enfin, la hauteur moyenne de l'objet est calculée par simple soustraction des deux hauteurs moyennes.

La Figure 4.11 montre deux histogrammes obtenus pour des fenêtres de 7x7 et 31x31. Elle illustre le fait que le nombre d'inliers augmente avec la taille de la fenêtre. Cette situation est attendue, surtout pour les points situés sur une zone uniforme. En croissant, la fenêtre va atteindre des zones de plus fort gradient et ainsi faire basculer un point d'outlier en inlier.

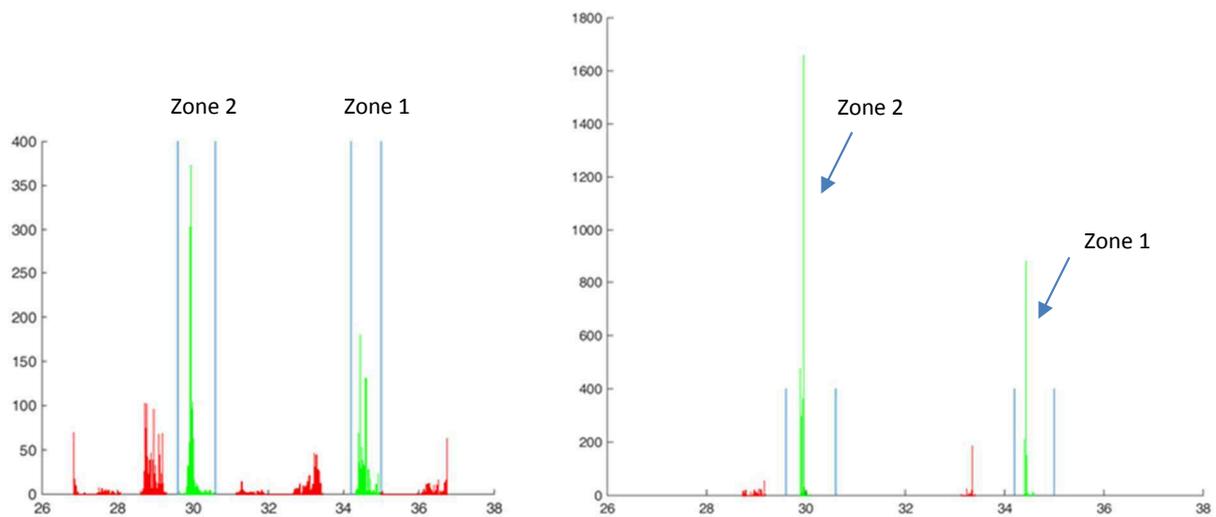


Figure 4.11 Histogrammes des mesures pour les zones 1 et 2 pour une taille de fenêtre de 7x7 (gauche) et 31x31 (droite).

Concernant les mesures de distances à la caméra, la Figure 4.12 montre que ces distances se stabilisent pour les deux zones étudiées à partir d'une taille de fenêtre de 25x25. Mais il est surtout intéressant de remarquer sur la Figure 4.13 que l'écart-type des mesures passe de plus de 1 mm à environ 0,25 mm. Cette évolution est également due à la croissance des fenêtres, puisque la plupart des fenêtres atteignent progressivement des zones de plus fort gradient, assurant ainsi une convergence des mesures.

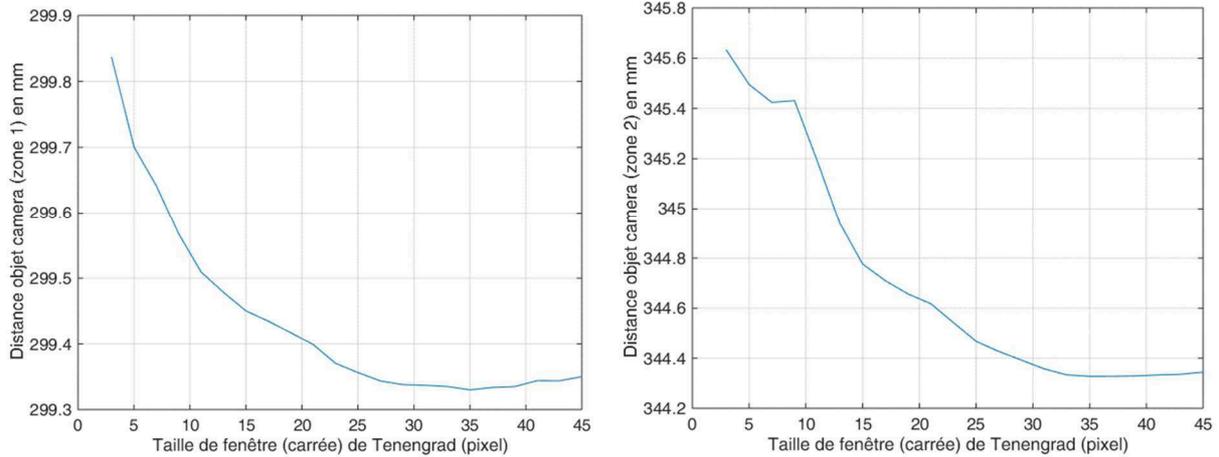


Figure 4.12 Distances objet caméra calculées pour les zones 1 et 2.

La Figure 4.13 droite représente la mesure de la hauteur de l'étalon. Cette mesure se stabilise autour de 45 mm pour une taille de fenêtre de 32x32 pixels.

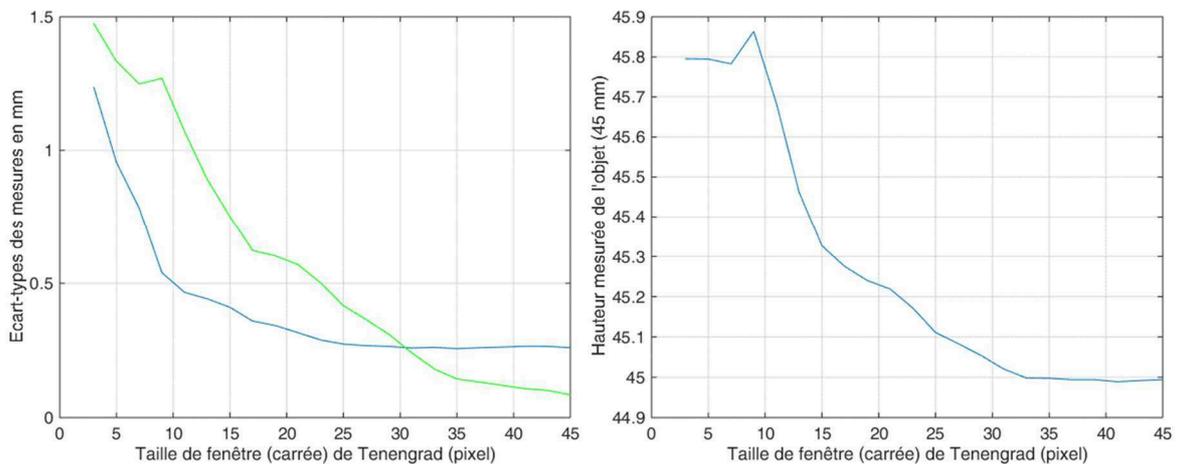


Figure 4.13 Ecart-type des deux mesures (gauche) et hauteurs de l'objet mesuré (droite) en fonction de la taille de la fenêtre de Tenengrad.

La conclusion immédiate qui conduit à penser que la meilleure mesure est obtenue si la taille de fenêtre est grande, est à notre avis en partie erronée. En effet, si un point de mesure est situé sur une zone à faible gradient, c'est le gradient d'une zone voisine qui va influencer le score du Tenengrad, et ainsi faire considérer de vrais outliers, comme des inliers. La Figure 4.14 met en évidence ce phénomène, pour une fenêtre de taille 7x7. Seules les zones à fort gradient sont détectées comme inliers. Les autres points, comme ici le U et N bénéficieront de l'augmentation de la taille de fenêtre pour être finalement vus comme des inliers. Le second effet de ce phénomène est un lissage des mesures, qui convergent vers la hauteur réelle de l'étalon avec un écart-type des mesures assez faible (cf. Figure 4.13 gauche).

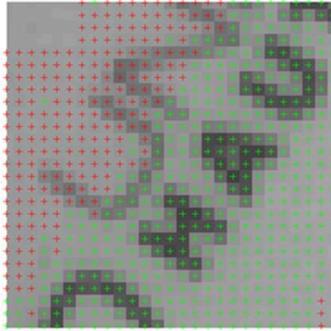


Figure 4.14 Illustration des *inliers* qui sont attachés aux zones à fort gradient pour une fenêtre de taille 7x7 pixels.

En conclusion, nous pouvons retenir de cette expérimentation que l'augmentation de la taille des fenêtres améliore la qualité de la mesure, mais ce point n'est vrai qu'à condition que la zone de mesure puisse être assimilable à une surface parallèle au plan image. Si ce n'est pas le cas, il est préférable de choisir des tailles de fenêtres étroites, même si l'incertitude de mesure se trouve augmentée.

L'incertitude de mesures est conditionnée par l'incertitude systématique de mesures dZ , dépendant du pas d'échantillonnage, calculable dans tous les cas de figure et estimable pour chaque mesure. En plus de dZ , l'écart-type lié à la mesure de Tenengrad influence l'incertitude de mesures. Si une mesure est ponctuelle, l'écart-type est plus élevé que celui obtenu avec un ensemble de mesures. Ainsi, pour diminuer l'écart-type, il faut multiplier les nombres de points pour chacune des mesures afin de tendre vers la valeur mesurée. De même, en bougeant la fenêtre de Tenengrad, la valeur mesurée est améliorée.

Dans la section suivante, nous allons présenter différentes applications de la refocalisation numérique calibrée en profondeur, telles que des détections conventionnelles ou bien des méthodes de mesures 3D « agiles ».

4.5 Exemples d'applications de la refocalisation numérique calibrée

Avec notre système composé de quatre mini-lentilles (présenté dans les chapitres 2 et 3), on considère qu'en ayant une distance de travail comprise entre 40 cm et 100 cm, une bonne perception visuelle est obtenue et chaque objet focalisé est réellement mis en valeur. Pour cette raison, nous pensons qu'il y a un réel intérêt des dispositifs *light-field* dans le domaine de la vision industrielle. L'idée principale est simple : si un objet n'est pas positionné à sa profondeur correcte, alors il peut apparaître non focalisé et l'inspection par vision artificielle échouera. De plus, nous avons montré que si l'effet de flou n'est pas perceptible visuellement, il peut être détecté par la mesure de focalisation. Celle-ci peut être utilisée afin de déterminer la meilleure profondeur de travail.

Afin de montrer l'intérêt des images refocalisées pour les traitements dans le domaine de la vision industrielle, deux types d'expérimentations ont été menés : le premier consiste à montrer l'intérêt de ces images pour effectuer des détections conventionnelles et le second groupe s'intéresse aux mesures 3D « agiles » à partir de la mesure du flou dans les images.

4.5.1. Mesures 2D : détections conventionnelles dans un plan de travail

Dans cette partie nous souhaitons montrer qu'un dispositif à refocalisation numérique peut avoir un intérêt dans le domaine du contrôle par vision. Pour cela, nous avons interfacé notre caméra avec refocalisation, à un environnement dédié à la vision industrielle (ici Vision Builder de National Instrument). Nous illustrons l'intérêt de la sélection d'un plan saillant dans le volume de travail (Figure 4.15) avec de la lecture optique de caractères ou OCR et de la lecture d'un code-barres.

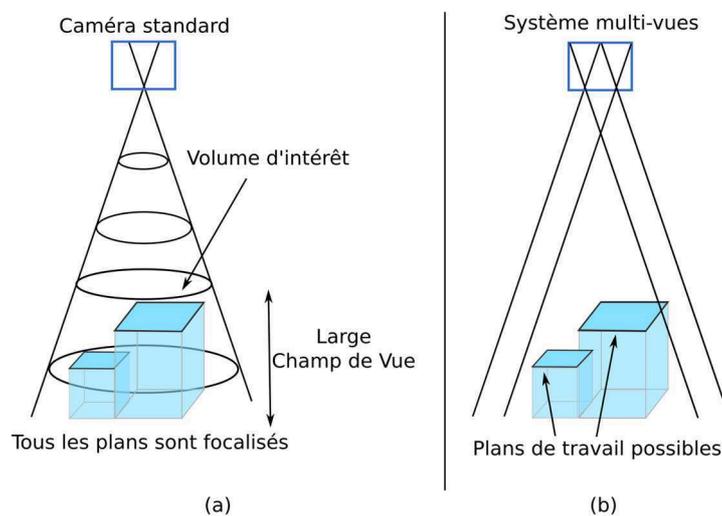


Figure 4.15 Représentation des plans d'intérêt définis par une image refocalisée ; à gauche plans d'intérêt pour une caméra standard et à droite, plans d'intérêt pour une caméra multi-vues.

4.5.1.1. Lecture de caractère

La première expérimentation consiste à lire un texte identique sur chacun des deux objets placés à deux distances différentes de la caméra, définissant ainsi deux plans de travail respectivement à 22,2 et 26,8 cm. La pile d'images refocalisées a été générée pour des distances comprises entre 20,2 cm et 30,2 cm par pas de 0,1 cm. La détection est effectuée avec le logiciel NI Vision Builder. La Figure 4.16 présente les deux images refocalisées pour lesquelles la détection est optimale pour chacun des cas (meilleur score de l'OCR du logiciel). Pour le plan de travail situé à 26,8 cm (Figure 4.16 gauche), la lecture du texte est possible avec une dizaine d'images de la pile, ce qui correspond à une plage de mesure sur 1 cm. Pour le second plan de travail, situé à 22,2 cm, le texte est lisible sur une plage d'environ 2 cm. Cette différence dans la plage de lecture s'explique par la taille des caractères, qui sont à la limite de la segmentation à la profondeur de 26,8 cm. L'erreur systématique de mesures est calculée au moyen de l'équation (4.9) et (4.10). Dans ce cas, elle est de l'ordre de 0,6 mm.

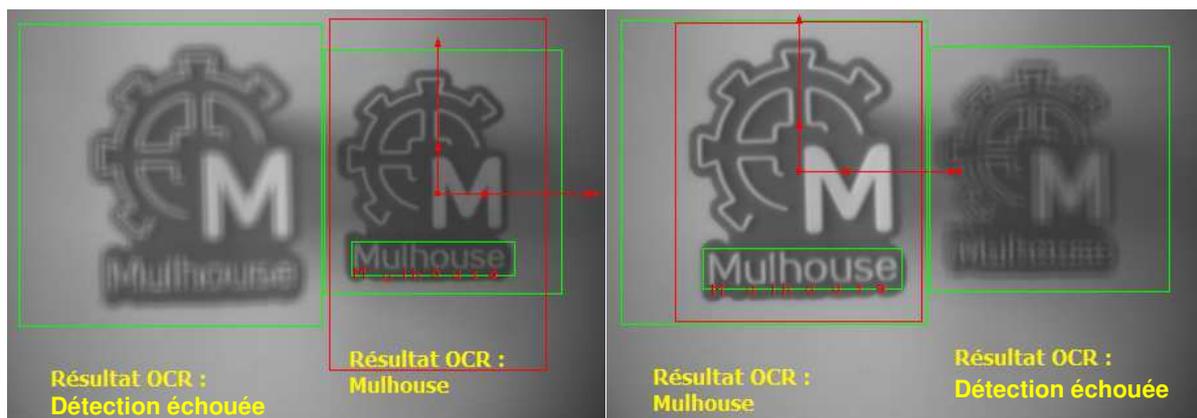


Figure 4.16 A gauche, image refocalisée à une distance de 26,8 cm, sur cette image, seul le texte du bas est détecté ; à droite, image refocalisée sur le plan situé à 22,6 cm.

Cet exemple montre que bien qu'identiques, les textes ne sont détectés que dans le plan de travail considéré. Nous avons ainsi introduit une sélectivité selon la profondeur.

4.5.1.2. Détection de code-barres

Une expérimentation similaire est effectuée avec un code-barres placé dans une scène à deux profondeurs différentes et distantes de 8,3 cm. Avec une caméra standard, comme sur la Figure 4.17 (a), les deux codes-barres sont correctement détectés. La pile d'images est calculée entre 27 cm et 44 cm avec un pas de 0,1 cm. Le code du dessous est détecté pour une profondeur de 32,5 cm tandis que pour celui du dessus, la profondeur est de 24,2 cm \pm 0,3 cm. Dans ce cas, l'erreur systématique de mesures est de l'ordre de 0,7 mm.

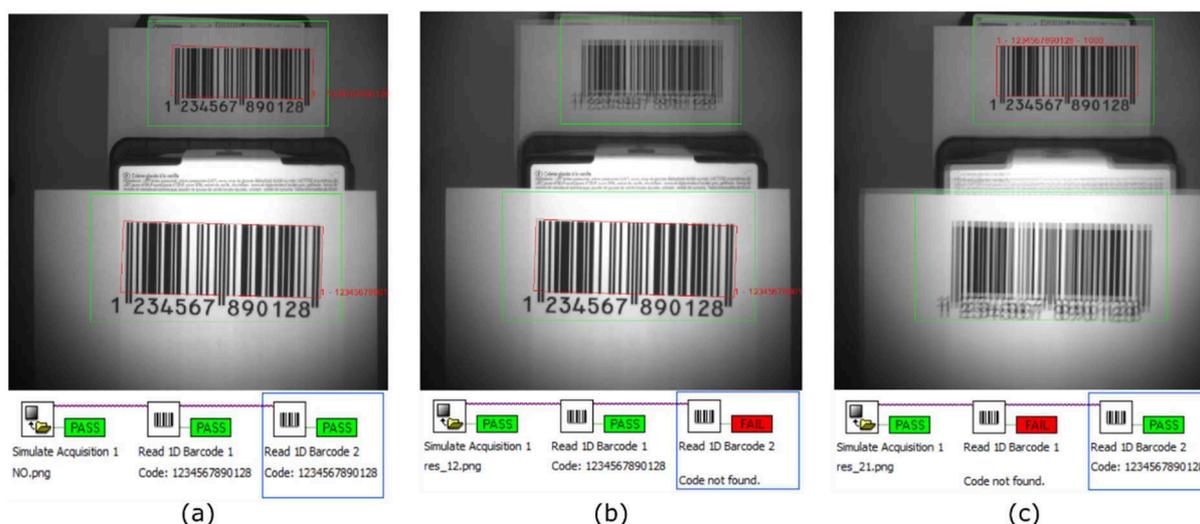


Figure 4.17 Détection de code barre (a) Vue totalement focalisée, correspondant à une vue acquise par notre caméra, les deux codes-barres sont détectés ; (b) l'image est refocalisée en bas (24,2 cm) ; (c) l'image est refocalisé sur le haut (32,5 cm). (b) et (c) en calculant les images refocalisées, la détection de code-barres est conditionnée par leur distance avec la caméra.

