

N° d'ordre : D.U. xxxxx
EDSPIC : xxx

Université d'Auvergne
Faculté de Médecine
Clermont Ferrand, France
*École Doctorale des Sciences
Pour l'Ingénieur*

Université de Tunis El Manar
École Nationale d'Ingénieur de Tunis
Tunisie
*École Doctorale des Sciences
et Techniques de l'Ingénieur*

Thèse

en cotutelle pour l'obtention du grade de

Docteur

de

*l'Université d'Auvergne
en vision par ordinateur*

de

*l'École Nationale d'Ingénieurs de Tunis
en génie électrique*

Modélisation du bruit et étalonnage de la mesure de profondeur des caméras Temps-de-Vol

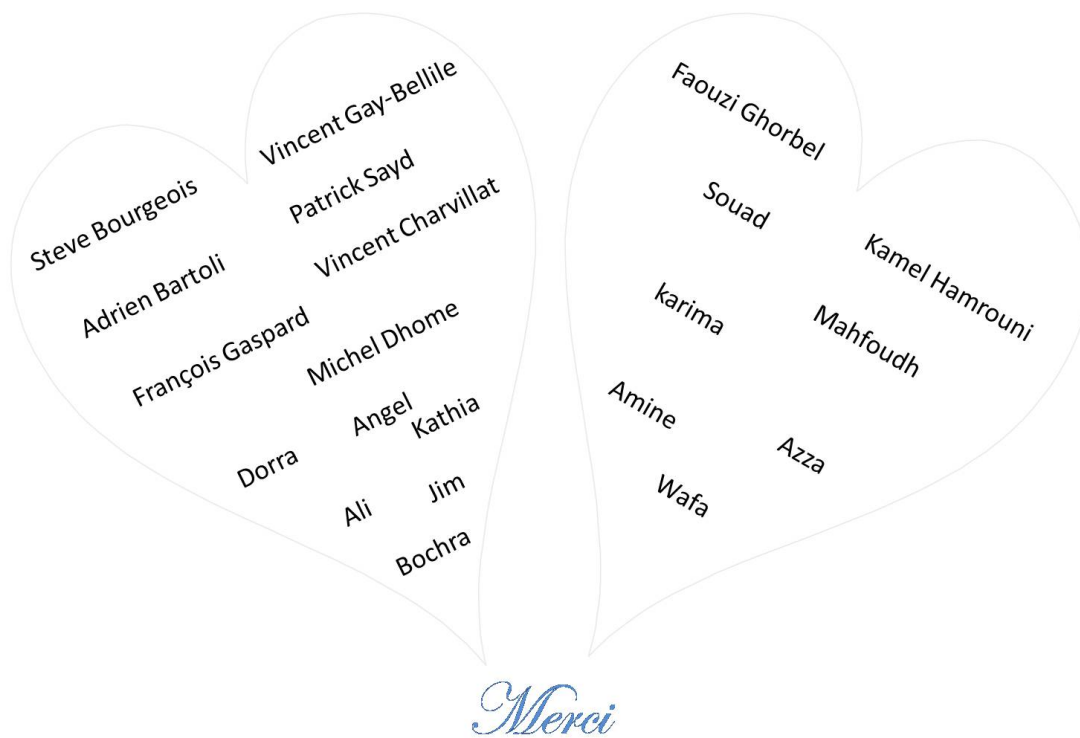
Présentée par

Amira BELHEDI

Soutenue publiquement le 4 juillet 2013 devant le jury :

| | | |
|------------------------|------------------------------------------|--------------------|
| Pr. Michel DHOME | Institut Pascal-Université Blaise Pascal | Président |
| Pr. Vincent CHARVILLAT | ENSEEIH-T-Université de Toulouse | Rapporteur |
| Pr. Faouzi GHORBEL | ENSI-Université La Manouba | Rapporteur |
| Pr. Adrien BARTOLI | ISIT-Université d'Auvergne | Directeur de thèse |
| Pr. Kamel HAMROUNI | ENIT-Université Tunis el Manar | Directeur de thèse |
| M. Patrick SAYD | LVIC-CEA,List | Co-encadrant |
| M. Steve BOURGEOIS | LVIC-CEA,List | Invité |
| M. Vincent GAY-BELLILE | LVIC-CEA,List | Invité |

Remerciements



Résumé

Avec l'apparition récente des caméras 3D, des perspectives nouvelles pour différentes applications de l'interprétation de scène se sont ouvertes. Cependant, ces caméras ont des limites qui affectent la précision de leurs mesures. En particulier pour les caméras Temps-de-Vol, deux types d'erreur peuvent être distingués : le bruit statistique de la caméra et la distorsion de la mesure de profondeur.

Dans les travaux de la littérature des caméras Temps-de-Vol, le bruit est peu étudié et les modèles de distorsion de la mesure de profondeur sont généralement difficiles à mettre en œuvre et ne garantissent pas la précision requise pour certaines applications. L'objectif de cette thèse est donc d'étudier, modéliser et proposer un étalonnage précis et facile à mettre en œuvre de ces deux types d'erreur des caméras Temps-de-Vol.

Pour la modélisation du bruit comme pour la distorsion de la mesure de profondeur, deux solutions sont proposées présentant chacune une solution à un problème différent. La première vise à fournir un modèle précis alors que le second favorise la simplicité de la mise en œuvre.

Ainsi, pour le bruit, alors que la majorité des modèles reposent uniquement sur l'information d'amplitude, nous proposons un premier modèle qui intègre aussi la position du pixel dans l'image. Pour encore une meilleure précision, nous proposons un modèle où l'amplitude est remplacée par la profondeur de l'objet et le temps d'intégration.

S'agissant de la distorsion de la mesure de profondeur, nous proposons une première solution basée sur un modèle non-paramétrique garantissant une meilleure précision. Ensuite, pour fournir une solution plus facile à mettre en œuvre que la précédente et que celles de l'état de l'art, nous nous basons sur la connaissance à priori de la géométrie planaire de la scène observée.

Mots clés : caméras Temps-de-Vol, modélisation du bruit, étalonnage de la mesure de profondeur, planaire, non-paramétrique.

Abstract

3D cameras open new possibilities in different fields such as 3D reconstruction, Augmented Reality and video-surveillance since they provide depth information at high frame-rates. However, they have limitations that affect the accuracy of their measures. In particular for TOF cameras, two types of error can be distinguished : the stochastic camera noise and the depth distortion.

In state of the art of TOF cameras, the noise is not well studied and the depth distortion models are difficult to use and don't guarantee the accuracy required for some applications. The objective of this thesis is to study, to model and to propose a calibration method of these two errors of TOF cameras which is accurate and easy to set up.