Ces deux exemples ont permis de montrer l'intérêt pour la vision industrielle des images refocalisées afin de faciliter les contrôles par vision. La vision 3D « agile » sera étudiée dans la partie suivante.

4.5.2. Vision 3D « agile »

Dans cette partie, nous montrons que la vision 3D peut être envisagée assez simplement en traitant une pile d'images refocalisées. Dans un certain nombre d'applications, une mesure 3D peut être utilisée pour vérifier un positionnement d'éléments dans un assemblage. Par exemple, l'utilisation de systèmes de vision 3D actifs (laser, ToF, stéréovision) dans ces cas de figure est peut-être surdimensionné et un système à refocalisation peut alors apporter une solution alternative et efficace.

Nous présentons ici quelques illustrations, dans un premier temps sur des exemples de mesures 3D, puis ensuite dans un contexte de détection d'obstacles sur une scène routière.

4.5.2.1. Mesures de hauteur

Dans cette partie, nous comparons deux approches pour mesurer une hauteur, égale à la différence entre deux profondeurs mesurées avec l'une des deux méthodes suivantes. Premièrement, en utilisant le critère de Tenengrad pour mesurer le degré de focalisation (cf. Section 4.4). Deuxièmement, en utilisant directement la pile d'images refocalisées mais avec un logiciel de vision industrielle, capable de faire de la reconnaissance de formes depuis une imagerie de référence.

Mesure 3D basée sur le critère de Tenengrad

Nous avons développé un algorithme de « *tracking* », capable à la fois de suivre un ou deux objets dans une scène, et d'estimer leurs distances par rapport à la caméra. Cet algorithme peut aussi être utilisé pour mesurer la hauteur de la pièce étalon que nous utilisons. L'algorithme peut se résumer à travers les étapes suivantes pour suivre une zone :

1. Sélection d'une zone d'intérêt ou d'un motif de référence, et calculs des descripteurs SURF et SIFT sur cette zone.
2. Les descripteurs SIFT et SURF sont ensuite calculés sur chacune des images refocalisées, afin de déterminer une sélection d'images, où le motif de référence est retrouvé, grâce à la mise correspondance des descripteurs.
3. Enfin, le critère de Tenengrad (cf. Section 4.4) est calculé au sein de la surface du motif, identifié dans la pile des images sélectionnées. Cette dernière étape permet de suivre l'évolution du critère de Tenengrad, suivant les profondeurs correspondantes aux images sélectionnées.

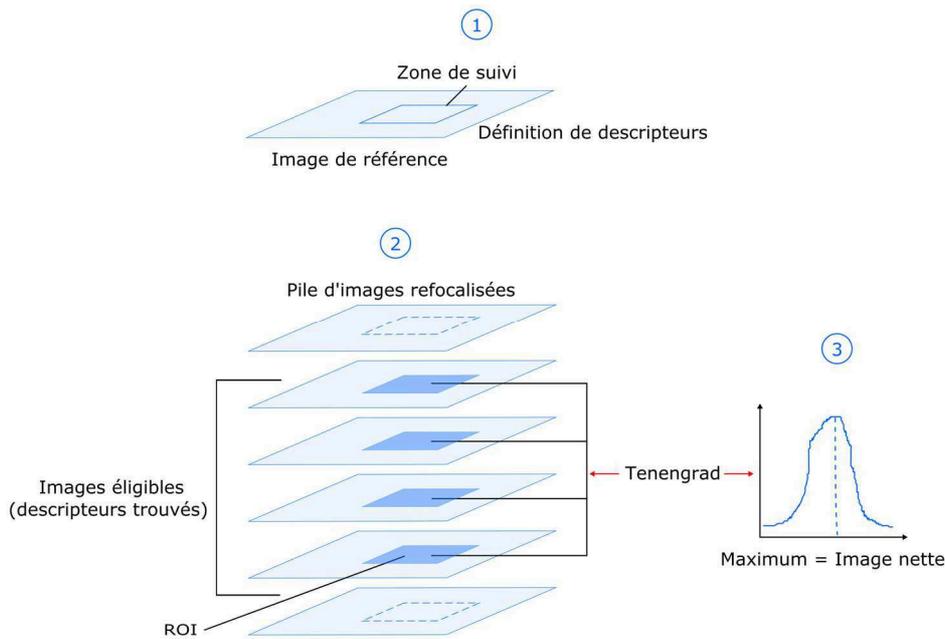


Figure 4.18 Explication de la méthode de suivi, basée sur Tenengrad. La première étape consiste à déterminer les images éligibles dans la pile d'images refocalisées à l'aide de la recherche de points caractéristiques en utilisant les méthodes SIFT ou SURF. Pour chacune des images éligibles, le critère de Tenengrad est calculé sur la ROI définie. Enfin, le résultat sélectionné correspond à l'image pour laquelle la valeur est maximale. Dans ce cas, l'image est nette.

L'évolution de la mesure de flou avec notre méthode de suivi basée sur le critère de Tenengrad est représentée pour les deux ROI sur la Figure 4.19. La pile d'images est calculée avec un pas de 1 mm. Une quinzaine d'images sont éligibles pour chacune de deux ROI. La profondeur recherchée correspondant aux maximums de ces courbes est égale à la distance entre la caméra et le plan contenant la ROI. La valeur maximale du score pour la première ROI est trouvée pour une profondeur égale à 29,9 cm et pour la seconde ROI de 34,4 cm. La hauteur mesurée avec notre méthode entre les plans des deux ROI est donc égale à 4,5 cm, ce qui équivaut à la hauteur de l'objet. L'erreur systématique de mesures est de l'ordre de 0,9 mm.

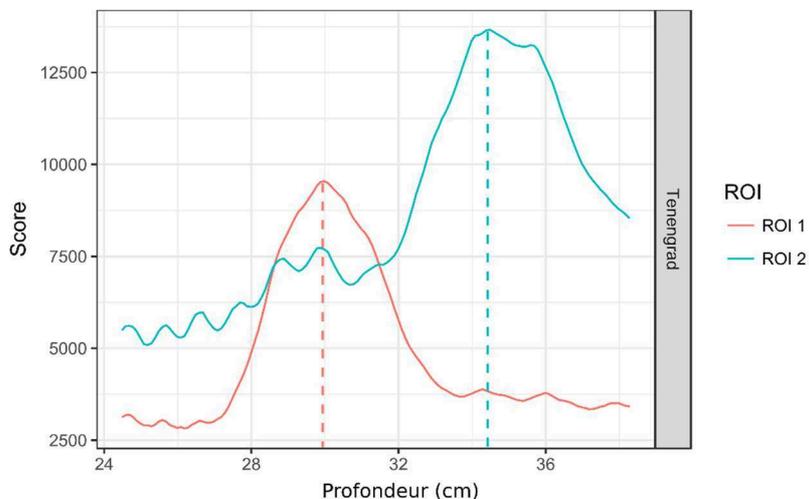


Figure 4.19 Evolution de la mesure du flou, pour les deux zones d'intérêts de la pièce mécanique, obtenue avec la méthode de suivi basée sur le critère de Tenengrad.

Mesure 3D par matching avec la méthode utilisant un logiciel de vision industrielle NI Builder

Le principe de mesure consiste en la comparaison de la ressemblance entre la ROI de l'image de référence et celle de l'image refocalisée étudiée. Dans ce cas, les images refocalisées sont les données d'entrée du logiciel. Cette expérimentation a pour but de montrer que les images refocalisées calibrées en profondeur peuvent être utilisées directement avec un logiciel de vision industrielle afin d'effectuer des mesures (Figure 4.20). L'objet d'étude est la pièce mécanique étalon, utilisée lors de l'étude des critères de mesure du degré de flou, présentée dans la partie 4.4.

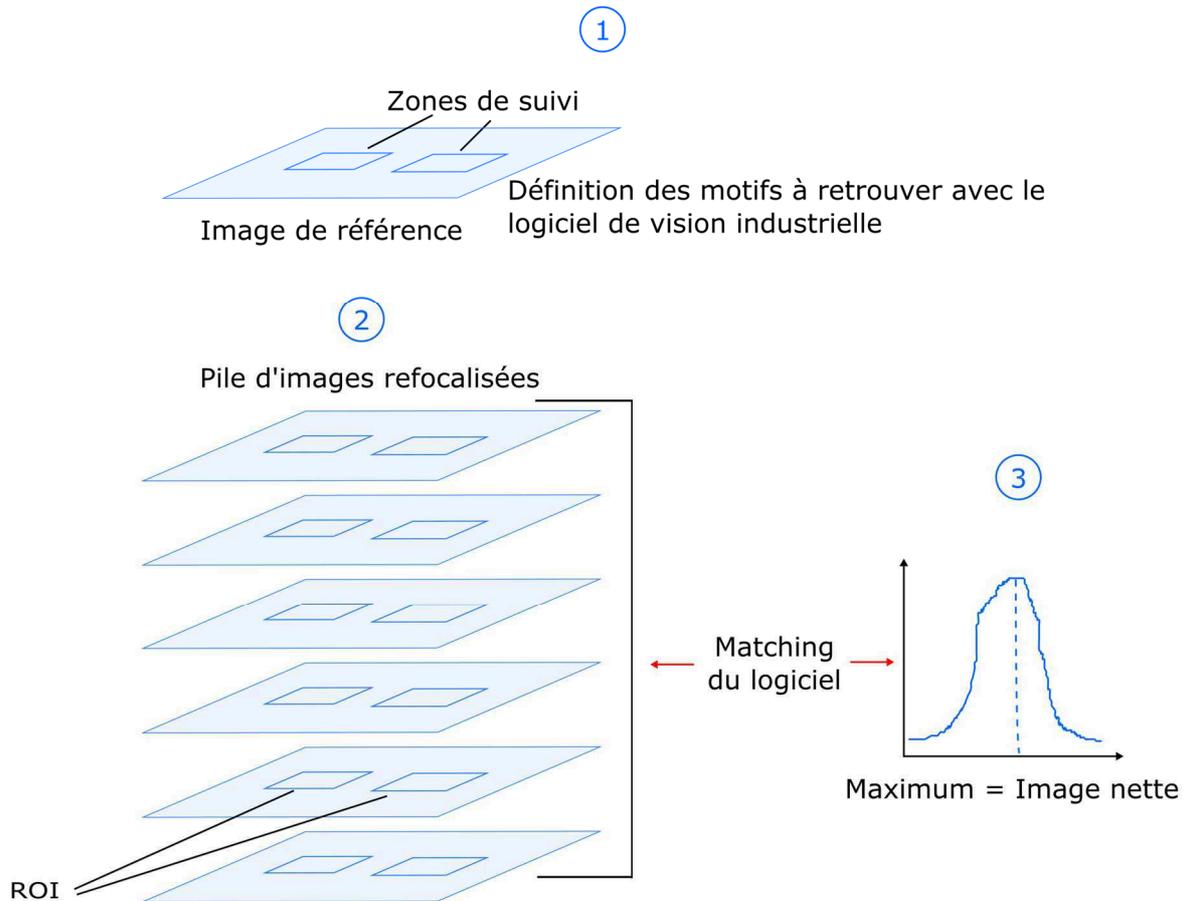


Figure 4.20 Explication de la méthode de suivi, utilisant le logiciel de vision industrielle NI Vision Builder. La première étape consiste à définir les motifs à rechercher sur une image de référence. Puis, dans la pile d'images, une mesure de correspondance par rapport aux zones de référence est effectuée avec le logiciel. Le résultat sélectionné correspond à l'image pour laquelle la valeur est maximale. Dans ce cas, l'image est nette.

Les résultats obtenus sont représentés sur la Figure 4.21. Les ROI sont identiques à celles définies précédemment dans la partie 4.4. La valeur maximale de la première ROI est obtenue pour une profondeur de 30,08 cm et pour la seconde ROI, la profondeur correspondant au maximum de la courbe est égale à 34,68 cm. La hauteur mesurée entre les deux plans est donc de 4,60 cm contre 4,5 cm pour l'étalon.

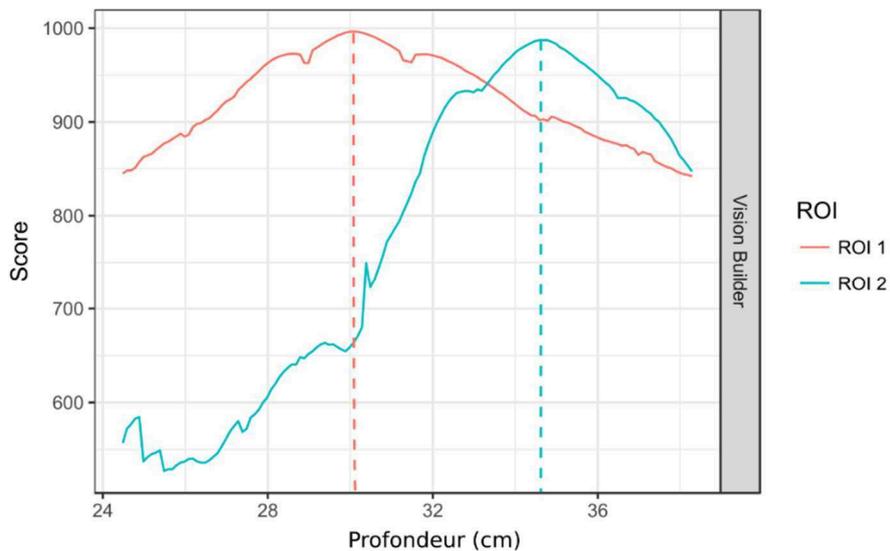


Figure 4.21 Evolution de la mesure de correspondance, pour les deux zones d'intérêts de la pièce mécanique, obtenue en utilisant le logiciel NI Vision Builder.

Cette expérimentation est réitérée sur un autre objet d'étude : une boîte de vitesse (Figure 4.22), correspondant à un cas plus réaliste. Dans ce cas, la mesure de la distance est effectuée entre deux zones d'intérêt visible sur la Figure 4.23. Ces dernières correspondent aux deux motifs à rechercher dans la pile d'images refocalisées, calculée entre 24,6 cm et 40 cm avec un pas de 0,1 cm. Les images refocalisées de la Figure 4.23 le sont autour de ces deux plans. L'erreur systématique de mesures est de l'ordre de 0,8 mm.



Figure 4.22 Boîte de vitesse. A gauche : une vue générale de l'objet. A droite : un zoom sur la hauteur à mesurer.

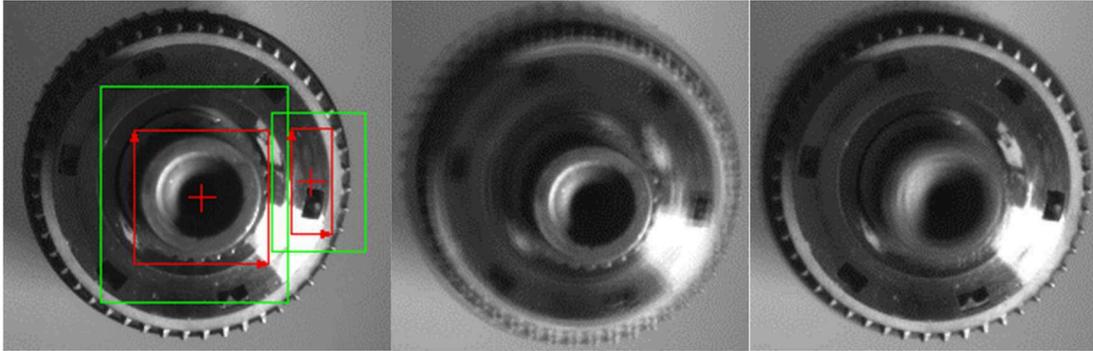


Figure 4.23 De gauche à droite : Image de référence avec les deux motifs (en rouge) de deux régions d'intérêt (en vert) de la boîte de vitesse ; image refocalisée sur la ROI du haut de la pièce et image refocalisée sur la ROI du bas de la pièce.

La Figure 4.24 représente l'évolution du score de correspondance entre le motif recherché et celui présent dans les images refocalisées. Elle montre que le maximum pour la ROI 1 (au-dessus) est obtenu pour une profondeur de 28,18 cm et pour la seconde, la distance est de 34,37 cm. Ainsi, la hauteur entre les deux ROI est de 6,19 cm contre une valeur mesurée avec des cales étalon de 6,22 cm. La mesure est conforme aux expérimentations précédentes, avec une mesure effectuée par pas de 1 mm.

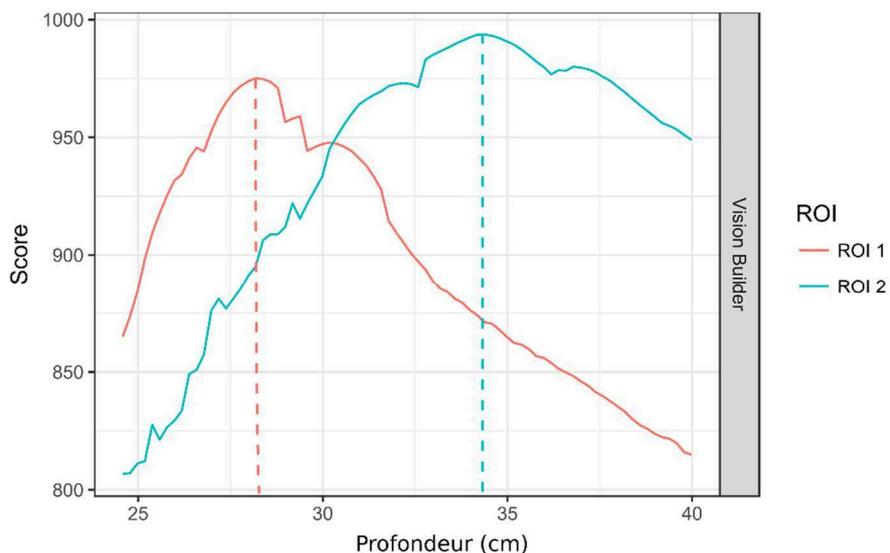


Figure 4.24 Evolution de la mesure de correspondance pour les deux zones d'intérêts de la boîte de vitesse. En rouge, la courbe représente les mesures pour la ROI 1 de la Figure 4.23 et en bleu, la courbe correspond aux résultats obtenus pour la ROI 2 (rectangle sur la Figure 4.23).

Comparaison entre les deux méthodes précédentes

Les deux précédentes expérimentations montrent que les performances en termes de mesures sont comparables. L'intérêt de la méthode se basant sur les mesures de focalisation repose sur le fait qu'un dispositif de référence n'est pas nécessaire. Ainsi, elle est compatible avec n'importe quel dispositif capable de générer des images refocalisées. Concernant la méthode basée sur le « *matching* », son intérêt réside dans sa capacité à exploiter les algorithmes de « *matching* » existants dans la plupart

des systèmes de vision industrielle. C'est un intérêt pour un déploiement aisé de cette approche 3D basée sur la refocalisation numérique.

4.5.2.2. Mesure 3D de diamètres à l'aide du critère de Tenengrad

Une expérimentation est effectuée avec un objet non plan correspondant à un composant mécanique (Figure 4.25). Le but est de mesurer le diamètre de deux cercles localisés à différentes profondeurs. Avec une caméra classique, les mesures peuvent être réalisées en pixels. Cependant, la correspondance métrique n'est pas possible tant que la localisation en profondeur Z des diamètres à mesurer est inconnue. Nous proposons d'utiliser une pile d'images refocalisées, et d'estimer la localisation en profondeur Z des cercles en mesurant le degré de focalisation des ROI choisies, proches de ces cercles.

Pour cet exemple, les images refocalisées de la pile sont générées pour une gamme de profondeur comprise entre 32,5 cm et 45,5 cm, avec un pas de 0,1 cm. Les images refocalisées sur le haut et le bas de la pièce sont présentées sur la Figure 4.25. Cette figure montre aussi la localisation des ROI 1 et 2. L'erreur systématique est de l'ordre de 1,4 mm.

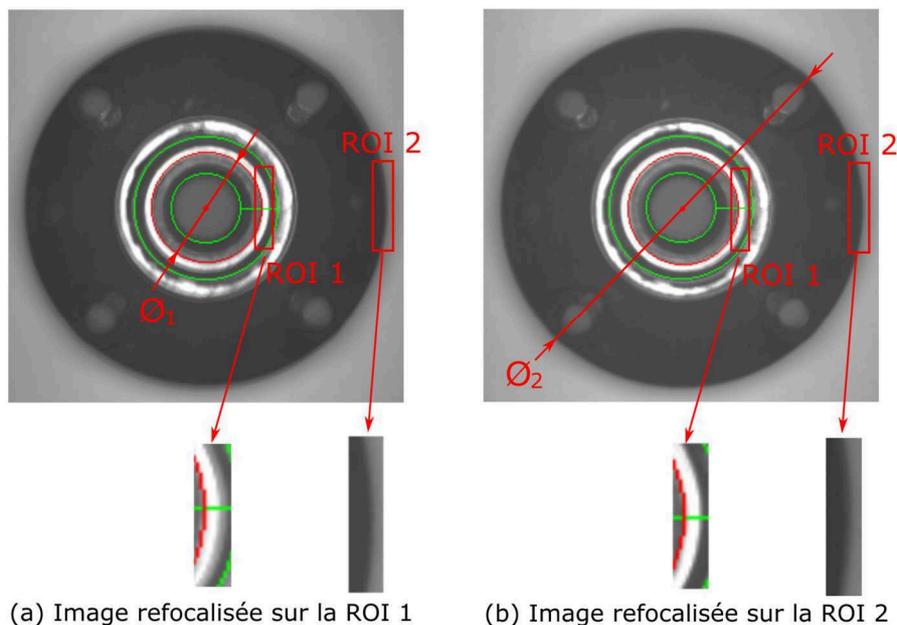


Figure 4.25 Images refocalisées et zooms sur les ROI respectivement à 41 cm et à 37 cm. La différence de focalisation est, dans ce cas, difficile à identifier. Elles peuvent être distinguées grâce à l'évolution de la mesure de la focalisation, calculée pour les deux ROI 1 et 2.

Les mesures de focalisation, avec la méthode utilisant le critère de Tenengrad, sont effectuées pour chacune des ROIs sur l'ensemble de la pile d'images. Les résultats sont reportés sur la Figure 4.26. Ces courbes indiquent la meilleure mesure trouvée pour les images refocalisées calculées à une profondeur $Z_1 = 42$ cm pour la ROI 1 et $Z_2 = 37$ cm pour la ROI 2. Avec ces valeurs, et en utilisant les paramètres intrinsèques préalablement déterminés pour la sous-image de référence, les diamètres des cercles correspondants peuvent aisément être convertis en unité métrique :

$$\phi_1 = 78,7 \text{ pixels} \rightarrow \phi_1 = 35,0 \text{ mm}$$

$$\phi_2 = 256,7 \text{ pixels} \rightarrow \phi_2 = 129,4 \text{ mm}$$

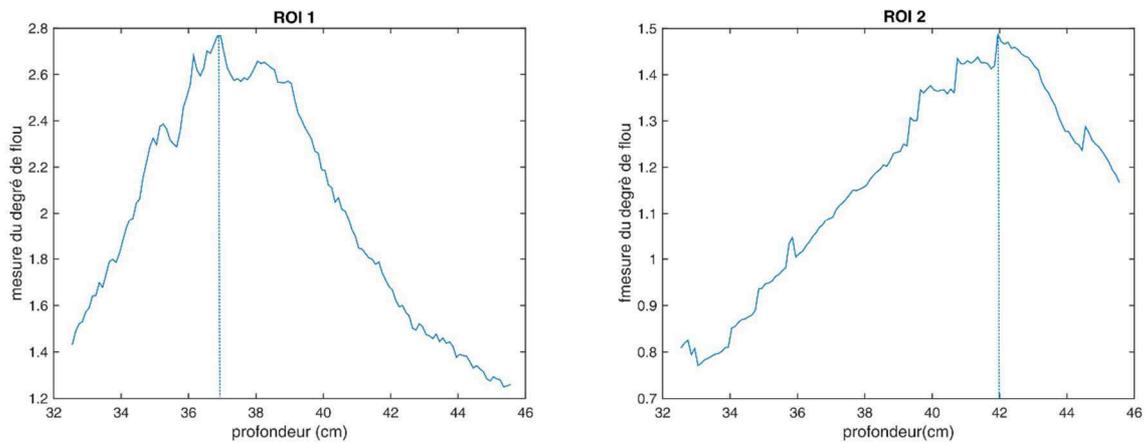


Figure 4.26 Evolution de la mesure de focalisation pour chacune des ROI, appliquée sur la pile d'images refocalisées entre 32,5 cm et 42,5 cm.

Bilan

Dans cette partie, nous avons exposé quelques exemples montrant comment l'information de profondeur inhérente aux images refocalisées peut être exploitée pour le contrôle par vision. Le fait de pouvoir mettre en évidence un plan de travail dans une scène est un nouveau concept, qui pourrait s'exporter dans divers cas d'usage.

Le paragraphe suivant étend les applications au domaine de la conduite automobile au travers d'un exemple de détection de piétons. Le principe d'images refocalisées et de détermination de la profondeur sont exploités pour la détection d'objets et la caractérisation de la scène routière en zones fonction de l'information de profondeur. Cette approche s'appuie sur le dispositif multi-caméras décrit au Chapitre 2.

4.5.2.3. Application liée à l'automobile

Dans ce paragraphe, nous montrons que les images refocalisées peuvent être utilisées dans un contexte différent de la vision industrielle, comme par exemple celui de l'automobile et en particulier l'aide à la conduite. Pour ces dispositifs, l'étape de perception de l'environnement est essentielle et s'appuie fréquemment sur des techniques d'imagerie. Ainsi, nous proposons dans cette section d'illustrer l'intérêt de la refocalisation d'images pour la détection d'obstacles. Cette expérimentation est menée avec ARTEMIPS, le véhicule à conduite autonome de notre laboratoire. Pour cet exemple, afin de disposer d'une *baseline* plus importante entre les vues, nous utilisons le système multi-caméras (Figure 4.27 droite), qui a été présenté dans le chapitre 2, partie 2.3.2.



Figure 4.27 Prototype multi-caméras, présentés dans le Chapitre 2. A gauche : système multi-caméras positionné dans le véhicule ARTEMIPS, véhicule expérimental du laboratoire. A droite : vue du dispositif depuis l'avant du véhicule ARTEMIPS.

Le principe de l'algorithme consiste à coder les pixels contours de l'image en fonction de leur distance par rapport à la caméra, puis à distinguer les pixels situés au sol des autres. Pour cela, le critère de Tenengrad est appliqué sur un jeu de n images refocalisées, afin d'attribuer une classification en distance pour chaque contour de l'image. La Figure 4.28 montre qu'au sol, la projection perspective permet de découper une image en bandes horizontales, où chaque bande est représentative d'une plage de profondeur. Grâce à la géométrie du système (Figure 4.29), les objets situés hors sol sont facilement isolés, si l'on considère qu'un objet se trouvant dans une bande de profondeur différente de la plage qui lui a été attribuée, est un objet « hors sol » et constitue donc un obstacle sur la route.

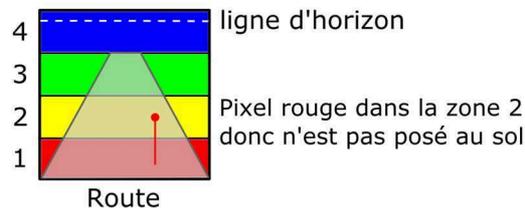


Figure 4.28 Représentation perspective d'une route sur une image découpée en quatre zones de profondeur. Le pixel rouge présent dans la zone 2 traduit qu'il n'est pas posé au sol.

Pour cet exemple illustré ici, nous avons choisi un découpage en quatre zones, nécessitant de ne calculer que quatre images refocalisées aux distances-seuils Z_1 , Z_2 , Z_3 et Z_4 , qui sont mentionnées sur la Figure 4.29. Un objet, positionné dans la zone jaune d'une certaine hauteur (supérieure au seuil défini), est détecté comme un obstacle en utilisant les pixels « hors-sol » de cet élément et présents dans la zone verte.

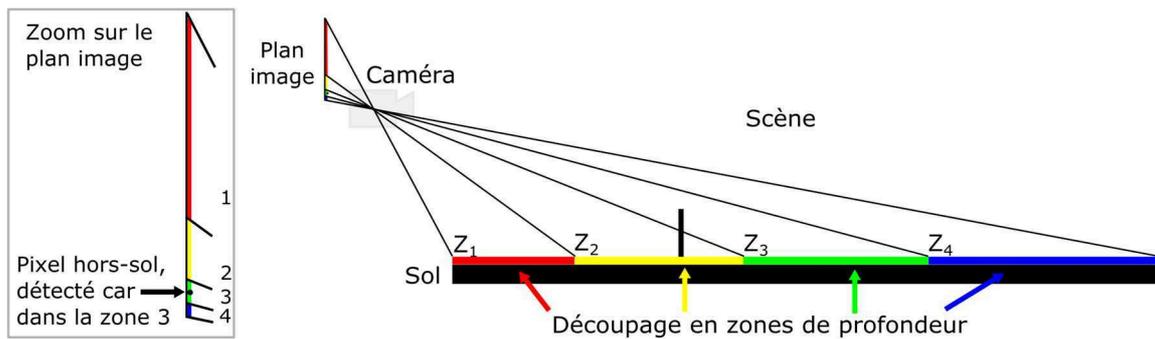


Figure 4.29 Découpage de la scène en 4 zones de profondeur et sa représentation sur l'image. Le sommet de l'obstacle de la zone 2 (jaune) au sol, se trouve projeté dans la zone 3 (verte) sur la caméra. Cette intrusion caractérise la présence d'un obstacle sur la route.

Les étapes de l'algorithme sont résumées ci-dessous :

1. Extraction des contours de la scène, sur l'image prise comme référence pour la refocalisation numérique. Chaque pixel contour est considéré comme un marqueur.
2. Calcul d'un jeu de n images refocalisées (ici $n=4$).
3. Calcul du critère de Tenengrad sur les n images, autour des positions de chaque pixel contour.
4. Chaque pixel contour est codé avec un numéro de plage, correspondant à l'index de l'image où le meilleur score du critère Tenengrad est trouvé. Le meilleur score donne ainsi une estimation approximative de la distance caméra-objet.
5. Pour chacune des bandes de l'image, les pixels contours portant un numéro identique à leur index de bande sont éliminés. Ainsi, seuls les contours « hors-sol » sont détectés sur la chaussée.

Le jeu d'images refocalisées à quatre distances différentes est présenté sur la Figure 4.30. Les plots y ont été positionnés tous les cinq mètres à partir de la position du système multi-caméras. Les zones de profondeur considérées dans cette expérimentation sont $Z_1 = 5$ à 15 m, $Z_2 = 15$ à 25 m, $Z_3 = 25$ à 35 m et $Z_4 \geq 35$ m. Ces zones ont été définies selon les portées de détection nécessaires pour certaines applications de détection d'obstacles, telles que l'aide au stationnement ou le guidage à basse vitesse, les dispositifs de pré-collision ou encore la régulation de vitesse intelligente jusqu'à 90 km/h [Perr08].



Figure 4.30 Les quatre images refocalisées utilisées comme données d'entrée pour associer chaque contour à une des quatre zones de profondeur. En haut : à gauche, image refocalisée à une profondeur de 10,22 m et à droite image refocalisée à une profondeur de 17,72 m. En bas : à gauche, image refocalisée à une distance de 31,22 m et à droite image refocalisée à une profondeur de 35,22 m.

Le résultat de l'étape 4 est présenté sur la Figure 4.31. Le découpage des pixels de l'image en quatre zones, obtenu à partir de l'estimation de la distance caméras-objet est clairement visible.

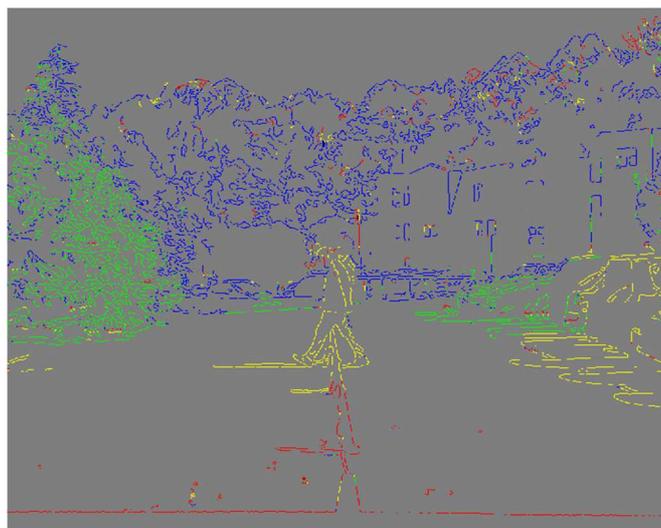


Figure 4.31 Découpage en quatre zones de profondeur des contours de l'image de référence.

Au final, cet exemple permet d'éliminer toutes les marques au sol situées sur la chaussée, et ainsi mettre en évidence les objets « hors-sol » devant le véhicule (cf. Figure 4.32 gauche). La hauteur de détection (pour cette figure est fixée à 30 cm), et les distances des obstacles sont ajustables grâce aux distances seuils Z_1 , Z_2 , Z_3 et Z_4 . Par exemple, la Figure 4.32 droite, montre un autre exemple où le seuil de détection est placé à 1 mètre. Nous montrons ainsi que l'information de profondeur d'un dispositif *light-field* peut être exploitée par le biais des images refocalisées, ce qui est une alternative aux approches rencontrées, qui se basent sur l'exploitation d'une carte des profondeurs.