Both for the noise and for the depth distortion, two solutions are proposed. Each of them gives a solution for a different problem. The former aims to obtain an accurate model. The latter, promotes the simplicity of the set up.

Thereby, for the noise, while the majority of the proposed models are only based on the amplitude information, we propose a first model which integrate also the pixel position in the image. For a better accuracy, we propose a second model where we replace the amplitude by the depth and the integration time.

Regarding the depth distortion, we propose a first solution based on a non-parametric model which guarantee a better accuracy. Then, we use the prior knowledge of the planar geometry of the observed scene to provide a solution which is easier to use compared to the previous one and to those of the litterature.

Key words : TOF cameras, noise model, depth calibration, planarity, non-parametric.

Géométrie

| | |
|---------------------------------------------|------------------------------------------------|
| \sim | Egalité à un facteur non-nul près |
| $(u \ v)$ | Coordonnées pixéliques d'un point dans l'image |
| d | Profondeur d'un point |
| \mathbf{m} | Point 2D |
| $\mathbf{p} = (u \ v)^\top$ | Coordonnées pixéliques de \mathbf{m} |
| $\check{\mathbf{p}} = (u \ v \ 1)^\top$ | Coordonnées homogènes |
| $\mathbf{m} = (x \ y)^\top$ | Coordonnées millimétriques de \mathbf{m} |
| $\check{\mathbf{m}} = (x \ y \ 1)^\top$ | Coordonnées homogènes |
| $\mathbf{q} = (u \ v \ d)^\top$ | Point 2.5D |
| $\check{\mathbf{q}} = (u \ v \ d \ 1)^\top$ | Coordonnées homogènes |
| M | Point 3D |
| $\mathbf{Q} = (X, Y, Z)^\top$ | Coordonnées de M dans le repère caméra |
| $\check{\mathbf{Q}} = (X, Y, Z, 1)^\top$ | Coordonnées homogènes |
| $\mathbf{P} = (U, V, W)^\top$ | Coordonnées de M dans le repère monde |
| $\check{\mathbf{P}} = (U, V, W, 1)^\top$ | Coordonnées homogènes |
| \mathbf{A} | Matrice |
| \mathbf{A}^{-1} | Inverse de \mathbf{A} |
| \mathbf{A}^\dagger | Pseudo-inverse de \mathbf{A} |

| | |
|-------|-------------------------------------------------------------------------|
| A^T | Transposée de A |
| R | Matrice de rotation 3×3 , $R^T R = \mathbf{I}$, $\det(R) = 1$ |
| T | Vecteur de translation 3×1 |
| K | Matrice des paramètres intrinsèques 3×3 |
| P | Matrice de projection perspective 3×4 |

Statistique et analyse

| | |
|-----------------------|---------------------------------------|
| \mathbf{x} | Variable Aléatoire |
| $\bar{\mathbf{x}}$ | Moyenne de \mathbf{x} |
| $\Sigma^{\mathbf{x}}$ | Matrice de covariance de \mathbf{x} |
| \mathcal{N} | Loi normale |
| J_T | Jacobienne d'une fonction notée T |

Acronymes

| | |
|-----------|-------------------------------------------------------|
| TOF | 'Time Of Flight' |
| TI | Temps d'Intégration |
| f_{mod} | Fréquence de modulation |
| FPN | 'Fixed Pattern Noise' |
| DLT | 'Direct Linear Transform' |
| TPS | 'Thin-Plate Spline' |
| FCD | Fonction de Correction de la Distorsion |
| CNP | Correction de la Non-Planarité |
| CA | Correction Affine |
| EF | Estimation de la FCD notée F |
| EP | Estimation de l'ensemble des plans noté \mathcal{P} |

Table des matières

| | |
|------------------------------------------------------------|-----------|
| Introduction | 1 |
| 1 Principe des caméras Temps-de-Vol | 5 |
| 1.1 Introduction | 6 |
| 1.2 Fonctionnement | 6 |
| 1.3 Données de sortie | 9 |
| 1.4 Paramètres de capture | 10 |
| 1.5 Limites du capteur Temps-de-Vol | 14 |
| 1.5.1 Résolution | 14 |
| 1.5.2 Portée | 16 |
| 1.5.3 Imprécision des mesures de profondeur | 17 |
| 1.5.3.1 Incertitude | 18 |
| 1.5.3.2 Distorsion | 18 |
| 1.6 Conclusion | 22 |
| 2 Etat de l'art | 25 |
| 2.1 Modélisation du bruit des caméras Temps-de-Vol | 26 |
| 2.1.1 Méthodes de la littérature | 26 |
| 2.1.2 Discussion et positionnement | 28 |
| 2.2 Étalonnage de la distorsion de la mesure de profondeur | 29 |
| 2.2.1 Introduction | 30 |
| 2.2.2 Modélisation de la FCD | 31 |
| 2.2.2.1 Un seul modèle | 31 |
| 2.2.2.2 Deux modèles additifs | 32 |

| | | |
|----------|-----------------------------------------------------------------------|-----------|
| 2.2.2.3 | Extension des modèles | 33 |
| 2.2.3 | Estimation de la FCD | 34 |
| 2.2.4 | Acquisition d'une vérité terrain | 36 |
| 2.2.5 | Discussion et positionnement | 37 |
| 2.2.5.1 | Limites des méthodes de la littérature | 37 |
| 2.2.5.2 | Méthodes proposées | 38 |
| 2.3 | Conclusion | 39 |
| 3 | Modélisation du bruit | 41 |
| 3.1 | Introduction | 42 |
| 3.2 | Caractérisation du bruit | 44 |
| 3.2.1 | Caractérisation du bruit pour un TI fixé | 44 |
| 3.2.2 | Dépendance du bruit au TI | 48 |
| 3.3 | Modélisation continue du bruit | 51 |
| 3.3.1 | Modèle position-amplitude | 51 |
| 3.3.2 | Modèle position-profondeur- TI | 53 |
| 3.4 | Évaluation des deux modèles | 55 |
| 3.5 | Exemple d'application : filtrage des images de profondeur | 56 |
| 3.5.1 | Principe | 56 |
| 3.5.2 | Résultats | 57 |
| 3.6 | Conclusion | 59 |
| 4 | Étalonnage de la distorsion à partir de mesures 2.5D | 61 |
| 4.1 | Introduction | 62 |
| 4.2 | Présentation de la méthode | 63 |
| 4.2.1 | Modèle de la FCD | 63 |
| 4.2.2 | Formulation variationnelle | 63 |
| 4.2.3 | Discrétisation | 65 |
| 4.2.4 | Résolution | 67 |
| 4.3 | Résultats expérimentaux | 70 |
| 4.3.1 | Protocole expérimental | 70 |
| 4.3.1.1 | Calcul de la vérité terrain | 70 |
| 4.3.1.2 | Données utilisées | 73 |
| 4.3.2 | Détermination des valeurs optimales de λ | 74 |
| 4.3.3 | Résultats expérimentaux de notre étalonnage | 74 |
| 4.3.4 | Comparaison avec des méthodes de l'état de l'art | 75 |
| 4.4 | Discussion et conclusion | 78 |