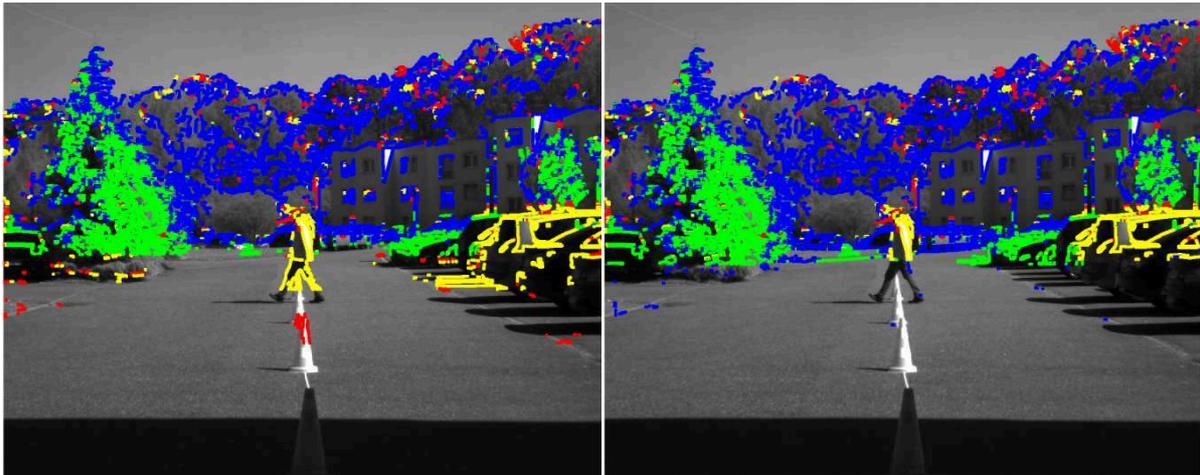


Figure 4.32 Images résultat de la détection d'obstacles : mise en valeur des obstacles sur la vue de référence du système. A gauche : seuil de détection d'obstacles fixé à une hauteur de 30 cm. A droite : seuil de détection d'obstacles fixé à une hauteur de 1 m. Les plots n'apparaissent pas comme des obstacles du fait du seuil choisi pour déterminer la position du sol au niveau des quatre zones dans l'image.

4.5.3. Conclusion

En exploitant la modélisation présentée dans le Chapitre 3, une méthode alternative et simple de calcul d'images refocalisées et calibrées en profondeur a été proposée. Ces images présentent plusieurs intérêts pour des secteurs applicatifs variés. Par exemple, en métrologie optique, la possibilité de choisir un plan de travail permet de combiner à la fois un contrôle de profondeur (hauteur) d'un objet et d'autres contrôles conventionnels à partir d'une ou plusieurs images refocalisées.

Une étude de la mesure de flou à partir des méthodes connues a été menée afin de déterminer l'image la plus adaptée aux images refocalisées. Cette sélection permet d'effectuer une mesure précise de la profondeur/ hauteur d'un objet dans une scène. Enfin, une utilisation de ces images afin de détecter les obstacles d'une scène automobile et de la catégoriser en quatre zones de profondeur a été proposée.

4.6. Images de profondeurs par propagation de contours

4.6.1. Introduction

Cette partie présente un autre volet de nos travaux. On travaille ici par couple de deux images pour générer des images de disparité suivant un principe de propagation de distances par rapport à un contour. Nous verrons que la méthode s'applique à des configurations simples, mais elle se démarque des approches connues par le fait qu'aucune mesure de similarité ne soit nécessaire. Avec les systèmes multi-vues, l'estimation de la profondeur est liée à la disparité entre deux pixels identiques projetés sur des coordonnées différentes dans la vue de référence. Cette tâche est souvent associée à une étape d'identification de la similarité entre les différentes vues, permettant de calculer les déplacements de points identifiés et de leurs positions tridimensionnelles (3D) correspondantes. L'identification de ces similarités est le plus souvent une tâche coûteuse en temps et est référencée en tant que méthodes de mise en correspondance, basée sur des approches locales ou globales [ScSz02]. Les méthodes locales comparent deux régions d'intérêt en mesurant les similarités avec un ou des critères bien connus tels que le *block matching* ou les approches d'extraction de caractéristiques (*feature approaches*) [Szel11]. Ces approches sont efficaces sur les images contenant un large pourcentage de zones texturées, mais échouent pour les régions uniformes, parce qu'elles contiennent trop peu d'informations pour le processus de mise en correspondance. La présence de régions occultées, dues à la différence de point de vue est une autre cause d'erreurs apparaissant lors de la mise en correspondance. Les approches globales peuvent surmonter certains de ces problèmes, en estimant la carte de disparité par minimisation du critère d'énergie sur l'image entière. Les méthodes *graph-cuts* [KoZa01], *belief propagation* [TsCC07] et de programmation dynamique [Veks05] sont les méthodes globales les plus connues. La complexité computationnelle de ces méthodes est souvent élevée, et elles ne sont pas capables de résoudre correctement les cas où les images contiennent de larges zones uniformes. Pour la robotique et le contrôle industriel, des motifs lumineux sont projetés sur les objets et le fond de la scène afin d'aider à la mise en correspondance. Dans ce cas, les motifs modifient l'aspect des objets homogènes, pour qu'ils semblent texturés. Ces méthodes actives sont efficaces pour le calcul de cartes de profondeur. Mais elles rendent les images inexploitable pour les autres contrôles tels que la détection de contours, la mesure d'intensité, la lecture de code barre ou la reconnaissance optique de caractères.

La méthode s'appuie sur un principe de propagation de contours. Elle est donc efficace pour des scènes assez simples (quelques objets sous une caméra) et plus performante en illumination naturelle. Il ne s'agit pas de concurrencer les principes existants, mais plutôt de poser les bases d'une approche différente, qui présente l'avantage d'être très peu gourmande en termes de calculs, au détriment de la résolution de mesures. Dans un second temps, le calibrage effectué avec l'homographie variable permettra d'établir une correspondance entre disparité et profondeur.

4.6.2. Calcul de la disparité avec le système à quatre mini-lentilles

Avec l'ensemble des images rectifiées à la distance de calibrage Z_c , nous proposons de calculer l'image de disparité (ou carte de disparité) correspondante, afin d'utiliser notre prototype comme une caméra de profondeur.

L'idée principale de cette solution est de soustraire les cartes de distances entre les sous-images afin d'obtenir la disparité. La méthode proposée est une méthode rapide de calcul, capable de fonctionner dans un environnement lumineux ambiant, et avec des objets uniformes.

4.6.2.1. Principe

La méthode repose sur la mesure de disparité entre un jeu de deux images : une image de référence et une autre image rectifiée. Nous verrons à la fin du paragraphe comment fusionner l'information si on travaille avec les quatre vues du système.

Nous proposons un algorithme composé de cinq étapes principales (les numéros d'étapes indiqués correspondent à ceux de la Figure 4.33) opérant entre deux sous-images :

- Extraction des contours binaires sur les deux images (étape 2);
- Calcul des Images de Distance ID^{xx} (étape 3) ;
- Calcul des Images de Pseudo-Soustraction IPS^{xx} (étape 4);
- Application d'un processus de fusion permettant d'obtenir une image IF (étape 5) ;
- Application d'un processus de décodage de l'Image Fusion (étape 6) ;
- Calcul d'une carte de disparité en utilisant les valeurs obtenues lors de l'étape précédente, (étape 7).

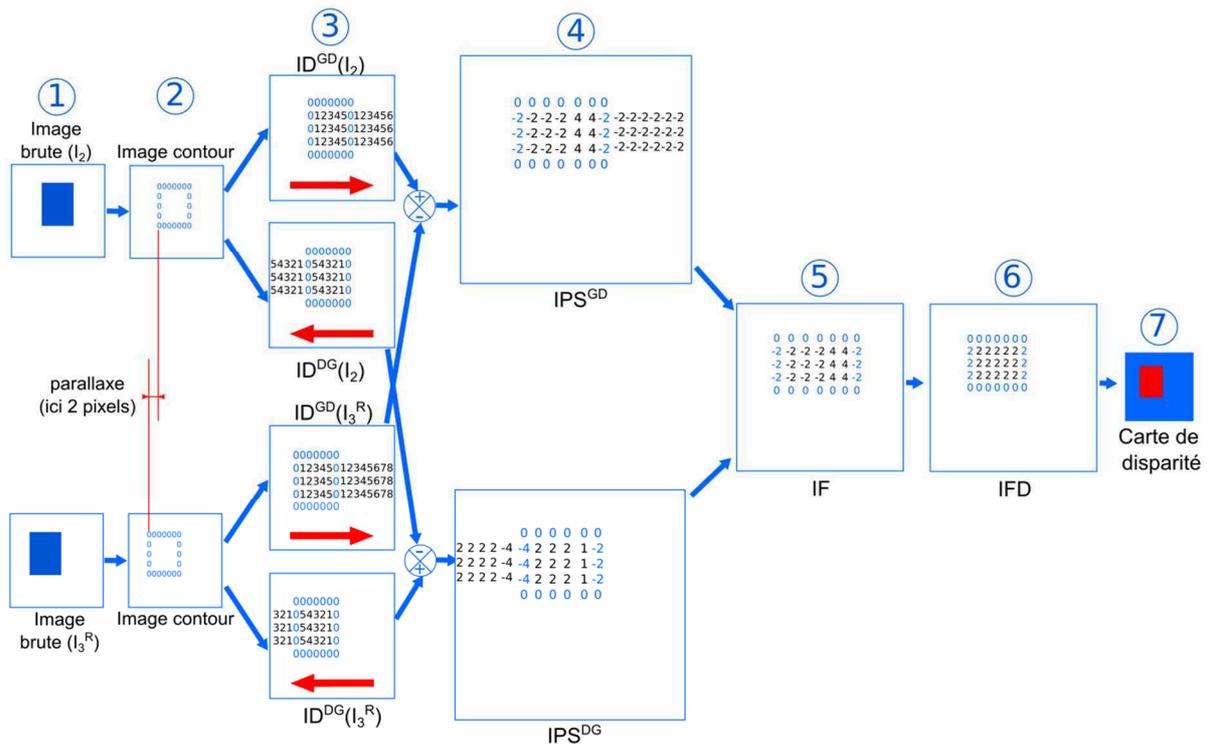


Figure 4.33 Etapes utilisées pour calculer une carte de disparité à partir des sous-images I_2 et I_3^R

Nous décrivons ci-après ces différentes étapes de traitement.

Images de distance ID^{XX}

Après l'étape d'acquisition et de seuillage, la troisième étape permet de générer les Images de Distance (ID) suivant des directions prédéterminées. Quand les sous-images utilisées sont horizontales, ID^{GD} (Image de Distance de Gauche à Droite) et ID^{DG} (Image de Distance de Droite à Gauche) sont calculées. De la même manière, les images de distance Haut/Bas sont générées quand les sous-images verticales sont utilisées. En suivant ce raisonnement, une extension suivant la diagonale est possible.

Les Images de Distance sont constituées des valeurs indiquant la distance en pixel entre les pixels et le précédent contour rencontré suivant la direction de balayage. Le marqueur x est utilisé tant qu'aucun contour n'a été rencontré au cours du balayage. La Figure 4.34 illustre des exemples de calcul des images ID^{GD} et ID^{DG} à partir des vues I_2 et I_3^R . Dans cet exemple, un objet rectangulaire est représenté par ces pixels contours, en rouge. Une disparité d (ou une parallaxe d_u) de 4 pixels est simulée et les points verts P_1 et P_2 représentent un point ayant les mêmes coordonnées dans les deux sous-images originelles. La propriété principale utilisée par cette méthode est basée sur le fait que la différence $ID^{GD}(I_2) - ID^{GD}(I_3^R)$ ou $ID^{DG}(I_2) - ID^{DG}(I_3^R)$ pour chaque pixel dans un objet dépend de la valeur de la disparité d .

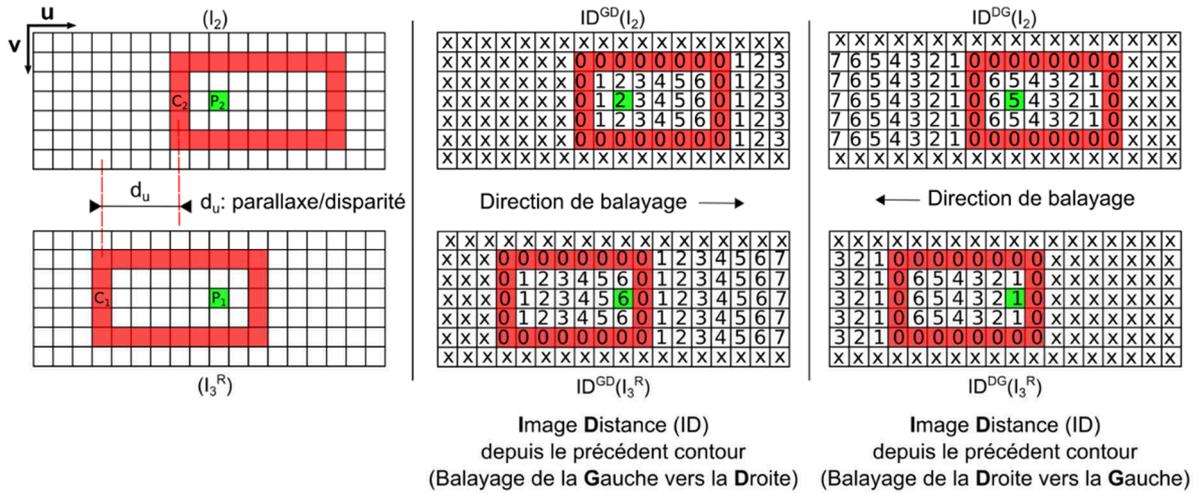


Figure 4.34 Calcul des images distances ID^{GD} et ID^{DG}

Images de Pseudo-Soustraction IPS^{XX}

Le résultat de l'étape 4 est basé sur la notion de pseudo-soustraction. Nous avons défini deux opérateurs distincts pour exploiter la propriété selon laquelle la soustraction de deux images de distance ID fournit la disparité. Cette propriété n'est pas vérifiée sur toute l'image, notamment quand le marqueur x est rencontré, d'où le terme de « pseudo » que nous allons utiliser. Nous avons établi que le meilleur résultat, où l'objet est rempli par la valeur de la disparité, est obtenu en utilisant deux types de soustraction appelés ici pseudo-soustraction, pour lesquels la valeur résultat peut être le marqueur x, le premier opérande ou la vraie différence. Ces opérateurs sont définis avec les arguments a et b, avec a représentant la valeur du pixel de l'image de référence et b la valeur de la seconde image utilisée. Les définitions de ces opérateurs sont pour les images IPS^{GD} :

$$IPS^{GD}(a, b) \rightarrow \begin{cases} (a - b) & \text{si } a, b \in \mathbb{N} \\ \times & \text{sinon} \end{cases} \quad (4.11)$$

Et pour les images IPS^{DG} :

$$IPS^{DG}(a, b) \rightarrow \begin{cases} (a - b) & \text{si } a, b \in \mathbb{N} \\ a & \text{si } b = \times \\ \times & \text{sinon} \end{cases} \quad (4.12)$$

Un exemple d'images IPS générées avec ces opérateurs est fourni par la Figure 4.35. Le résultat le plus important concerne les valeurs des images IPS^{GD} , et plus spécifiquement celles dans l'objet.

Image Fusion IF

La cinquième étape, la fusion utilise les images IPS^{GD} et IPS^{DG} pour remplacer sur les images IPS^{GD} , les valeurs -d apparaissant en dehors des objets par le marqueur x. Cette opération est simplement définie par :

$$IF(a, b) \rightarrow \begin{cases} a & \text{si } a, b \in \mathbb{N}, \\ \times & \text{sinon} \end{cases} \quad (4.13)$$

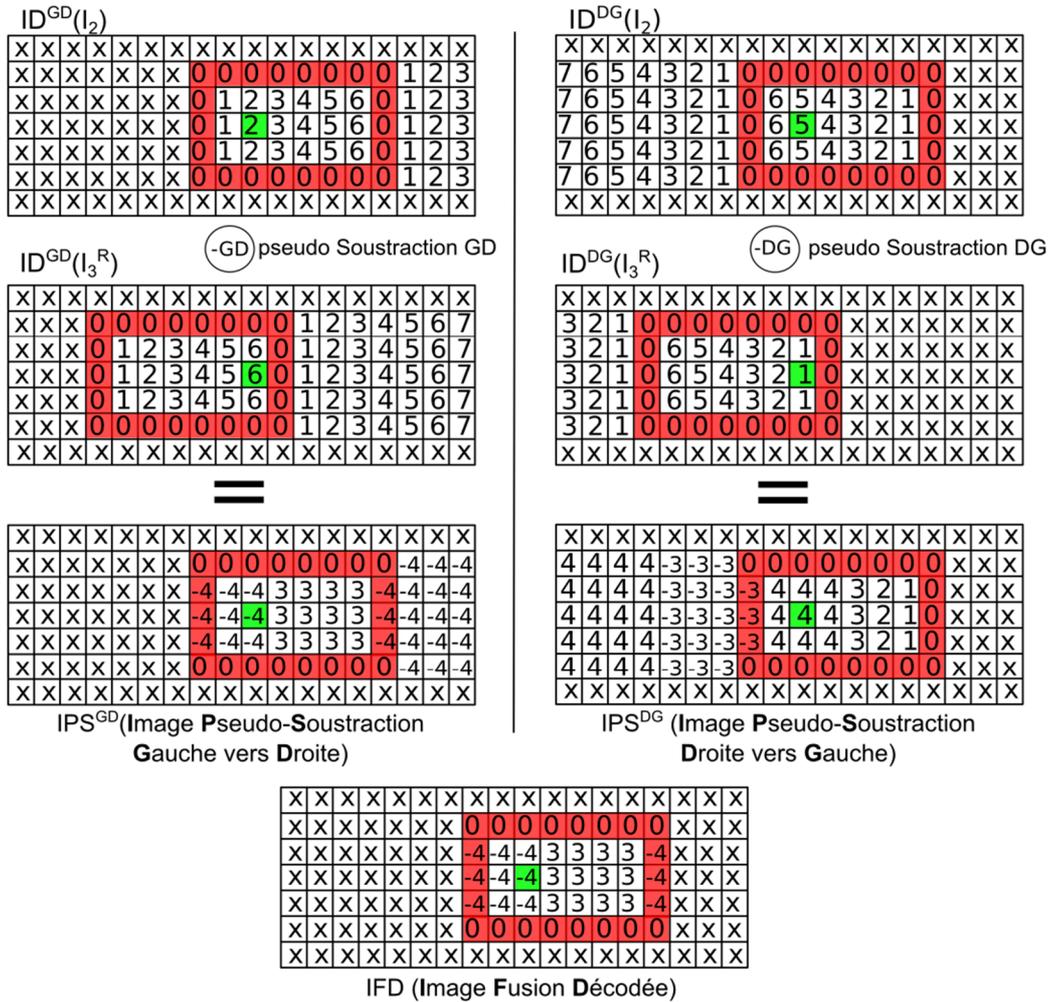


Figure 4.35 Exemples de pseudo-soustraction $IPSGD$ et $IPSDG$ pour un objet mesurant une largeur $T=8$ et pour une disparité $d=4$. Sur les images $IPSGD$, les valeurs trouvées sont $-d=-4$ et $T-d-1=3$.

Décodage de l'image fusion IFD

Le but de cette sixième étape est de décoder le contenu de l'image fusion afin d'identifier chaque objet par sa disparité, sa position et sa taille. Grâce aux étapes précédentes, plus spécialement celles de la propagation de la distance et de la pseudo-soustraction, une description complète des objets contenus dans les sous-images d'origine peut être extraite à partir de la valeur des pixels de l'image fusion. Ces valeurs sont constituées de combinaisons de positions, tailles et disparités des objets. Pour N objets séparés, nous avons établi par simulations que chaque valeur de pixel peut être décodée. Trois grandeurs A_i , B_i et C_i peuvent être définies pour chaque objet i :

$$A_i = -d_i \quad (4.14)$$

$$B_i = l_i - d_i - 1 \quad (4.15)$$

$$C_i = p_i - d_i - p_{i-1} - l_{i-1} + 1 \quad (4.16)$$

Où d_i , l_i et p_i sont respectivement la disparité, la taille et la première position de l'objet i .

Ces trois variables sont présentes dans l'image fusion suivant deux types de séquences définies. Le décodage commence toujours suivant la même séquence Seq_1 puis on retrouve la séquence Seq_i au tant de fois que d'objets présents dans l'image. Ces deux séquences sont définies comme suit :

$$Seq_1 = |A_1 \dots A_1 B_1 \dots B_1 A_1 \dots A_1| \quad (4.17)$$

$$Seq_i = |C_i \dots C_i A_i \dots A_i B_i \dots B_i A_i \dots A_i| \quad (4.18)$$

Chaque ligne de l'image fusion est ainsi constituée de l'arrangement de séquences suivant :

$$[\times \dots \times; Seq_1; Seq_2; \dots Seq_i; \times \dots \times] \quad (4.19)$$

Le nombre d'occurrences de chaque terme A_i , B_i et C_i à l'intérieur de chaque séquence dépend de la disparité, de la taille et de la position de chaque objet dans la sous-image. Cette information n'est pas utilisée pour décoder l'image fusion. Par exemple, dans la Figure 4.35, avec $d_1=4$ et $l_1=8$, où l'image fusion est constituée de l'arrangement suivant, il est possible de reconnaître la séquence Seq_1 .

$$[\times \dots \times; -4; -4; -4; 3; 3; 3; 3; -4; \times \dots \times] \quad (4.20)$$

Enfin, le rôle de cette étape de décodage est d'utiliser cette propriété pour générer les cartes de disparité.

Carte de disparité

Le but de l'étape finale est de remplir les objets avec la valeur de la disparité. A partir des conclusions précédentes, la disparité est extraite à partir des valeurs $-d_i$ issues du décodage. La carte de disparité est finalement générée en remplaçant chaque pixel à l'intérieur de l'objet i par la valeur d_i .

Carte de profondeur

Les images de profondeur sont ensuite établies grâce au calibrage présenté précédemment, et à l'expression suivante reliant la profondeur et la disparité :

$$d = \frac{bf_1}{Z} - \frac{bf_1}{Z_c} \quad (4.21)$$

Cette méthode peut donc être également appliquée au système de stéréovision. Ainsi, comme notre dispositif est composé de quatre vues, il est possible de fusionner les cartes de disparité ou de profondeur obtenues pour chacun des couples d'images associés à l'image de référence. La carte de profondeur est aussi appelée image de profondeur, notée IP. Les plages de mesures dépendent de la profondeur de calibrage, comme illustré par la Figure 3.11 du chapitre 3. De plus, la disparité est mesurée par pas de 1 pixel, ce qui équivaut à échantillonner la profondeur et de mesurer la profondeur par palier.

4.6.2.2. Combinaison des cartes de disparités ou des cartes de profondeur

Comme trois cartes de disparité peuvent être calculées à partir de l'image de référence, une étape finale fusionne ces trois cartes afin d'éliminer d'éventuels artéfacts. L'algorithme présenté permet de calculer une carte de disparité à partir de deux sous-images. Une vue complète d'ensemble est fournie, résumant toutes les connexions entre les étapes, et est représentée sur la Figure 4.36. Ce schéma fonctionne pour n'importe quel couple de sous-images. Comme cette approche est uniquement basée sur les contours détectés dans chaque sous-image, la qualité des contours initialement détectés est essentielle. Leur détection constitue une étape initiale et primordiale de prétraitement. Aucune étape de mise en correspondance n'est effectuée entre les sous-images, ce qui constitue un des avantages de cette approche. Même si la mise en correspondance pourrait être utile pour s'assurer que chaque point contour dans la sous-image i dispose bien d'un point correspondant dans la sous-image $i+1$, nous avons décidé de ne pas effectuer cette étape. Cependant, sans processus de mise en correspondance, certains contours de la sous-image i n'ont pas de correspondants dans la sous-image $i+1$. Les images ID peuvent en être affectées, et ainsi des artéfacts dans les images de disparité peuvent apparaître. Suivant la direction de balayage, ils sont caractérisés par des traînées horizontales, verticales ou diagonales. La solution pour atténuer ces artéfacts a été trouvée en exploitant les multiples vues de notre système.

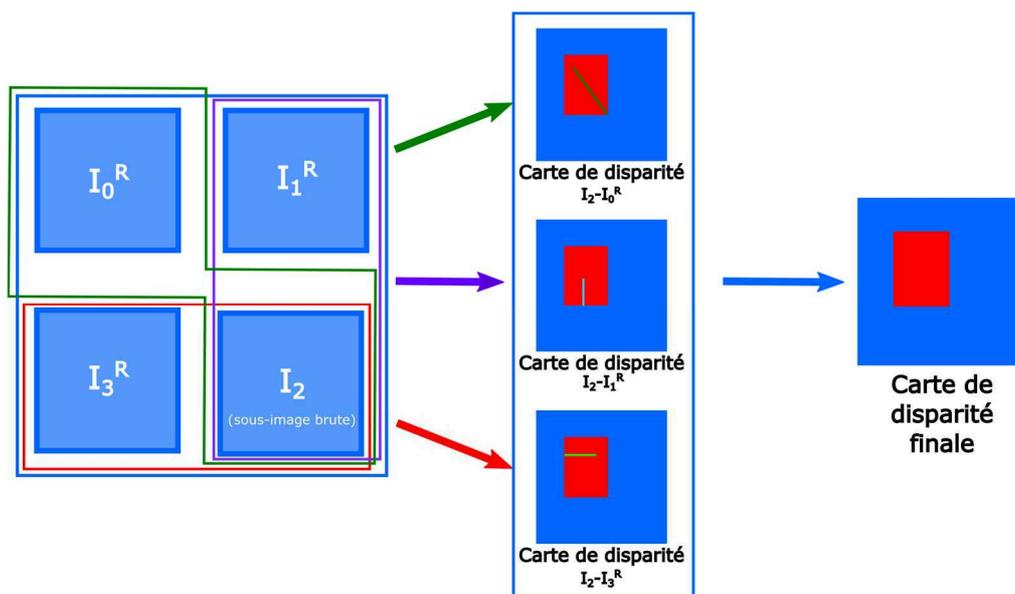


Figure 4.36 Combinaison des 3 cartes de disparité intermédiaire afin d'obtenir une carte de disparité finale sans artéfacts.

4.6.3. Résultats expérimentaux

Dans cette partie, nous allons présenter des résultats obtenus avec des images synthétiques ainsi que des résultats expérimentaux obtenus avec des images acquises avec notre prototype. Nous avons choisi d'effectuer deux types d'expérimentation. Les premiers résultats ayant permis de valider l'intérêt de notre approche et sa cohérence tandis que la seconde expérimentation permet de valider sa capacité de fonctionner pour des images obtenues avec un éclairage naturel et contenant des objets plats et uniformes.

4.6.3.1. Cas des images synthétiques

Au cours de cette première expérimentation, nous avons voulu vérifier que notre approche fonctionne correctement et fournit les résultats escomptés. Les images synthétiques utilisées sont composées des contours de différents objets présents dans la scène. La Figure 4.37 présente quelques-uns des résultats obtenus sur des images synthétiques montrant des objets localisés à différentes profondeurs.

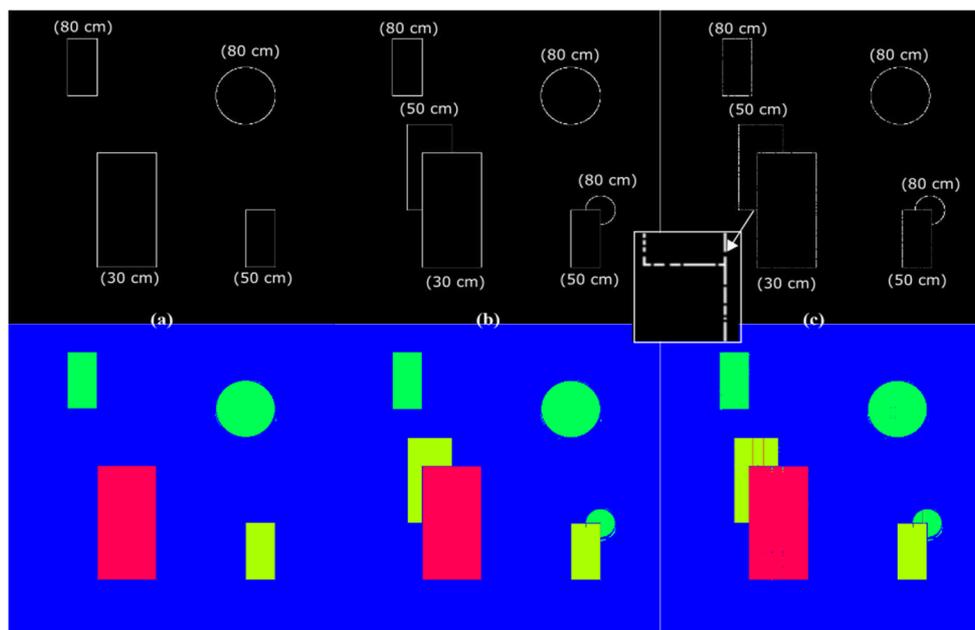


Figure 4.37 Images synthétiques (seule la sous-image I_2 est montrée) contenant des objets à différentes profondeurs et leur carte de profondeur correspondante. (a) 4 objets sans superposition sont placés dans la scène ; (b) des objets superposés sont placés ; (c) les contours dans les sous-images sont dégradés aléatoirement.

Le cas (a) montre une configuration simple, où les objets sont localisés éloignés les uns des autres. Le cas (b) montre des objets superposés. Pour ces deux cas, l'étape finale de fusion n'est pas nécessaire puisque chaque contour a un correspondant dans l'autre sous-image utilisée pour générer la carte de profondeur. Ainsi, seule une carte de disparité est suffisante pour générer la carte de profondeur finale. Pour le cas (c), nous avons choisi une situation dans laquelle tous les pixels ne disposent pas d'un correspondant dans les sous-images, comme c'est le cas dans les conditions expérimentales. L'image de profondeur est ensuite calculée grâce au processus de fusion utilisant ici les images de disparité (I_2 ,

I_3^R) et (I_2, I_1^R) comme source. Dans ce cas également, l'image finale de profondeur reste très proche du résultat attendu et convient pour la plupart des processus d'inspection 3D dans le domaine de la vision industrielle.

Ainsi, après avoir montré que notre approche est cohérente dans des conditions de synthèses, nous avons pu passer à la seconde expérience permettant de tester la méthode dans des conditions expérimentales.

4.6.3.2. Cas d'images acquises avec le prototype multi-vues

Au lieu de présenter les résultats sous la forme de carte de profondeur, ils sont présentés avec des images de hauteur, ce qui facilite la comparaison avec les objets de référence. Une image de hauteur I_H est définie comme suit :

$$I_H = Z_c - IP \quad (4.22)$$

Avec IP = Image de Profondeur.

Les différences correspondent juste à une translation des valeurs mesurées par la distance de calibrage Z_c . La Figure 4.38 montre des images de hauteur calculées à partir de scènes réelles. Nous explorons deux situations avec des objets ayant de grandes zones uniformes en un éclairage naturel, correspondant à des cas pour lesquels les méthodes standards de mise en correspondance ne fonctionnent pas quand aucune lumière structurée n'est utilisée. Pour cela, la distance de calibrage Z_c est adaptée à l'échelle de la scène comme expliqué dans le Chapitre 3, partie 3.4.6. Les hauteurs mesurées pour les quatre objets de référence sont ainsi bien retrouvées. Cependant, quelques artefacts sont encore visibles sur les cartes de hauteur. La principale caractéristique de la méthode proposée est d'être capable de calculer des cartes de disparité/hauteur sans mise en correspondance. Avec un seul capteur image qui acquiert simultanément quatre vues, les images capturées présentent la même luminosité et le même contraste. On peut donc espérer que la détection de contours fournit les mêmes contours pour les quatre vues. En pratique, à cause de l'échantillonnage du capteur image, ce n'est pas toujours le cas : quelques correspondances inexactes entre les contours détectés sont rencontrées.

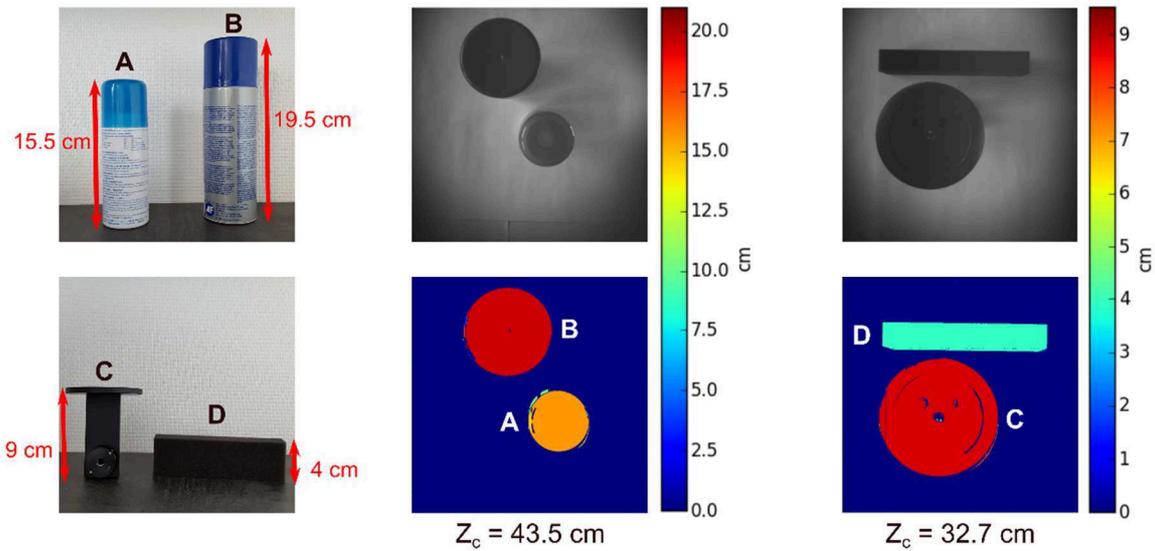


Figure 4.38 Résultats expérimentaux (seule la sous-image I_2 est montrée) avec quatre objets. Les hauteurs mesurées sont : A : 15,6 cm ; B ; 19,4 cm ; C : 8,7 cm ; D : 3,9 cm.

La Figure 4.39 présente un exemple décrivant cette situation quand la correspondance entre les contours détectés est inexacte, altérant la carte de hauteur/profondeur. Comme on peut le voir sur la partie (a) de la figure, il existe deux zones de contours, notées A et B, qui n'apparaissent que dans une sous-image, respectivement I_3^R et I_0^R . Ces deux zones correspondent aux deux zones considérées comme du fond sur la partie (b) de la Figure 4.39. D'autres artefacts peuvent être causés par la différence de point de vue, soit par occultation soit en cachant des contours entre les sous-images. Avec notre méthode, cette sorte d'artefacts est réduite significativement par le processus de fusion et ce n'est pas celui-ci le plus rencontré au cours de nos expériences.

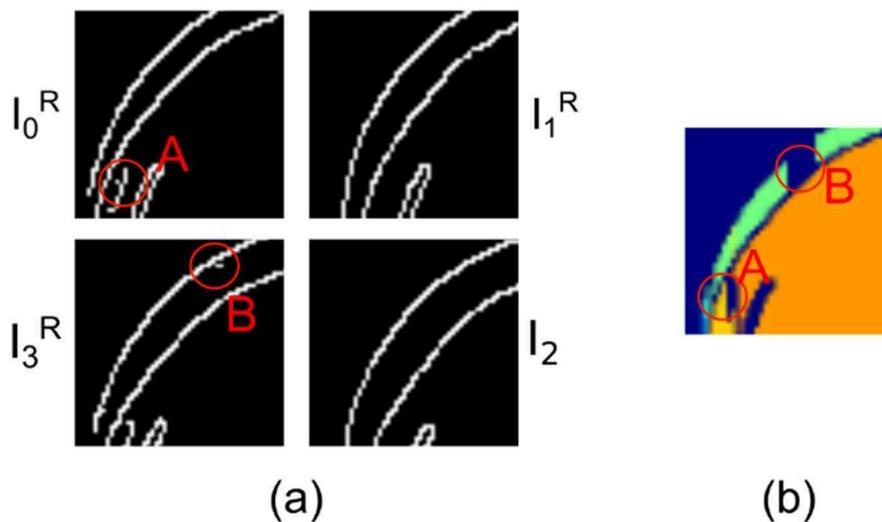


Figure 4.39 Zooms sur la détection de contours non similaires et les artefacts correspondants sur la carte de disparité. (a) quatre sous-images contours ; (b) zones d'artefacts où les contours ne sont pas identiques.