| | | |
|----------|------------------------------------------------------------------------------|------------|
| 5 | Étalonnage de la distorsion à partir de structures planes | 81 |
| 5.1 | Motivation | 82 |
| 5.2 | Présentation de la méthode | 83 |
| 5.2.1 | Introduction | 83 |
| 5.2.2 | Modèle de la CNP | 85 |
| 5.2.3 | Formulation variationnelle de la CNP | 86 |
| 5.2.4 | Résolution itérative de la CNP | 88 |
| 5.3 | Modélisation et estimation de la FCD | 90 |
| 5.3.1 | Modélisation de la FCD avec une fonction TPS 3D | 91 |
| 5.3.2 | Estimation de la FCD | 92 |
| 5.4 | Résultats expérimentaux | 94 |
| 5.4.1 | Expériences sur des données simulées | 95 |
| 5.4.2 | Expériences sur des données réelles | 98 |
| 5.5 | Conclusion | 103 |
| 6 | Application : mesure 3D et incertitude associée | 105 |
| 6.1 | Principe | 106 |
| 6.1.1 | Incertitude du point 2.5D | 107 |
| 6.1.2 | Propagation de l'incertitude du 2.5D au 3D | 108 |
| 6.1.3 | Prise en compte de l'erreur d'étalonnage | 108 |
| 6.1.4 | Propagation de l'incertitude des points 3D aux mesures de distance | 110 |
| 6.2 | Résultats | 111 |
| 6.3 | Conclusion | 114 |
| | Conclusion | 117 |
| | Annexes | 119 |
| | A Analyse du bruit de la caméra Kinect | 121 |
| | B Notions de base | 125 |
| B.1 | Outils de vision par ordinateur | 126 |
| B.1.1 | Modélisation de la caméra | 126 |
| B.1.1.1 | Modèle sténopé | 126 |
| B.1.1.2 | Modèle de distorsion optique | 129 |
| B.1.2 | Résolution de systèmes linéaires | 130 |
| B.1.2.1 | Décomposition en valeurs singulières | 130 |
| B.1.2.2 | Pseudo-inverse | 131 |
| B.1.3 | Étalonnage et calcul de pose | 131 |

| | | |
|---------|-----------------------------------------------------------------------|------------|
| B.1.3.1 | Étalonnage standard d'une caméra | 131 |
| B.1.3.2 | Calcul de pose d'une caméra | 133 |
| B.1.3.3 | Étalonnage stéréo | 135 |
| B.2 | Outils statistiques et mathématiques | 136 |
| B.2.1 | Propagation d'incertitude | 136 |
| B.2.2 | Fonction de type <i>B-spline</i> | 138 |
| B.2.3 | Fonction spline de type plaque mince TPS 3D | 139 |
| B.3 | Transformations utilisées | 141 |
| B.3.1 | Transformation de l'espace 2.5D vers l'espace 3D | 142 |
| B.3.2 | Transformation entre les fonctions de correction f et F | 143 |
| | Bibliographie | 144 |

Table des figures

| | | |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | Principe du Temps-de-Vol. | 7 |
| 1.2 | Données fournies par une caméra Temps-de-Vol : (a) image de profondeur, (b) image d'amplitude, (c) image d'intensité et (d) nuage de points 3D. | 10 |
| 1.3 | Acquisitions dans différents contextes en faisant varier TI dans une scène statique (a) 0,14 ms, (b) 1,4 ms, (c) 14 ms et dans une scène dynamique (d) 14 ms (mouvement de la main). TI influence la qualité de l'image de profondeur et doit être ajusté à la nature de la scène (statique ou dynamique). | 11 |
| 1.4 | Acquisitions extérieures en faisant varier TI (a) 0,14 ms, (b) 1,4 ms et (c) 14 ms. Les images sont très bruitées. Ceci est dû principalement à l'interférence avec plusieurs autres sources de lumière. | 12 |
| 1.5 | Acquisitions en faisant varier f_{mod} (a) 18 MHz, (b) 19 MHz, (c) 20 MHz et (d) 21 MHz. f_{mod} n'a pas d'influence sur la qualité de l'image de profondeur. Par contre, elle change les valeurs des mesures de profondeur ainsi que la portée de la caméra. | 13 |
| 1.6 | Quelques exemples de caméras Temps-de-Vol basées sur la démodulation de type " <i>lock-in pixels</i> ". Leur résolution spatiale est faible : elle varie de 64×48 à 204×204 (actuellement la plus haute résolution disponible sur le marché). | 15 |

| | | |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.7 | Exemple d'image de profondeur (a) avec des objets situés à des distances supérieures à la distance ambiguë $\frac{c}{2f_{mod}}$. Le phénomène de dépliement de phase (" <i>unwrapping phase</i> ") est observé aux bords de l'image (les zones noires). La même image est présentée dans (b) après correction de dépliement de phase. | 17 |
| 1.8 | La profondeur du point de la surface du cube d_{TOF} issue du capteur Temps-de-Vol correspond à une mesure distordue de la profondeur réelle d_{GT} à laquelle s'additionne un bruit stochastique. | 19 |
| 1.9 | Distorsion de la mesure de profondeur calculée pour différents TI . Chaque courbe correspond à un TI donné (<i>extrait de [Kahlmann et al. 2006]</i>). | 20 |
| 1.10 | Carte de profondeur d'un damier de calibrage. La profondeur mesurée par la caméra varie selon l'indice de réflectivité de l'objet. Des objets (les carrés du damier) situés à la même distance de la caméra n'ont pas les mêmes mesures de profondeur. | 21 |
| 1.11 | Un exemple de graphe (a) représentant la distorsion de la mesure de profondeur d'une caméra Temps-de-Vol en fonction de la vérité terrain (en gris). La courbe qui approxime au mieux ces données oscille autour de zéro (fonction <i>B-spline</i> tracée en noir) (<i>extrait de [Lindner & Kolb 2006]</i>). (b) représente la mesure de profondeur d'une caméra Temps-de-Vol en fonction de la vérité terrain (en noir). La ligne dans le graphe représente la fonction qui approxime au mieux les points 2.5D issues du capteur Temps-de-Vol (<i>extrait de [Schiller et al. 2008]</i>). | 22 |
| 1.12 | Exemple de la distorsion de la mesure de profondeur calculée pour tous les pixels de l'image d'un mur situé à une distance égale à 2,45 m de la caméra (<i>extrait de [Kahlmann et al. 2006]</i>). | 23 |
| 2.1 | L'écart-type σ de la mesure de profondeur pour une image d'un mur blanc situé à une distance approximativement égale à 4 m, calculé pour 100 mesures de profondeur par pixel et un TI égal à 14 ms. | 26 |
| 2.2 | Comparaison de la variation en fonction de la profondeur de (a) l'écart-type σ de la mesure de profondeur calculé pour 100 mesures par pixel et (b) la fonction de modélisation du bruit $\frac{\sqrt{h}}{a}$ proposée dans [Edeler et al. 2010]. Les données utilisées correspondent à un mur blanc. | 28 |
| 2.3 | Principe général des méthodes de correction de la distorsion de la mesure de profondeur de la littérature. | 30 |