Cette approche pour le calcul de carte de hauteur/profondeur est validée par ces résultats synthétiques et expérimentaux. Ce travail fournit quelques nouvelles perspectives pour établir des cartes de

disparité sans procédure de mise en correspondance : principal intérêt de notre approche. Les artefacts rencontrés sur certains résultats devraient disparaître ou être atténués en améliorant et en adaptant la méthode de détection de contours au système multi-lentilles.

4.6.4. Conclusion

Cette seconde partie présente une nouvelle approche pour calculer des cartes de disparité, suivant la définition évoquée dans la section 4.6. Elle correspond à une mesure de décalage de contours en pixels, capable uniquement de mesurer des paliers. L'intérêt majeur de cette technique est de proposer un nouveau cadre ne nécessitant pas de procédure de mise en correspondance, qui fonctionne avec les zones uniformes et peu texturées. Cette approche a été mise en œuvre pour mesurer la hauteur d'objets placés devant une caméra pour des contrôles industriels par vision. En utilisant les contours, cette méthode est efficace sur des objets plats et uniformes en lumière naturelle. Notre solution peut être considérée comme une méthode d'interpolation de l'information de disparité ou de profondeur dans les zones uniformes, et fournissant un rendu réaliste pour des objets simples. La comparaison des résultats ainsi obtenus avec d'autres méthodes de calcul de carte de disparité n'est pas nécessaire puisque les résultats associés à ces dernières seront meilleurs. Cependant, comme nous avons travaillé avec un capteur bas coût, nous avons montré qu'une méthode sans mise en correspondance peut fournir des informations suffisantes pour générer des cartes de profondeur.

4.7. Conclusion

Ce chapitre a décrit l'application dans des domaines variés de notre système multi-vues, tels que le contrôle par vision. Notre dispositif multi-caméras peut fonctionner pour détecter des objets à une profondeur supérieure à deux mètres et présente un intérêt pour le domaine automobile. Les principales contributions consistent en une approche originale pour calculer des cartes de disparité ou des images de profondeur, et aussi pour le calcul d'images refocalisées calibrées en profondeur. Ces dernières présentent des intérêts dans les domaines de la vision industrielle et de l'automobile. Pour ce domaine, les expérimentations sont menées avec le système multi-caméras.

Comme la photométrie de la scène est conservée, contrairement aux mesures effectuées en illumination active, la plupart des contrôles standards, comme par exemple la lecture de code barre, la reconnaissance de forme, peut être fait conjointement avec les mesures 3D. A partir de ce constat et de notre modélisation des caméras *ligh-field*, nous avons proposé une nouvelle méthode de calcul d'images refocalisées calibrées en profondeur. Afin de déterminer quelle image est la plus nette parmi la pile d'images, une comparaison de plusieurs méthodes de détection de flou notamment dans le domaine du *shape-from-focus* a été effectuée. Elle a permis de sélectionner un critère pour caractériser le degré de flou dans nos images. De plus, une étude sur l'incertitude de mesures a été menée.

Différents intérêts de ces images sont proposés comme la sélection d'un plan de travail permettant à la fois de connaître la hauteur d'un objet et d'effectuer des contrôles conventionnels en vision industrielle comme la lecture de caractères ou de code-barres. De plus, le concept de mesure 3D agile a été développé notamment pour la mesure de hauteur et de diamètre d'objets. Pour cela, trois approches ont été envisagées : une première basée sur l'utilisation du critère de Tenengrad sélectionné, une seconde ajoutant une étape préliminaire de *matching* avant le calcul du degré du flou et une troisième consistant à une mise en correspondance à l'aide d'un logiciel de vision industrielle. Enfin, une méthode pour classifier les obstacles dans une scène automobile en différentes zones de profondeur a été mise en œuvre. Les bases d'une mesure 3D basée sur la refocalisation *light-field* ont ainsi été posées.

La deuxième partie de ce chapitre présente une nouvelle approche pour calculer des cartes de disparité, suivant la définition rappelée précédemment. L'intérêt majeur de cette technique est de ne pas nécessiter de procédure de mise en correspondance, et de fonctionner avec les zones uniformes et peu texturées. Nous avons développé notre approche mesurant la hauteur d'objets placés devant une caméra pour des contrôles industriels par vision. Basée sur l'extraction des contours, cette méthode est efficace sur des objets plats et uniformes en lumière naturelle. Notre solution peut être considérée comme une méthode d'interpolation de l'information de disparité ou de profondeur dans les zones uniformes, et fournissant un rendu réaliste pour des objets simples.

Ce travail valide les intérêts de nos approches. Quelques perspectives intéressantes sont envisagées. Notre méthode de calcul de cartes de disparité peut être étendue à des scènes naturelles plus complexes. De plus, l'amélioration de la détection de contours afin de réduire les artéfacts constitue une autre perspective. Enfin, une troisième perspective concernant les images refocalisées est à envisager. Il s'agirait d'adapter cette méthode aux caméras plénoptiques et aux dispositifs multi-vues plénoptiques, notamment en prenant en compte les différences de modélisation qu'il existe entre ces dispositifs et notre dispositif multi-vues. De plus, d'autres domaines d'applications de ces images pourraient être envisagés comme l'odométrie visuelle. Enfin, à partir de ces images, il pourrait être intéressant de générer des images multi-focalisées pour lesquelles il serait possible de choisir les distances auxquelles l'image est nette tandis que les autres plans de l'image apparaîtront flous. Ces images pourraient trouver un intérêt dans le domaine de l'art, notamment dans le cinéma.

Conclusions et perspectives

Dans cette thèse, les architectures et les apports de systèmes de vision *light-field* pour la vision par ordinateur ont été présentés. Plus précisément, les principaux dispositifs étudiés sont les caméras multi-vues, multi-vues plénoptique et les systèmes multi-caméras. Dans le chapitre 2, la conception optique et l'étude des systèmes optiques développés sont détaillés. Dans le chapitre 3, un calibrage simple à partir de notre modélisation est proposé. Dans le chapitre 4, les apports pour la vision industrielle de ces dispositifs sont présentés.

Conclusions

Le but de cette thèse est d'utiliser les caméras multi-vues et plénoptiques dans un autre domaine que la photographie. Pour répondre à cet objectif, plusieurs axes ont été étudiés comme expliqué ci-dessus.

Afin de comprendre la formation des images dans les systèmes plénoptiques, nos propres dispositifs ont été développés : un dispositif multi-caméras, un multi-vues et un prototype multi-vues plénoptique. Les deux premiers ont été mis en œuvre grâce à des équations découlant, soit de la triangulation, soit de notre modélisation de ces systèmes, présentée dans le chapitre 3. Dans les deux cas, leur caractérisation s'effectue au moyen de la mesure de la profondeur en fonction de la disparité. Le dispositif multi-vues plénoptique est composé d'un objectif principal, d'un capteur et de mini-lentilles, positionnées entre les deux autres éléments. Ce type de système étant plus complexe que les deux autres, un outil de simulation de tracé de rayons est développé pour prédire le comportement des rayons lumineux à travers les éléments optiques. De plus, en fonction de la géométrie du dispositif, des caractéristiques ont été définies : le taux de recouvrement, la parallaxe, le grandissement et l'ouverture numérique des mini-lentilles. Une comparaison entre les résultats théoriques et expérimentaux est effectuée. Elle a permis de vérifier l'exactitude des définitions.

Ensuite, dans le Chapitre 3, notre modélisation des caméras plénoptiques, à partir du principe de l'homographie variable, est expliquée. A partir de ce modèle, un calibrage simple et aisé a été proposé. Il nécessite uniquement deux acquisitions, chacune dans un plan différent mais dont l'écart en profondeur est connu. Puis, ce calibrage est perfectionné pour prendre en compte la distorsion. Pour ces deux méthodes de calibrage, une étude de la précision des mesures concernant la profondeur a été menée. De plus, un critère de calibrage est proposé afin de vérifier que ces données sont cohérentes. Cependant, il n'est ni utilisé pour prévenir les éventuelles erreurs de mesures ni pour affiner les résultats. Enfin, une comparaison entre nos méthodes et celle proposée par Zhang est menée. Les résultats associés nous permettent de conclure que notre méthode est plus simple à mettre en œuvre et qu'elle est plus précise pour le calcul des homographies entre les différentes vues.

Enfin, dans le Chapitre 4, nos dispositifs sont utilisés comme des caméras de profondeur. Deux principaux apports sont présentés : le calcul de carte de disparité et la refocalisation numérique calibrée en profondeur. Pour les cartes de profondeur, l'approche proposée repose sur la propagation de l'information de la profondeur/disparité entre les contours. Elle ne nécessite pas de mise en correspondance. Elle fonctionne pour des objets plats présentant de larges zones uniformes, observés en lumière naturelle. De plus, la mesure de la hauteur d'objets pour des contrôles par vision est possible en utilisant ce procédé. A partir de notre modélisation des caméras *light-field*, une nouvelle méthode de calcul d'images refocalisées numériquement et calibrées en profondeur est proposée. Un critère de mesure du flou, utilisé notamment dans le domaine du *shape-from-focus*, a été sélectionné pour faciliter la détermination de l'image la plus nette parmi la pile d'images. Comme les images sont acquises en lumière naturelle, la photométrie des images est conservée, permettant d'effectuer des contrôles conventionnels en vision par ordinateur. Ainsi, nous avons utilisé ces images refocalisées pour réaliser la plupart des contrôles standards comme la lecture de code-barres, la reconnaissance optique de caractères dans un plan de travail (concept établi à partir de ces images). De plus, plusieurs méthodes de mesure 3D agile ont été proposées. Chacune permet de réaliser des mesures 3D. Enfin, une dernière application dans le domaine de l'automobile permet de classifier les obstacles en différentes zones de profondeur.

Perspectives

A partir des différents sujets couverts dans cette thèse, des perspectives sont envisagées.

Plusieurs perspectives sont envisagées telles que l'étude de la faisabilité des systèmes multi-vues plénoptique à une échelle moins expérimentale et leur production à grande échelle. De plus, si ces deux points sont prometteurs, la propriété intellectuelle sera une conséquence logique à donner à ces travaux. Sachant que l'on obtient une matrice résultat, il est intéressant d'étudier la possibilité qu'il existe un modèle équivalent des matrices (A,B,C,D) pour les matrices résultats de dimension (3 x 3). Enfin, l'utilisation de ce modèle afin de fournir une configuration optique et mécanique d'un dispositif multi-vues plénoptique à partir d'un cahier des charges est la plus importante suite à donner à ces travaux.

Concernant le calibrage, il est à simplifier pour les systèmes multi-caméras. Le principe est applicable à ce type de dispositif, mais il est difficile à mettre en œuvre. Il faut donc proposer une méthode alternative facile à utiliser dans le domaine de l'automobile.

Concernant les apports des systèmes *light-field*, la détection des contours pour le calcul de carte de disparité est à améliorer. Ce calcul peut être étendu aux scènes naturelles plus complexes. Au sujet des images refocalisées calibrées en profondeur, la première perspective consiste à étudier de manière plus approfondie les limites de la précision des mesures. De plus, une étude de la mesure du flou, basée sur le critère de Tenengrad est à poursuivre, celle-ci a été commencée au cours d'un stage de

master. Dans ce cas, les observations sont effectuées avec différents motifs projetés sur l'objet étudié. Les résultats préliminaires montrent que le score mesuré est influencé par la forme du motif et par la taille de la fenêtre d'analyse. Les prochaines conclusions de ce travail pourraient conduire à optimiser une mesure de flou pour les systèmes *light-field*.

Enfin, le dispositif multi-vues est testé en conditions réelles dans une usine pour des applications de vision industrielle.

Production scientifique

- **Revue international**

- C. Riou et al., "Calibration and disparity maps for a depth camera based on a four-lens device," Journal of Electronic Imaging [doi:10.1117/1.JEI.24.6.061108].

- **Conférences internationales**

- C. Riou et al., "Interests of refocused images calibrated in depth with a multi-view camera for control by vision," conference QCAV 2017, Tokyo, Japon, 14-16 Mai 2017.
- C. Riou et al., "A four-lens based plenoptic camera for depth measurements," proceeding de la conference QCAV 2015 (3-5 juin 2015, Le Creusot, France)

- **Conférence nationale**

- C. Riou et al., "Une caméra de profondeur plénoptique basée sur 4 mini-objectifs," proceedings de la conf. ORASIS 2015 (Journées francophones des jeunes chercheurs en vision par ordinateur). Amiens 15-18 juin 2015

- **Autres**

- Depth plenoptic camera based on 4 mini-objectives à l'école d'été VISUM 2015 (Poster)
- GDR ISIS Journée Systèmes de Vision Grand Angle, Multi-Caméra et Plénoptique 18/11/2015 (Présentation orale)
- Calibrage et refocalisation à partir de systèmes multi-vues et multi-caméras pour les JEDS 2 Juin 2017 (Poster)
- GDR ISIS Journée Co-conception : capteurs hybrides et algorithmes pour des systèmes innovants 20/10/2017 (Présentation orale)

Bibliographie

- [00] *Extended Depth of Field - EPFL*. URL <http://bigwww.epfl.ch/demo/edf/#ref>. - abgerufen am 2017-10-17
- [16] *Google Camera*. URL <http://www.androidcentral.com/app/google-camera>. - abgerufen am 2017-02-07. — Android Central
- [ABVT13] AUBRETON, OLIVIER ; BAJARD, ALBAN ; VERNEY, BENJAMIN ; TRUCHETET, FREDERIC: Infrared system for 3D scanning of metallic surfaces. In: *Machine vision and applications* Bd. 24 (2013), Nr. 7, S. 1513–1524
- [AcGL01] ACCORSI, ROBERTO ; GASPARINI, FRANCESCA ; LANZA, RICHARD C.: Optimal coded aperture patterns for improved SNR in nuclear medicine imaging. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* Bd. 474 (2001), Nr. 3, S. 273–284
- [ADAD04] AGARWALA, ASEEM ; DONTCHEVA, MIRA ; AGRAWALA, MANEESH ; DRUCKER, STEVEN ; COLBURN, ALEX ; CURLESS, BRIAN ; SALESIN, DAVID ; COHEN, MICHAEL: Interactive digital photomontage. In: *ACM Transactions on Graphics (TOG)*. Bd. 23 : ACM, 2004, S. 294–302
- [AdBe91] ADELSON, EDWARD H. ; BERGEN, JAMES R.: The plenoptic function and the elements of early vision (1991)
- [AdWa92] ADELSON, EDWARD H. ; WANG, JOHN Y. A.: Single lens stereo with a plenoptic camera. In: *IEEE transactions on pattern analysis and machine intelligence* Bd. 14 (1992), Nr. 2, S. 99–106
- [BaDa06] BARRETO, JOAO P. ; DANIILIDIS, KOSTAS: Epipolar geometry of central projection systems using veronese maps. In: *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. Bd. 1 : IEEE, 2006, S. 1258–1265
- [Bail02] BAILEY, TIM: *Mobile robot localisation and mapping in extensive outdoor environments*, Citeseer, 2002
- [BASH15] BARRON, JONATHAN T. ; ADAMS, ANDREW ; SHIH, YICHANG ; HERNÁNDEZ, CARLOS: Fast bilateral-space stereo for synthetic defocus. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015*, S. 4466–4474
- [BBHL15] BUGARIN, FLORIAN ; BARTOLI, ADRIEN ; HENRION, DIDIER ; LASSERRE, JEAN-BERNARD ; ORTEU, JEAN-JOSÉ ; SENTENAC, THIERRY: Rank-constrained fundamental matrix estimation by polynomial global optimization versus the eight-point algorithm. In: *Journal of Mathematical Imaging and Vision* Bd. 53 (2015), Nr. 1, S. 42–60
- [BETV08] BAY, HERBERT ; ESS, ANDREAS ; TUYTELAARS, TINNE ; VAN GOOL, LUC: Speeded-up robust features (SURF). In: *Computer vision and image understanding* Bd. 110 (2008), Nr. 3, S. 346–359
- [BGYC13] BROXTON, MICHAEL ; GROSENICK, LOGAN ; YANG, SAMUEL ; COHEN, NOY ; ANDALMAN, AARON ; DEISSEROTH, KARL ; LEVOY, MARC: Wave optics theory and 3-D deconvolution for the light field microscope. In: *Optics express* Bd. 21 (2013), Nr. 21, S. 25418–25439
- [BiFa11] BISHOP, TOM E. ; FAVARO, PAOLO: Full-resolution depth map estimation from an aliased plenoptic light field. In: *Computer Vision–ACCV 2010* : Springer, 2011, S. 186–200