| | | |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.4 | Damier d'étalonnage avant (image à gauche) et après (image à droite) correction de la distorsion liée à la variation de la réflectivité (<i>extrait de [Lindner et al. 2010]</i>). | 34 |
| 2.5 | Exemple de caméras utilisées pour l'étalonnage de la mesure de profondeur : une caméra Temps-de-Vol couplée (a) avec trois caméras de haute résolution et (b) avec une caméra de haute résolution. | 37 |
| 3.1 | Réponse du modèle linéaire en fonction de $\frac{1}{a}$. σ prédit par ce modèle est tracé en fonction de σ empirique pour différents points. Les données correspondent à un mur blanc. Pour des grandes valeurs de σ , les points s'écartent de la droite $y = x$ | 42 |
| 3.2 | Diagramme de probabilité normale correspondant à un point 2.5D des données d'entrée. Les mesures (tracées en bleu) suivent approximativement la ligne droite de la distribution normale théorique (tracée en rouge). | 45 |
| 3.3 | Histogrammes de la distribution de quelques points 2.5D (d'un mur blanc). Chaque graphe représente la distribution de 100 mesures de profondeur. Graphiquement, ces histogrammes correspondent à des distributions normales, comme le confirme le test de Kolmogorov-Smirnov. | 46 |
| 3.4 | Distribution de l'écart-type à chaque pixel des images de profondeur (d'un mur blanc). σ est calculé pour chaque pixel à partir de 100 mesures de profondeur. σ varie approximativement de 0 à 30 mm avec une valeur moyenne égale à 12,5 mm. | 47 |
| 3.5 | Exemple montrant la dépendance du bruit de la réflectivité. (a) Image d'intensité d'un damier d'étalonnage noir et blanc. (b) L'écart type σ calculé pour chaque pixel à partir de 100 mesures de profondeur. Les carrés du damier de différentes couleurs situés à la même distance de la caméra n'ont pas le même σ (bruit plus élevé sur les carrés noirs). | 48 |
| 3.6 | Écart-type σ (mm) (d'une image de profondeur à une distance approximative de 3 m) pour différents <i>TIs</i> : (a) 14 ms, (b) 7 ms, (c) 3,5 ms, (d) 2 ms et (e) 1 ms. σ augmente de façon significative du centre de l'image, où l'illumination des objets observées est plus élevée, vers les bords de l'image. | 49 |

| | | |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.7 | Écart-type σ (mm) correspondant à des images de profondeur d'un mur blanc tracé en fonction de la profondeur, pour différents TIs : (a) 14 ms, (b) 7 ms, (c) 3,5 ms, (d) 2 ms et (e) 1 ms. La valeur moyenne de σ est tracée en points rouges et la fonction approximant ces données en ligne bleue. La valeur moyenne de σ varie approximativement de (a) 6 mm jusqu'à 22 mm, (b) 6 mm jusqu'à 35 mm, (c) 7 mm jusqu'à 64 mm, (d) 11 mm jusqu'à 110 mm et (e) 20 mm jusqu'à 200 mm respectivement. Les fonctions approximant les données, pour les différents TIs , sont tracées dans un même graphe dans (f). | 50 |
| 3.8 | Approximation de $\sigma_{TI=7}$, $\sigma_{TI=3,5}$, $\sigma_{TI=2}$ et $\sigma_{TI=1}$, respectivement, en fonction de $\sigma_{TI=14}$ entre 2 m et 6,5 m | 54 |
| 3.9 | Comparaison de (a) notre premier modèle position-amplitude noté $TPS_{(u,v,a)}$ avec (b) notre deuxième modèle position-profondeur- TI noté $TPS_{(u,v,d)}$. Les données correspondent à un mur blanc. Pour un modèle idéal, les points doivent être sur la droite $y = x$ | 56 |
| 3.10 | Comparaison du filtre proposé avec le filtre bilatéral sur des données réelles. Les résultats sont mieux visibles avec l'extraction des contours effectuée après le filtrage. Le filtrage bilatéral est effectué en faisant varier les valeurs de σ_{bf} : (g) 0,5 mm, (h) 1 mm, (i) 1,5 mm et (f) 3 mm. | 58 |
| 4.1 | Espace \mathbb{H} obtenu par discrétisation de l'espace continu \mathbb{G} avec un pas de distance régulier Δd . \mathbb{H} (ensemble des images de profondeur en bleu) obtenu à partir des observations (les images de profondeur en jaune). | 66 |
| 4.2 | Structure de la matrice \mathbf{M}_u . Cette matrice a une structure très creuse. Seules les lignes de la matrice ayant au moins un élément non-nul sont représentées. Les matrices \mathbf{M}_v et \mathbf{M}_d ont la même structure. | 68 |
| 4.3 | Structure des matrices \mathbf{L} et \mathbf{M} . Seules les lignes des matrices ayant au moins un élément non-nul sont représentées. Ces matrices ont une structure très creuse. | 69 |
| 4.4 | Principe du calcul de la vérité terrain par l'approche d'estimation de pose utilisant la paire de caméra stéréo (RGB-TOF) et un mur instrumenté. | 72 |

| | | |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.5 | Distorsion de la mesure de profondeur calculée pour trois différents points 2.5D de l'image et à différentes distances. La courbe en rouge correspond au pixel central et les courbes bleue et verte correspondent respectivement au pixel du coin supérieur gauche et au pixel du coin inférieur droit de la carte de profondeur. | 73 |
| 4.6 | Erreurs médianes avec différentes valeurs de $\lambda_u = \lambda_v$ et λ_d | 75 |
| 4.7 | Comparaison de l'erreur avant (a),(c) et après (b),(d) correction de la mesure de profondeur. | 76 |
| 4.8 | Comparaison de la distribution de l'erreur avant (a),(c) et après (b),(d) correction de la mesure de profondeur. | 77 |
| 5.1 | Les approches classiques nécessitent un grand ensemble de mesures de vérité terrain obtenu par un système de rail de mesure " <i>track line system</i> " ou un système d'extraction de cibles. Le premier système est coûteux. Le second système nécessite une mise en œuvre délicate, longue et ne permet pas d'obtenir des mesures d'une grande précision à de longues distances. L'approche proposée utilise des vues d'un plan et nécessite très peu de correspondances de points entre la vérité terrain et les images du capteur (4 points suffisent). | 82 |
| 5.2 | Données d'entrée de notre méthode d'étalonnage projetées dans le plan XZ : un grand ensemble de vues d'un mur plan issues du capteur Temps-de-Vol qui s'intersectent pour couvrir tout l'espace d'étalonnage. | 85 |
| 5.3 | Exemple d'une des vues du mur utilisée lors de l'étalonnage. Les résultats des deux premières itérations sont illustrés. La surface constituée des points 3D est présentée en rouge. Le plan Π_i qui approxime au mieux ces points est présenté en vert. | 90 |
| 5.4 | Résultats avec les données simulées durant le processus d'étalonnage de la mesure de profondeur. (a) Une partie de l'espace d'étalonnage allant de 1 m à 2,5 m. Une section de (b) est considérée pour montrer les résultats obtenus à (c) la première itération, (d) la seconde itération et (e) la dernière itération de l'étape de CNP. (f) Données d'entraînement obtenues après l'étape de CNP et les points utilisés pour estimer A. Comparaison des résultats (g) avant et (h) après l'étape de CA. | 96 |
| 5.5 | Une vue fronto-parallèle d'un plan (a) non distordu (vérité terrain) (b) avant correction, (c) après CNP et (d) après CA. | 99 |
| 5.6 | Comparaison de la distribution de la distance point-plan pour tous les points des images de validation avant et après correction. | 101 |

| | | |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 5.7 | Une vue d'un mur blanc (à gauche) avant la correction de la distorsion de la mesure de profondeur et (à droite) après l'étape de CNP. La couleur représente la distance d'un point à un plan mesurée en mm. | 101 |
| 6.1 | Calcul des dimensions du damier d'étalonnage (longueur et largeur) à partir de points 2.5D : (a) image d'intensité (non utilisée dans les calculs) et (b) image de profondeur après correction de la distorsion optique. Les valeurs de la longueur et de la largeur calculées sont respectivement égales à $1211 \pm 11,18$ mm et à $903 \pm 8,97$ mm. Ces mesures sont obtenues en utilisant notre deuxième méthode d'étalonnage. (c) l'ellipse d'incertitude de \mathbf{p}_1 et l'ellipsoïde d'incertitude de (d) \mathbf{q}_1 et (e) de \mathbf{Q}_1 sont présentées. Leurs ellipses d'incertitude associées sont tracées en bleu autour des ellipsoïdes. | 112 |
| A.1 | Écart-type σ (mm) d'une image de profondeur à une distance approximative de : (a) 1,5 m, (b) 2 m, (c) 3 m et (d) 3,5 m. Les formes des 4 figures sont différentes. Contrairement aux caméras Temps-de-Vol, la position du pixel dans l'image n'est pas un facteur de variation de σ pour les caméras Kinect. | 122 |
| A.2 | Écart-type σ (mm) correspondant à des images de profondeur d'un mur blanc tracé en fonction de la profondeur. La valeur moyenne de σ est tracée en points rouge et la fonction approximant ces données en ligne bleue. La valeur moyenne de σ augmente en fonction de la profondeur approximativement de 2,5 mm jusqu'à 25 mm. | 123 |
| B.1 | Modèle sténopé et repères associés. Le repère associé à la caméra est choisi de telle sorte que son centre soit placé en C , son axe z confondu avec l'axe optique et l'axe x parallèle aux bords horizontaux de l'image. Le repère image (dans lequel les coordonnées sont exprimées en pixels) a son centre A dans le coin supérieur gauche de l'image, ses axes u et v sont orientés respectivement de la gauche vers la droite et du haut vers le bas. | 126 |
| B.2 | Image issue de la caméra Temps-de-Vol utilisée dans ce mémoire : (a) avant et (b) après correction de la distorsion optique. | 130 |
| B.3 | Erreur de reprojection : distance entre l'observation 2D (m de coordonnées \mathbf{p}) d'un point 3D (M de coordonnées \mathbf{P}) et sa projection \hat{m} dans l'image de coordonnées $\mathbf{P}\check{\mathbf{P}}$ | 134 |

| | | |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| B.4 | Étalonnage stéréo : utilisation d'une mire d'étalonnage plane. Les paramètres stéréo $\mathbf{R}_{L \rightarrow R}$ et $\mathbf{T}_{L \rightarrow R}$ sont déterminés en minimisant l'erreur de reprojection δ . δ exprimée pour un point M est la distance géométrique dans l'image entre la projection du point m_L dans l'image droite notée \widehat{m}_R et le point m_R réellement observé dans l'image. | 136 |
| B.5 | Exemple de déformation par fonction de type B-spline de dimension 2D. (a) Modèle de la déformation synthétique : rotation d'une région d'intérêt de taille 2×2 d'une grille dont les coins sont fixés. (b) Déformation de la grille par ce modèle synthétique.(c) Déformation estimée avec la fonction B-spline. (<i>extrait de [Gay-Bellile 2008]</i>) . . . | 140 |
| B.6 | Exemple de déformation par fonction de type plaque mince de dimension 2D. (<i>extrait de [Gay-Bellile 2008]</i>) | 141 |
| B.7 | Transformation d'un point 2.5D \mathbf{q} vers un point 3D \mathbf{Q} | 142 |