- [BoJK16] BOK, YUNSU ; JEON, HAE-GON ; KWEON, IN SO: Geometric calibration of micro-lens-based light field cameras using line features. In: *IEEE transactions on pattern analysis and machine intelligence* (2016)
- [Boot07] BOOTH, MARTIN J.: Adaptive optics in microscopy. In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* Bd. 365 (2007), Nr. 1861, S. 2829–2843
- [BSCB11] BERNET, SACHA ; STURM, PETER F. ; CUDEL, CHRISTOPHE ; BASSET, MICHEL ; ALPES, INRIA RHÔNE: Study on the interest of hybrid fundamental matrix for head mounted eye tracker modeling. In: *BMVC*, 2011, S. 1–10
- [BTAT10] BELDEN, JESSE ; TRUSCOTT, TADD T. ; AXIAK, MICHAEL C. ; TECHET, ALEXANDRA H.: Three-dimensional synthetic aperture particle image velocimetry. In: *Measurement Science and Technology* Bd. 21 (2010), Nr. 12, S. 125403
- [BuPr66] BUSBRIDGE, I. W. ; PREISENDORFER, R. W.: *Radiative transfer on discrete spaces* : JSTOR, 1966
- [CCPR09] CHIABRANDO, FILIBERTO ; CHIABRANDO, ROBERTO ; PIATTI, DARIO ; RINAUDO, FULVIO: Sensors for 3D imaging: Metric evaluation and calibration of a CCD/CMOS time-of-flight camera. In: *Sensors* Bd. 9 (2009), Nr. 12, S. 10080–10096
- [CDBV16] COSSU, KEVIN ; DRUART, GUILLAUME ; BONNEFOIS, AURELIE MONTMERLE ; VELLUET, MARIE-THERESE: Caméras plénoptiques pour l'imagerie tridimensionnelle. In: *Techniques de l'Ingenieur* (2016), Nr. R1393, S. 23–pages
- [CGSM04] CORNILLE, N. ; GARCIA, D. ; SUTTON, M. A. ; MCNEILL, S. ; ORTEU, J. J.: Calibrage d'imageurs avec prise en compte des distorsions. In: *Intrumentation, Mesure, Métrologie* Bd. 4 (2004), Nr. 3-4, S. 105–124
- [Cham05] CHAMBON, SYLVIE: *Mise en correspondance stéréoscopique d'images couleur en présence d'occultations*, Université toulouse III-Paul Sabatier, 2005
- [CLKT13] CHO, DONGHYEON ; LEE, MINHAENG ; KIM, SUNYEONG ; TAI, YU-WING: Modeling the calibration pipeline of the Lytro camera for high quality light-field image reconstruction. In: *Computer Vision (ICCV), 2013 IEEE International Conference on* : IEEE, 2013, S. 3280–3287
- [CILa04] CLARE, RICHARD M. ; LANE, RICHARD G.: Comparison of wavefront sensing with the Shack-Hartmann and pyramid sensors. In: *SPIE Astronomical Telescopes+ Instrumentation* : International Society for Optics and Photonics, 2004, S. 1211–1222
- [CLYB14] CHEN, CAN ; LIN, HAITING ; YU, ZHAN ; BING KANG, SING ; YU, JINGYI: Light field stereo matching using bilateral statistics of surface cameras. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, S. 1518–1525
- [CSDN87] CAROLI, E. ; STEPHEN, J. B. ; DI COCCO, G. ; NATALUCCI, L. ; SPIZZICHINO, A.: Coded aperture imaging in X-and gamma-ray astronomy. In: *Space Science Reviews* Bd. 45 (1987), Nr. 3-4, S. 349–403
- [CYAB14] COHEN, NOY ; YANG, SAMUEL ; ANDALMAN, AARON ; BROXTON, MICHAEL ; GROSENICK, LOGAN ; DEISSEROTH, KARL ; HOROWITZ, MARK ; LEVOY, MARC: Enhancing the performance of the light field microscope using wavefront coding. In: *Optics express* Bd. 22 (2014), Nr. 20, S. 24817–24839

- [DaBr03] DANSEREAU, DON ; BRUTON, LEN: A 4D frequency-planar IIR filter and its application to light field processing. In: *Circuits and Systems, 2003. ISCAS'03. Proceedings of the 2003 International Symposium on*. Bd. 4 : IEEE, 2003, S. IV–476
- [DaLD12] DAVIS, ABE ; LEVOY, MARC ; DURAND, FREDO: Unstructured light fields. In: *Computer Graphics Forum*. Bd. 31 : Wiley Online Library, 2012, S. 305–314
- [Dans14] DANSEREAU, DONALD G: *Plenoptic Signal Processing for Robust Vision in Field Robotics*, 2014
- [DaPW13] DANSEREAU, DONALD G. ; PIZARRO, OSCAR ; WILLIAMS, STEFAN B.: Decoding, calibration and rectification for lenselet-based plenoptic cameras. In: *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on* : IEEE, 2013, S. 1027–1034
- [DaPW15] DANSEREAU, DONALD G. ; PIZARRO, OSCAR ; WILLIAMS, STEFAN B.: Linear Volumetric Focus for Light Field Cameras. In: *ACM Trans. Graph.* Bd. 34 (2015), Nr. 2, S. 15:1–15:20
- [DMPW11] DANSEREAU, DONALD G. ; MAHON, IAN ; PIZARRO, OSCAR ; WILLIAMS, STEFAN B.: Plenoptic flow: Closed-form visual odometry for light field cameras. In: *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on* : IEEE, 2011, S. 4455–4462
- [Dubr09] DUBROFSKY, ELAN: *Homography estimation*, UNIVERSITY OF BRITISH COLUMBIA (Vancouver, 2009
- [EnLa93] ENS, JOHN ; LAWRENCE, PETER: An investigation of methods for determining depth from focus. In: *IEEE Transactions on pattern analysis and machine intelligence* Bd. 15 (1993), Nr. 2, S. 97–108
- [FaSo02] FAVARO, PAOLO ; SOATTO, STEFANO: Learning shape from defocus. In: *European Conference on Computer Vision* : Springer, 2002, S. 735–745
- [FiCS14] FISS, JULIET ; CURLESS, BRIAN ; SZELISKI, RICHARD: Refocusing plenoptic images using depth-adaptive splatting. In: *Computational Photography (ICCP), 2014 IEEE International Conference on* : IEEE, 2014, S. 1–9
- [Fore04] FOREST COLLADO, JOSEP: *New Methods for Triangulation-based Shape Acquisition using Laser Scanners*, Universitat de Girona, 2004
- [Garc01] GARCIA, DORIAN: *Mesure de formes et de champs de déplacements tridimensionnels par stéréo-corrélation d'images*, Institut National Polytechnique de Toulouse-INPT, 2001
- [GaVS12] GARBE, C. ; VOSS, BJÖRN ; STAPF, JULIAN: Plenoptic particle streak velocimetry (ppsv): 3d3c fluid flow measurement from light fields with a single plenoptic camera. In: *16th international symposium on applications of laser techniques to fluid mechanics, Instituto Superior Técnico, Lisbon* : Citeseer, 2012, S. 1–12
- [GeLG11] GEORGIEV, TODOR ; LUMSDAINE, ANDREW ; GOMA, SERGIO: Plenoptic principal planes. In: *Imaging Systems and Applications* : Optical Society of America, 2011, S. JTuD3
- [GeLu09] GEORGIEV, TODOR G. ; LUMSDAINE, ANDREW: Resolution in Plenoptic Cameras. In: *Frontiers in Optics 2009/Laser Science XXV/Fall 2009 OSA Optics & Photonics Technical Digest* : Optical Society of America, 2009, S. CTuB3

- [GeLu10] GEORGIEV, TODOR ; LUMSDAINE, ANDREW: Focused plenoptic camera and rendering. In: *Journal of Electronic Imaging* Bd. 19 (2010), Nr. 2, S. 021106–021106
- [GeLu12] GEORGIEV, TODOR ; LUMSDAINE, ANDREW: The multifocus plenoptic camera. In: *IS&T/SPIE Electronic Imaging* : International Society for Optics and Photonics, 2012, S. 829908–829908
- [Geng11] GENG, JASON: Structured-light 3D surface imaging: a tutorial. In: *Advances in Optics and Photonics* Bd. 3 (2011), Nr. 2, S. 128–160
- [GGSC96] GORTLER, STEVEN J. ; GRZESZCZUK, RADEK ; SZELISKI, RICHARD ; COHEN, MICHAEL F.: The lumigraph. In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* : ACM, 1996, S. 43–54
- [Hahn16] HAHNE, CHRISTOPHER: The standard plenoptic camera: applications of a geometrical light field model (2016)
- [HAHV14] HAHNE, CHRISTOPHER ; AGGOUN, AMAR ; HAXHA, SHYQYRI ; VELISAVLJEVIC, VLADAN ; FERNÁNDEZ, JUAN CARLOS JÁCOME: Light field geometry of a standard plenoptic camera. In: *Optics Express* Bd. 22 (2014), Nr. 22, S. 26659
- [HaNg06] HANRAHAN, PAT ; NG, REN: Digital correction of lens aberrations in light field photography. In: *International Optical Design Conference* : Optical Society of America, 2006, S. WB2
- [HAVF16] HAHNE, CHRISTOPHER ; AGGOUN, AMAR ; VELISAVLJEVIC, VLADAN ; FIEBIG, SUSANNE ; PESCH, MATTHIAS: Refocusing distance of a standard plenoptic camera. In: *Optics Express* Bd. 24 (2016), Nr. 19, S. 21521–21540
- [HaZi03] HARTLEY, RICHARD ; ZISSERMAN, ANDREW: *Multiple view geometry in computer vision* : Cambridge university press, 2003
- [HCBO15] HARVENT, J. ; COUDRIN, BENJAMIN ; BRÈTHES, L. ; ORTEU, J.-J. ; DEVY, MICHEL: Shape measurement using a new multi-step stereo-DIC algorithm that preserves sharp edges. In: *Experimental Mechanics* Bd. 55 (2015), Nr. 1, S. 167–176
- [HeRP13] HEBER, STEFAN ; RANFTL, RENE ; POCK, THOMAS: Variational Shape from Light Field. In: *Energy Minimization Methods in Computer Vision and Pattern Recognition* : Springer, 2013, S. 66–79
- [HFTG17] HALL, ELISE M. ; FAHRINGER, TIMOTHY W. ; THUROW, BRIAN S. ; GULDENBECHER, DANIEL R.: Volumetric calibration of a plenoptic camera. In: *55th AIAA Aerospace Sciences Meeting* : American Institute of Aeronautics and Astronautics, 2017
- [HRHL13] HEIDE, FELIX ; ROUF, MUSHFIQUR ; HULLIN, MATTHIAS B. ; LABITZKE, BJORN ; HEIDRICH, WOLFGANG ; KOLB, ANDREAS: High-quality computational imaging through simple lenses. In: *ACM Transactions on Graphics (TOG)* Bd. 32 (2013), Nr. 5, S. 149
- [HSHP15] HEINZE, CHRISTIAN ; SPYROPOULOS, STEFANO ; HUSSMANN, STEPHAN ; PERWASS, CHRISTIAN: Automated robust metric calibration of multi-focus plenoptic cameras. In: *2015 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2015, S. 2038–2043

- [HSKK96] HATA, SEIJI ; SAITOH, YOKO ; KUMAMURA, SYOJI ; KAIDA, KEN'ICHI: Shape extraction of transparent object using genetic algorithm. In: *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*. Bd. 4 : IEEE, 1996, S. 684–688
- [IRMS12] IHRKE, IVO ; RESHETOUSKI, ILYA ; MANAKOV, ALKHAZUR ; SEIDEL, HANS-PETER: Three-dimensional kaleidoscopic imaging. In: *Computational Optical Sensing and Imaging* : Optical Society of America, 2012, S. CTu4B–8
- [Ives03] IVES, FREDERIC E.: *Parallax stereogram and process of making same.* : Google Patents, 1903
- [Ives30] IVES, HERBERT E.: Parallax panoramagrams made with a large diameter lens. In: *Journal of the Optical Society of America* Bd. 20 (1930), Nr. 6, S. 332–340
- [JaBL12] JACOBS, DAVID E. ; BAEK, JONGMIN ; LEVOY, MARC: Focal stack compositing for depth of field control. In: *Stanford Computer Graphics Laboratory Technical Report* Bd. 1 (2012), Nr. 1, S. 2012
- [JHGP13] JOHANNSEN, OLE ; HEINZE, CHRISTIAN ; GOLDLUECKE, BASTIAN ; PERWALD, CHRISTIAN: On the calibration of focused plenoptic cameras. In: *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications* : Springer, 2013, S. 302–317
- [JoSG16] JOHANNSEN, OLE ; SULC, ANTONIN ; GOLDLUECKE, BASTIAN: Occlusion-Aware Depth Estimation Using Sparse Light Field Coding. In: *German Conference on Pattern Recognition* : Springer, 2016, S. 207–218
- [JPCP15] JEON, HAE-GON ; PARK, JAESIK ; CHOE, GYEONGMIN ; PARK, JINSUN ; BOK, YUNSU ; TAI, YU-WING ; SO KWEON, IN: Accurate depth map estimation from a lenslet light field camera. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015*, S. 1547–1555
- [KaLH08] KANG, YUN-SUK ; LEE, CHEON ; HO, YO-SUNG: An efficient rectification algorithm for multi-view images in parallel camera array. In: *3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video, 2008* : IEEE, 2008, S. 61–64
- [KFEW02] KOCH, REINHARD ; FRAHM, JAN-MICHAEL ; EVERS-SENNE, JAN-FRISO ; WOETZEL, JAN: Plenoptic modeling of 3d scenes with a sensoraugmented multi-camera rig. In: *Tyrrhenian International Workshop on Digital Communication (IWDC): proceedings, 2002*
- [KHPV99] KOCH, REINHARD ; HEIGL, BENNO ; POLLEFEYS, MARC ; VAN GOOL, LUC ; NIEMANN, HEINRICH: A geometric approach to light field calibration. In: *International Conference on Computer Analysis of Images and Patterns* : Springer, 1999, S. 596–603
- [KiOK14] KIM, MIN-JUNG ; OH, TAE-HYUN ; KWEON, IN So: Cost-aware depth map estimation for Lytro camera. In: *Image Processing (ICIP), 2014 IEEE International Conference on* : IEEE, 2014, S. 36–40
- [KoZa01] KOLMOGOROV, VLADIMIR ; ZABIH, RAMIN: Computing visual correspondence with occlusions using graph cuts. In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. Bd. 2 : IEEE, 2001, S. 508–515
- [KZPS13] KIM, CHANGIL ; ZIMMER, HENNING ; PRITCH, YAEL ; SORKINE-HORNUNG, ALEXANDER ; GROSS, MARKUS H.: Scene reconstruction from high spatio-angular resolution light fields. In: *ACM Trans. Graph.* Bd. 32 (2013), Nr. 4, S. 73–1

- [Lam15] LAM, EDMUND Y.: Computational photography with plenoptic camera and light field capture: tutorial. In: *Journal of the Optical Society of America A* Bd. 32 (2015), Nr. 11, S. 2021
- [LeHa96] LEVOY, MARC ; HANRAHAN, PAT: Light field rendering. In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* : ACM, 1996, S. 31–42
- [LeZM09] LEVOY, MARC ; ZHANG, ZHENGYUN ; MCDOWALL, IAN: Recording and controlling the 4D light field in a microscope using microlens arrays. In: *Journal of microscopy* Bd. 235 (2009), Nr. 2, S. 144–162
- [LFDF07] LEVIN, ANAT ; FERGUS, ROB ; DURAND, FRÉDO ; FREEMAN, WILLIAM T.: Image and depth from a conventional camera with a coded aperture. In: *ACM transactions on graphics (TOG)* Bd. 26 (2007), Nr. 3, S. 70
- [Lipp08] LIPPMANN, GABRIEL: Epreuves reversibles. photographies integrals. In: *Comptes-Rendus Academie des Sciences* Bd. 146 (1908), S. 446–451
- [LiZL15] LIU, PENG ; ZHANG, RUMIN ; LIU, DIJUN: Digital depth of field control in photography based on focused plenoptic camera. In: *2015 International Conference on Intelligent Systems Research and Mechatronics Engineering* : Atlantis Press, 2015
- [LLWL08] LIANG, CHIA-KAI ; LIN, TAI-HSU ; WONG, BING-YI ; LIU, CHI ; CHEN, HOMER H.: Programmable aperture photography: multiplexed light field acquisition. In: *ACM Transactions on Graphics (TOG)*. Bd. 27 : ACM, 2008, S. 55
- [LNAF06] LEVOY, MARC ; NG, REN ; ADAMS, ANDREW ; FOOTER, MATTHEW ; HOROWITZ, MARK: Light field microscopy. In: *ACM Transactions on Graphics (TOG)* Bd. 25 (2006), Nr. 3, S. 924–934
- [Lowe99] LOWE, DAVID G.: Object recognition from local scale-invariant features. In: *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. Bd. 2 : IEEE, 1999, S. 1150–1157
- [LuGe09] LUMSDAINE, ANDREW ; GEORGIEV, TODOR: The focused plenoptic camera. In: *Computational Photography (ICCP), 2009 IEEE International Conference on* : IEEE, 2009, S. 1–8
- [LuMF13] LU, CHIEN-HUNG ; MUENZEL, STEFAN ; FLEISCHER, JASON: High-resolution light-field microscopy. In: *Computational Optical Sensing and Imaging* : Optical Society of America, 2013, S. CTh3B–2
- [LyFT12] LYNCH, KYLE ; FAHRINGER, TIM ; THUROW, BRIAN: Three-dimensional particle image velocimetry using a plenoptic camera. In: *50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition, 2012*, S. 1056
- [MiRI17] MIGNARD-DEBISE, LOIS ; RESTREPO, JOHN ; IHRKE, IVO: A Unifying First-Order Model for Light-Field Cameras: the Equivalent Camera Array. In: *IEEE Transactions on Computational Imaging* Bd. PP (2017), Nr. 99, S. 1–1
- [MNST13] MAENO, KAZUKI ; NAGAHARA, HAJIME ; SHIMADA, ATSUSHI ; TANIGUCHI, RIN-ICHIRO: Light field distortion feature for transparent object recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013*, S. 2786–2793
- [MoKu05] MORRIS, N. J. W. ; KUTULAKOS, K. N.: Dynamic refraction stereo. In: *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*. Bd. 2 : IEEE, 2005, S. 1573–1580