Liste des tableaux

| | | |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.1 | Comparaison avec deux approches de la littérature | 77 |
| 5.1 | Erreur RMS calculée après CNP en faisant varier les valeurs des hyper-paramètres : l allant de 3^5 à 8^3 et λ de 10^{-4} à 10^4 en échelle logarithmique. Les cas où le système ne converge pas sont représentés par N. A partir de $l = 5^3$ (5 centres suivant $X-$, $Y-$ et $Z-$), le système est stable pour un large intervalle de λ | 98 |
| 5.2 | Erreur RMS (mm). | 98 |
| 5.3 | Comparaison avec la première méthode et deux approches de la littérature. | 102 |
| 5.4 | Comparaison de la correction de la planarité avec la première méthode proposée. | 103 |
| 6.1 | Comparaison des dimensions du damier (largeur et longueur) après correction de la mesure de profondeur avec nos deux méthodes d'étalonnage et de l'incertitude associée. | 113 |
| 6.2 | Erreur résiduelle moyenne des dimensions du damier (largeur et longueur) calculée avant et après correction de la mesure de profondeur. La correction est estimée avec les deux méthodes d'étalonnage proposées ainsi que les principales méthodes de l'état de l'art [Lindner & Kolb 2006] et [Kahlmann <i>et al.</i> 2006]. Un ensemble de 15 images du damier d'étalonnage positionné à différentes distances et à différentes positions dans l'image est considéré. | 114 |

Introduction

Une étape primordiale pour différentes applications d'interprétation visuelle consiste à reconstruire en 3D la scène observée. L'une des difficultés de cette étape est la construction en temps réel d'une carte de profondeur précise pour le point de vue courant. Il y a peu de temps, les systèmes disponibles étaient principalement des systèmes laser n'offrant pas le temps réel ou des systèmes stéréoscopiques (ou multi-vues) fonctionnant uniquement pour des scènes texturées. Au cours des dernières années, de nouvelles familles de caméras 3D sont apparues. Parmi celles-ci, on distingue généralement deux familles de technologie : les caméras Temps-de-Vol (ou caméras TOF) et les caméras à lumière structurée (*e.g.* la caméra Kinect). L'apparition de ces caméras ouvre des perspectives nouvelles pour l'interprétation de scène. En effet, ces systèmes de capture offrent une carte de profondeur en temps réel d'un environnement, que celui-ci soit texturé ou non, dynamique ou statique. Ceci permet d'améliorer la précision de certaines applications :

- Gestion des occultations dans les applications de suivi de personnes et d'objets [Zhang *et al.* 2012];
- Gestion des scènes non texturées en localisation et reconstruction 3D [Castaneda *et al.* 2011];
- Gestion des occultations et des collisions entre l'environnement réel et les objets virtuels introduits dans la scène en Réalité Augmentée [Dong & Kamat 2010];
- Robotique autonome [Biswas & Veloso 2012].

L'apparition récente des caméras actives 3D à bas coût, principalement la Kinect, a entraîné leur démocratisation. Leur expansion devrait d'autant plus s'accroître avec l'arrivée des caméras de faible encombrement, telles que la camera

Temps-de-Vol CamBoard pico de la marque PMD ([PMDTechnologies 2013]) et la caméra Capri de la marque PrimeSens ([PrimeSense 2013]).

Dans ce mémoire nous nous intéressons à l'étude des caméras de la première famille (Temps-de-Vol) et plus précisément celles basées sur la technologie de calcul 'lock-in pixels' qui sont les plus répandues. Les cartes de profondeur fournies par les capteurs Temps-de-Vol sont généralement entachées d'erreurs. En effet, ce type de capteur a des limites propres à sa technologie affectant la qualité des données qu'il fournit. La limite principale est le manque de précision des mesures de profondeur. Cette imprécision est due à deux facteurs :

- Incertitude de la mesure de profondeur dû au bruit statistique du capteur ;
- Distorsion de la mesure de profondeur principalement systématique.

Cependant, dans les travaux de la littérature, le bruit est très peu étudié et les méthodes d'étalonnage de la mesure de profondeur pour la correction de la distorsion systématique sont généralement difficiles à mettre en œuvre et ne garantissent pas la précision requise pour certaines applications. Dans de nombreux cas une connaissance plus approfondie de ces deux types d'erreur permettrait d'améliorer les performances de l'application finale utilisant les données des caméras Temps-de-Vol. Il est donc important de modéliser correctement un capteur actif 3D afin de prendre en compte au mieux le bruit et la distorsion de la mesure de profondeur.

L'objectif de ce mémoire est d'abord d'étudier les deux types d'erreur des caméras Temps-de-Vol susmentionnés : le bruit de la caméra et la distorsion de la mesure de profondeur et de fournir ensuite une modélisation et un étalonnage précis et simple à mettre en œuvre.

Contributions. Nos contributions s'articulent autour de deux parties distinctes :

- Étude empirique et modélisation du bruit des caméras Temps-de-Vol. L'incertitude des mesures due au bruit signifie que différentes mesures d'un même point effectuées en maintenant la position de la caméra fixe ne seront pas identiques. Il est important d'avoir un modèle de bruit qui donne pour chaque mesure de profondeur l'incertitude associée afin de pouvoir qualifier la qualité de la mesure.
- Modélisation et étalonnage des mesures de profondeur des caméras Temps-de-Vol. La distorsion de ces mesures est due à plusieurs facteurs. Le principal facteur est l'*erreur systématique* inhérente à la technologie Temps-de-Vol.

Pour la première partie, une caractérisation du bruit des caméras Temps-de-Vol est d'abord présentée où nous démontrons empiriquement que ce bruit suit une distribution normale. S'agissant de la modélisation du bruit, les travaux de la littérature considèrent que ce dernier ne dépend que de l'amplitude du signal reçu.

Dans ce mémoire, nous proposons donc de vérifier cette hypothèse. Nous montrons que l'amplitude seule n'est pas suffisante et nous intégrons d'autres paramètres, tels que la position dans l'image, la profondeur, le Temps d'Intégration (TI) et la réflectivité pour modéliser le bruit.

Dans la seconde partie, deux méthodes d'étalonnage de la mesure de profondeur sont proposées, chacune présentant une solution différente. La première vise à fournir un modèle précis quitte à nécessiter une mise en œuvre délicate, alors que la seconde privilégie la simplicité de mise en œuvre. En effet, la première est basée entièrement sur des mesures de vérité terrain (nécessitant une mise en œuvre délicate pour les obtenir), tandis que la deuxième est basée principalement sur une vérité terrain structurelle :

- La première méthode que nous présentons est basée sur une modélisation non-paramétrique permettant d'augmenter la précision de l'étalonnage.
- La deuxième méthode que nous présentons est basée principalement sur l'intégration d'une connaissance a priori de la géométrie plane de la scène observée. Elle présente l'avantage d'une plus grande facilité de mise en œuvre et la possibilité de faire l'étalonnage en temps réel.

A travers un exemple d'application, nous montrons que le modèle de la distorsion de la mesure de profondeur et le modèle de bruit peuvent être combinés afin d'augmenter la précision des mesures et de quantifier leur précision en fournissant l'incertitude associée. Nous comparons également nos méthodes avec celles de l'état de l'art. Les résultats obtenus montrent leurs validités et leurs précisions.