- [MSCS14] MREJEN, SARAH ; SATO, TAKU ; CURCIO, CHRISTINE A. ; SPAIDE, RICHARD F.: Assessing the cone photoreceptor mosaic in eyes with pseudodrusen and soft drusen in vivo using adaptive optics imaging. In: *Ophthalmology* Bd. 121 (2014), Nr. 2, S. 545–551
- [Naya89] NAYAR, SHREE K.: *Shape from Focus* (1989)
- [Ng06] NG, REN: *Digital light field photography*, stanford university, 2006
- [NLBD05] NG, REN ; LEVOY, MARC ; BRÉDIF, MATHIEU ; DUVAL, GENE ; HOROWITZ, MARK ; HANRAHAN, PAT: Light field photography with a hand-held plenoptic camera. In: *Computer Science Technical Report CSTR* Bd. 2 (2005), Nr. 11
- [PaTB14] PARK, JAE YOUNG ; TOŠIĆ, IVANA ; BERKNER, KATHRIN: Plenoptic system identification using random calibration patterns. In: *Image Processing (ICIP), 2014 IEEE International Conference on* : IEEE, 2014, S. 51–55
- [Pent87] PENTLAND, ALEX PAUL: A new sense for depth of field. In: *IEEE transactions on pattern analysis and machine intelligence* (1987), Nr. 4, S. 523–531
- [PePG13] PERTUZ, SAID ; PUIG, DOMENEC ; GARCIA, MIGUEL ANGEL: Analysis of focus measure operators for shape-from-focus. In: *Pattern Recognition* Bd. 46 (2013), Nr. 5, S. 1415–1432
- [Pére84] PEREZ, JOSE-PHILIPPE: Optique. Optique geometrique matricielle et ondulatoire. In: *Paris: Masson, 1984* (1984)
- [Perr08] PERROLLAZ, MATHIAS: *Détection d'obstacles multi-capteurs supervisée par stéréovision*, Université Pierre et Marie Curie-Paris VI, 2008
- [PeST00] PEDERSINI, FEDERICO ; SARTI, AUGUSTO ; TUBARO, STEFANO: Automatic monitoring and 3D reconstruction applied to cultural heritage. In: *Journal of Cultural Heritage* Bd. 1 (2000), Nr. 3, S. 301–313
- [PeWi12] PERWASS, CHRISTIAN ; WIETZKE, LENNART: Single lens 3D-camera with extended depth-of-field. In: *IS&T/SPIE Electronic Imaging* : International Society for Optics and Photonics, 2012, S. 829108–829108
- [PPRM12] PÉREZ, FERNANDO ; PÉREZ, ALEJANDRO ; RODRIGUEZ, MANUEL ; MAGDALENO, EDUARDO: Fourier slice super-resolution in plenoptic cameras. In: *Computational Photography (ICCP), 2012 IEEE International Conference on* : IEEE, 2012, S. 1–11
- [PVVV04] POLLEFEYS, MARC ; VAN GOOL, LUC ; VERGAUWEN, MAARTEN ; VERBIEST, FRANK ; CORNELIS, KURT ; TOPS, JAN ; KOCH, REINHARD: Visual modeling with a hand-held camera. In: *International Journal of Computer Vision* Bd. 59 (2004), Nr. 3, S. 207–232
- [RCNF08] RODRÍGUEZ-RAMOS, J. M. ; CASTELLÁ, B. FEMENÍA ; NAVA, F. PÉREZ ; FUMERO, S.: Wavefront and distance measurement using the CAFADIS camera. In: *SPIE Astronomical Telescopes+ Instrumentation* : International Society for Optics and Photonics, 2008, S. 70155Q–70155Q
- [Rebe17] REBERT, MARTIN: Comparaison de décompositions de la matrice homographique et essentielle pour l'estimation du mouvement de caméra. In: . Colleville-sur-Mer, France, 2017

- [ReBR13] REDDY, DIKPAL ; BAI, JIAMIN ; RAMAMOORTHY, RAVI: External mask based depth and light field camera. In: *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on* : IEEE, 2013, S. 37–44
- [RiSá10] RICOLFE-VIALA, CARLOS ; SÁNCHEZ-SALMERÓN, ANTONIO-JOSÉ: Robust metric calibration of non-linear camera lens distortion. In: *Pattern Recognition* Bd. 43 (2010), Nr. 4, S. 1688–1699
- [SaAB02] SALVI, JOAQUIM ; ARMANGUÉ, XAVIER ; BATLLE, JOAN: A comparative review of camera calibrating methods with accuracy evaluation. In: *Pattern recognition* Bd. 35 (2002), Nr. 7, S. 1617–1635
- [SaPB04] SALVI, JOAQUIM ; PAGES, JORDI ; BATLLE, JOAN: Pattern codification strategies in structured light systems. In: *Pattern recognition* Bd. 37 (2004), Nr. 4, S. 827–849
- [SaTD09] SANSONI, GIOVANNA ; TREBESCHI, MARCO ; DOCCHIO, FRANCO: State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation. In: *Sensors* Bd. 9 (2009), Nr. 1, S. 568–601
- [SCDS06] SEITZ, STEVEN M. ; CURLESS, BRIAN ; DIEBEL, JAMES ; SCHARSTEIN, DANIEL ; SZELISKI, RICHARD: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: *Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on*. Bd. 1 : IEEE, 2006, S. 519–528
- [ScSz02] SCHARSTEIN, DANIEL ; SZELISKI, RICHARD: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In: *International journal of computer vision* Bd. 47 (2002), Nr. 1-3, S. 7–42
- [Sett01] SETTLES, GS: *Schlieren and Shadowgraph Techniques* (2001)
- [Shan16] SHAN, XU: *4D light field reconstruction and scene depth estimation from plenoptic camera raw images*, 2016
- [Sola07] SOLA, JOAN: *Towards visual localization, mapping and moving objects tracking by a mobile robot: a geometric and probabilistic approach*, Institut National Polytechnique de Toulouse-INPT, 2007
- [SSDS14] SABATER, NEUS ; SEIFI, MOZHDEH ; DRAZIC, VALTER ; SANDRI, GUSTAVO ; PÉREZ, PATRICK: Accurate Disparity Estimation for Plenoptic Images. In: *Computer Vision-ECCV 2014 Workshops* : Springer, 2014, S. 548–560
- [Subb88] SUBBARAO, MURALI: Parallel Depth Recovery by Changing Camera Parameters. In: *ICCV*, 1988, S. 149–155
- [SuHS15] SUWAJANAKORN, SUPASORN ; HERNANDEZ, CARLOS ; SEITZ, STEVEN M.: Depth from focus with your mobile phone. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, S. 3497–3506
- [Surr02] SURREL, YVES: Images optiques; mesures 2D et 3D. In: *cours du Conservatoire National des Arts et Métiers,(2002-2003)* (2002)
- [SvMP05] SVOBODA, TOMÁŠ ; MARTINEC, DANIEL ; PAJDLA, TOMÁŠ: A convenient multicamera self-calibration for virtual environments. In: *PRESENCE: teleoperators and virtual environments* Bd. 14 (2005), Nr. 4, S. 407–422

- [Szel10] SZELISKI, RICHARD: *Computer vision: algorithms and applications* : Springer Science & Business Media, 2010
- [Szel11] SZELISKI, RICHARD: *Computer Vision*, 2011. — DOI: 10.1007/978-1-84882-935-0
- [TCLI11] TROUVE, PAULINE ; CHAMPAGNAT, FREDERIC ; LE BESNERAIS, GUY ; IDIER, JEROME: Single image local blur identification. In: *Image Processing (ICIP), 2011 18th IEEE International Conference on* : IEEE, 2011, S. 613–616
- [THMR13] TAO, MICHAEL W. ; HADAP, SUNIL ; MALIK, JITENDRA ; RAMAMOORTHY, RAVI: Depth from Combining Defocus and Correspondence Using Light-Field Cameras. In: *Computer Vision (ICCV), 2013 IEEE International Conference on* : IEEE, 2013, S. 673–680
- [ThTF14] THOMASON, CHELSEA M. ; THUROW, BRIAN S. ; FAHRINGER, TIMOTHY W.: Calibration of a micro-lens array for a plenoptic camera. In: *52nd Aerospace Sciences Meeting*, 2014, S. 0396
- [TKYM01] TANIDA, JUN ; KUMAGAI, TOMOYA ; YAMADA, KENJI ; MIYATAKE, SHIGEHIRO ; ISHIDA, KOUICHI ; MORIMOTO, TAKASHI ; KONDOU, NORIYUKI ; MIYAZAKI, DAISUKE ; U. A.: Thin observation module by bound optics (TOMBO): concept and experimental verification. In: *Applied optics* Bd. 40 (2001), Nr. 11, S. 1806–1813
- [TLGS05] TARINI, MARCO ; LENSCH, HENDRIK PA ; GOESELE, MICHAEL ; SEIDEL, HANS-PETER: 3D acquisition of mirroring objects using striped patterns. In: *Graphical Models* Bd. 67 (2005), Nr. 4, S. 233–259
- [ToBe14] TOSIC, IVANA ; BERKNER, KATHRIN: Light field scale-depth space transform for dense depth estimation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, S. 435–442
- [TsCC07] TSENG, YU-CHENG ; CHANG, NELSON ; CHANG, TIAN-SHEUAN: Low memory cost block-based belief propagation for stereo correspondence. In: *Multimedia and Expo, 2007 IEEE International Conference on* : IEEE, 2007, S. 1415–1418
- [Tyso10] TYSON, ROBERT: *Principles of Adaptive Optics* : CRC Press, 2010
- [UeTo03] UESHIBA, TOSHIO ; TOMITA, FUMIAKI: Plane-based calibration algorithm for multi-camera systems via factorization of homography matrices. In: *null* : IEEE, 2003, S. 966
- [USKK13] UENO, R. ; SUZUKI, K. ; KOBAYASHI, M. ; KWON, H. ; HONDA, H. ; FUNAKI, H.: Compound-Eye Camera Module as Small as 8.5 8.5 6.0 mm for 26 k-Resolution Depth Map and 2-Mpix 2D Imaging. In: *IEEE Photonics Journal* Bd. 5 (2013), Nr. 4, S. 6801212–6801212
- [Veks05] VEKSLER, OLGA: Stereo correspondence by dynamic programming on a tree. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Bd. 2 : IEEE, 2005, S. 384–390
- [VGTA07] VAISH, V. ; GARG, G. ; TALVALA, E.-V. ; ANTUNEZ, E. ; WILBURN, B. ; HOROWITZ, M. ; LEVOY, M.: Synthetic Aperture Focusing Using Dense Camera Arrays. In: *3D Imaging for Safety and Security* : Springer, 2007, S. 159–172
- [VRAM07] VEERARAGHAVAN, ASHOK ; RASKAR, RAMESH ; AGRAWAL, AMIT ; MOHAN, ANKIT ; TUMBLIN, JACK: Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. In: *ACM Transactions on Graphics* Bd. 26 (2007), Nr. 3, S. 69

- [VWJL04] VAISH, VAIBHAV ; WILBURN, BENNETT ; JOSHI, NEEL ; LEVOY, MARC: Using plane+ parallax for calibrating dense camera arrays. In: *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. Bd. 1 : IEEE, 2004, S. 1–2
- [WaFJ11] WANNER, SVEN ; FEHR, JANIS ; JÄHNE, BERND: Generating EPI representations of 4D light fields with a single lens focused plenoptic camera. In: *Advances in Visual Computing* : Springer, 2011, S. 90–101
- [WaGo12] WANNER, SVEN ; GOLDLUECKE, BASTIAN: Globally consistent depth labeling of 4D light fields. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* : IEEE, 2012, S. 41–48
- [WaGo14] WANNER, SVEN ; GOLDLUECKE, BASTIAN: Variational light field analysis for disparity estimation and super-resolution. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* Bd. 36 (2014), Nr. 3, S. 606–619
- [WCHO92] WENG, JUYANG ; COHEN, PAUL ; HERNIOU, MARC ; OTHERS: Camera calibration with distortion models and accuracy evaluation. In: *IEEE Transactions on pattern analysis and machine intelligence* Bd. 14 (1992), Nr. 10, S. 965–980
- [WeRH11] WETZSTEIN, GORDON ; RASKAR, RAMESH ; HEIDRICH, WOLFGANG: Hand-held schlieren photography with light field probes. In: *Computational Photography (ICCP), 2011 IEEE International Conference on* : IEEE, 2011, S. 1–8
- [WJVT05] WILBURN, BENNETT ; JOSHI, NEEL ; VAISH, VAIBHAV ; TALVALA, EINO-VILLE ; ANTUNEZ, EMILIO ; BARTH, ADAM ; ADAMS, ANDREW ; HOROWITZ, MARK ; U. A.: High performance imaging using large camera arrays. In: *ACM Transactions on Graphics (TOG)* Bd. 24 (2005), Nr. 3, S. 765–776
- [WRHR11] WETZSTEIN, GORDON ; ROODNICK, DAVID ; HEIDRICH, WOLFGANG ; RASKAR, RAMESH: Refractive shape from light field distortion. In: *Computer Vision (ICCV), 2011 IEEE International Conference on* : IEEE, 2011, S. 1180–1186
- [XCKF11] XU, JUN ; CUDEL, CHRISTOPHE ; KOHLER, SOPHIE ; FONTAINE, STÉPHANE ; HAEBERLÉ, OLIVIER ; KLOTZ, MARIE-LOUISE: Using variable homography to measure emergent fibers on textile fabrics. In: *10th International Conference on Quality Control by Artificial Vision* : International Society for Optics and Photonics, 2011, S. 800003–800003
- [XCKF12] XU, JUN ; CUDEL, CHRISTOPHE ; KOHLER, SOPHIE ; FONTAINE, STÉPHANE ; HAEBERLÉ, OLIVIER ; KLOTZ, MARIE-LOUISE: Original method to compute epipoles using variable homography: application to measure emergent fibers on textile fabrics. In: *Journal of Electronic Imaging* Bd. 21 (2012), Nr. 2, S. 021103–1
- [XCKF13] XU, JUN ; CUDEL, CHRISTOPHE ; KOHLER, SOPHIE ; FONTAINE, STÉPHANE ; HAEBERLÉ, OLIVIER ; KLOTZ, MARIE LOUISE: A New Method to Measure 3D Textile Defects by Using Dual-lens Camera. In: *Advanced Materials Research*. Bd. 821 : Trans Tech Publ, 2013, S. 237–242
- [XMNT15] XU, YICHAO ; MAENO, KAZUKI ; NAGAHARA, HAJIME ; TANIGUCHI, RIN-ICHIRO: Camera array calibration for light field acquisition. In: *Frontiers of Computer Science* Bd. 9 (2015), Nr. 5, S. 691–702
- [Xu12] XU, JUN: *Using Variable Homography to measurement Protruding Fibers on Textile Fabrics*, Université de Haute-Alsace, 2012

- [YGLL13] YU, ZHAN ; GUO, XINQING ; LIN, HAIBING ; LUMSDAINE, ANDREW ; YU, JINGYI: Line assisted light field triangulation and stereo matching. In: *Proceedings of the IEEE International Conference on Computer Vision*, 2013, S. 2792–2799
- [YYLG13] YU, ZHAN ; YU, JINGYI ; LUMSDAINE, ANDREW ; GEORGIEV, TODOR: Plenoptic depth map in the case of occlusions. In: *IS&T/SPIE Electronic Imaging : International Society for Optics and Photonics*, 2013, S. 86671S–86671S
- [ZeOt96] ZELLER, CYRIL ; OTHERS: *Calibration projective, affine et Euclidienne en vision par ordinateur et application a la perception tridimensionnelle*, 1996
- [ZeQS14] ZELLER, NICLAS ; QUINT, FRANZ ; STILLA, UWE: Applying a Traditional Calibration Method to a Focused Plenoptic Camera. In: *BW-CAR/ SINCOM (2014)*, S. 19
- [ZeQS15] ZELLER, N. ; QUINT, F. ; STILLA, U.: Narrow field-of-view visual odometry based on a focused plenoptic camera. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* Bd. 2 (2015), Nr. 3, S. 285
- [ZeQS16] ZELLER, NICLAS ; QUINT, FRANZ ; STILLA, UWE: Depth estimation and camera calibration of a focused plenoptic camera for visual odometry. In: *ISPRS Journal of Photogrammetry and Remote Sensing* Bd. 118 (2016), S. 83–100
- [Zhan00] ZHANG, ZHENGYOU: A flexible new technique for camera calibration. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* Bd. 22 (2000), Nr. 11, S. 1330–1334
- [ZhGr07] ZHANG, SHAWN ; GREENSPAN, MICHAEL: *Variable homography compensation of parallax along mosaic seams* : Springer, 2007
- [ZhHu04] ZHANG, SONG ; HUANG, PEISEN: High-resolution, real-time 3D shape acquisition. In: *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on* : IEEE, 2004, S. 28–28
- [ZhJW16] ZHANG, CHUNPING ; JI, ZHE ; WANG, QING: Decoding and calibration method on focused plenoptic camera. In: *Computational Visual Media* Bd. 2 (2016), Nr. 1, S. 57–69
- [ZhNa11] ZHOU, CHANGYIN ; NAYAR, SHREE K.: Computational cameras: convergence of optics and processing. In: *IEEE Transactions on Image Processing* Bd. 20 (2011), Nr. 12, S. 3322–3340
- [ZoNa06] ZOMET, ASSAF ; NAYAR, SHREE K.: Lensless imaging with a controllable aperture. In: *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. Bd. 1 : IEEE, 2006, S. 339–346