Organisation du mémoire. En premier lieu, le chapitre 1 présente et caractérise les caméras Temps-de-Vol et plus précisément celles basées sur la technologie de calcul '*lock-in pixels*'. Cette dernière correspondant à la caméra utilisée aux cours de nos travaux. En particulier, les facteurs affectant la qualité de leurs données sont détaillés. Le chapitre 2 présente l'état de l'art de la modélisation du bruit interne des caméras Temps-de-Vol et de l'étalonnage de la mesure de profondeur.

La structure de la suite du mémoire est guidée par les trois contributions proposées. Dans le chapitre 3 nous commençons par caractériser le bruit des caméras Temps-de-Vol et déterminer ses facteurs de variation. Ensuite, nous proposons deux modèles qui donnent la valeur du bruit pour chaque mesure du capteur Temps-de-Vol. Chacune de nos deux méthodes d'étalonnage de la mesure de profondeur est détaillée, respectivement dans le chapitre 4 et le chapitre 5. Le chapitre 6 présente une simple application illustrant l'intérêt de nos travaux.

Nous dressons enfin un bilan des travaux réalisés et présentons différentes perspectives envisageables pour les travaux futurs.

CHAPITRE 1

Principe des caméras Temps-de-Vol

Ce chapitre a pour but de présenter les caméras Temps-de-Vol et plus précisément celle utilisée dans nos travaux. Nous commençons par présenter leur principe et les différentes technologies de calcul du Temps-de-Vol. Nous nous concentrons ensuite sur les caméras basées sur la technologie de calcul "lock-in pixels". C'est la technologie de la caméra utilisée dans nos travaux. Nous détaillons leur principe de fonctionnement, leurs données de sorties et leurs paramètres de capture. Finalement, nous énumérons leurs principales limites : la faible résolution, la courte portée et le manque de précision des mesures de profondeur. La dernière limitation est due principalement à deux facteurs : le bruit stochastique et la distorsion systématique.

1.1 Introduction

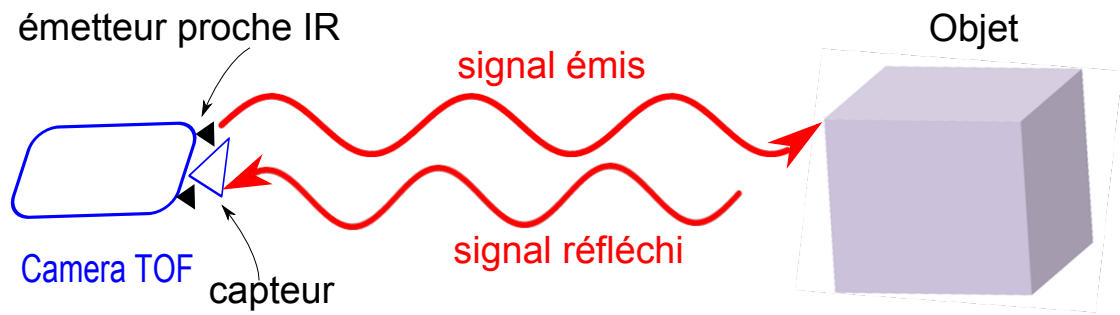
Les caméras Temps-de-Vol ("Time-Of-Flight", TOF) sont des capteurs actifs qui fournissent des images de profondeur en temps-réel. Comme l'illustre la figure 1.1a, une caméra Temps-de-Vol envoie un signal optique proche infrarouge (IR). Le signal réfléchi par les objets de la scène est ensuite détecté par le capteur qui calcule la profondeur [Lange 2000]. Il existe deux technologies de calcul du Temps-de-Vol. La première se base sur des impulsions (pulsed TOF) : elle mesure le temps que met le signal pour effectuer le trajet entre l'objet et la caméra. La deuxième se base sur des ondes de modulation continues (Continuous-Wave (CW)) : elle mesure le déphasage entre le signal émis et celui réfléchi par démodulation synchrone du signal réfléchi. Nous nous intéressons dans nos travaux aux caméras qui sont les plus commercialisées, les caméras de la deuxième technologie (CW modulation), et plus précisément à celles qui sont basées sur la démodulation de type "*lock-in pixels*" ([Spirig *et al.* 1995]). Leur principe est basé sur la détection et la sauvegarde des charges photo-électriques par synchronisation des électrodes des pixels. Les caméras de cette technologie sont les plus utilisées : elles présentent l'avantage d'acquérir les images à haute fréquence, de consommer peu d'énergie et d'avoir un plus faible poids. Nous nous concentrons dans ce chapitre sur les caméras Temps-de-Vol de type "*lock-in pixels*".

1.2 Fonctionnement

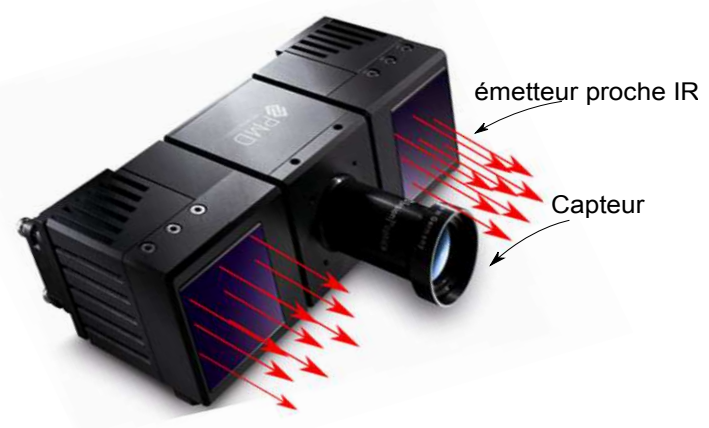
La figure 1.1b présente un exemple de caméra basée sur la technologie de type "*lock-in pixels*". Le capteur de cette caméra est situé au milieu de ses deux émetteurs proche Infra Rouge (IR). Le signal réfléchi est modélisé par un signal sinusoïdale de phase φ et d'amplitude a . Pour estimer ces deux paramètres, la démodulation du signal réfléchi est réalisée. Cela est effectué par corrélation avec le signal de modulation. La fonction de corrélation est définie comme suit :

$$c(\tau) = s(t) \otimes g(t) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{+T/2} s(t)g(t + \tau)dt, \quad (1.1)$$

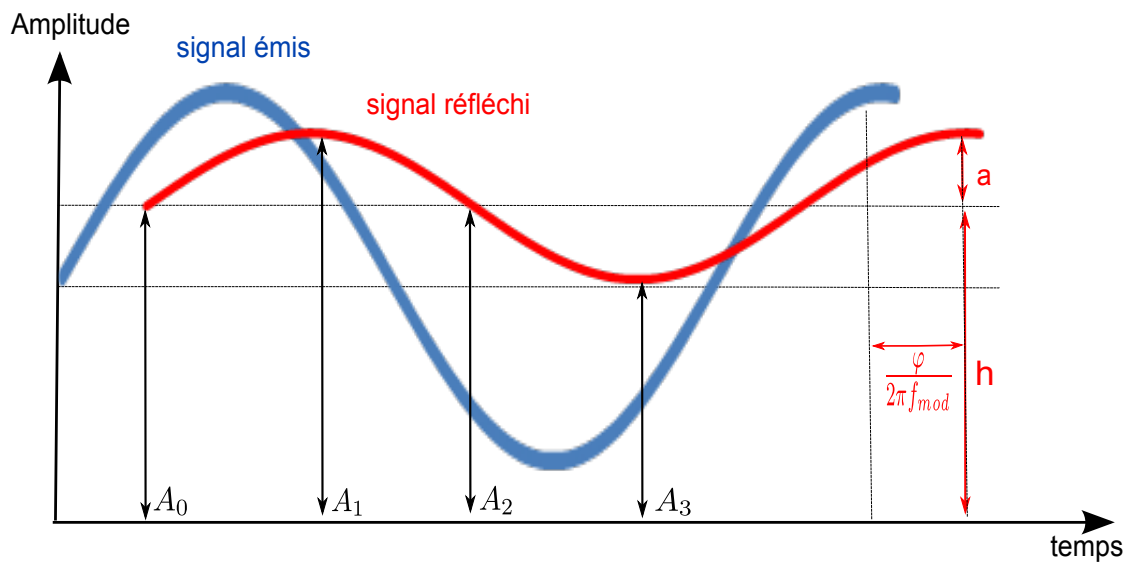
où T est la période du signal, $s(t) = b + a \cos(2\pi f_{mod}t - \varphi)$ représente le signal réfléchi d'amplitude de modulation a et de phase φ et $g(t) = \cos(2\pi f_{mod}t)$ est le signal de démodulation. Notons que f_{mod} est la fréquence de modulation qui est un paramètre de capture à configurer dans les caméras Temps-de-Vol (présenté plus en détail dans la section 1.4).



(a) Principe de fonctionnement des caméras Temps-de-Vol : émission d'un signal proche IR qui est par la suite réfléchi et traité par le capteur pour calculer la profondeur.



(b) Exemple de caméra Temps-de-Vol : PMD Vision CamCube3.0, 204×204 . Elle est basée sur la technique de démodulation de type "lock-in pixels".



(c) Le déphasage φ est calculé par démodulation (réalisée par corrélation avec le signal de modulation) du signal réfléchi (en rouge). En bleu le signal émis, a l'amplitude, h l'intensité, A_0, A_1, A_2, A_3 différentes valeurs de la fonction de corrélation.

FIGURE 1.1 – Principe du Temps-de-Vol.

La fonction de corrélation peut être calculée comme suit :

$$\begin{aligned} c(\tau) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{+T/2} [b + a \cos(2\pi f_{mod}t - \varphi)][\cos(2\pi f_{mod}t + 2\pi f_{mod}\tau)] dt \\ &= \frac{a}{2} \cos(2\pi f_{mod}\tau + \varphi). \end{aligned} \quad (1.2)$$

La plupart du temps, dans le signal réfléchi $s(t)$, il y a un offset qui représente l'illumination du fond de l'image. Un offset h doit alors être ajouté à la fonction de corrélation pour obtenir la valeur réelle mesurée. Pour calculer l'amplitude a et la phase φ du signal $s(t)$, la fonction de corrélation doit être évaluée pour différentes valeurs de phases $\tau_0 = 0, \tau_1 = \frac{\pi}{2}, \tau_2 = \pi, \tau_3 = \frac{3\pi}{2}$. Notons A_0, A_1, A_2, A_3 les différentes valeurs de la fonction de corrélation correspondant respectivement à ces valeurs de phases. La figure 1.1c illustre ces grandeurs :

$$A_0 = c(\tau_0) + h = \frac{a}{2} \cos(\varphi) + h \quad (1.3)$$

$$A_1 = c(\tau_1) + h = -\frac{a}{2} \sin(\varphi) + h \quad (1.4)$$

$$A_2 = c(\tau_2) + h = -\frac{a}{2} \cos(\varphi) + h \quad (1.5)$$

$$A_3 = c(\tau_3) + h = \frac{a}{2} \sin(\varphi) + h \quad (1.6)$$

A partir des valeurs de $A_0, A_1, A_2, A_3, \varphi$ et a sont calculés comme suit :

$$\varphi = \arctan \frac{A_3 - A_1}{A_0 - A_2} \quad (1.7)$$

$$a = \frac{\sqrt{[A_3 - A_1]^2 + [A_0 - A_2]^2}}{2} \quad (1.8)$$

Cette technique de démodulation est connue sous le nom de l'échantillonnage "four-bucket". Une fois φ mesuré, la profondeur d pour un pixel donné est obtenue comme suit :

$$d = \frac{c\varphi}{4\pi f_{mod}}, \quad (1.9)$$

où c représente la vitesse de la lumière. L'offset h de la fonction de corrélation représente la moyenne du signal reçu et est donné par :

$$h = \frac{A_0 + A_1 + A_2 + A_3}{4} \quad (1.10)$$

Il est fréquent d'appeler h l'intensité du signal, qui est aussi la somme de a et b .

En plus de la profondeur d , l'amplitude a et l'intensité h , le capteur Temps-de-Vol fournit le nuage de points 3D. Cependant, si les paramètres intrinsèques de la caméra sont calculés par l'utilisateur (et sont généralement plus précis que ceux fixés par le constructeur), il est aussi possible de convertir l'image de profondeur en un nuage de points 3D en utilisant l'équation (B.37) introduite dans la section B.3.1 de l'Annexe B.

1.3 Données de sortie

Les données présentées précédemment (la distance d , l'amplitude a , l'intensité h et les coordonnées 3D) sont calculées pour chaque pixel du capteur Temps-de-Vol. Ce dernier les manipule sous forme matricielle. Par conséquent, il peut fournir, à chaque période de modulation, trois images différentes et un nuage de points 3D. Les données sont illustrées par la figure 1.2 :

- L'image de profondeur (figure 1.2a) contient les profondeurs d représentant la distance en mètres entre les objets de la scène et le centre optique du capteur : plus l'objet est noir plus il est proche.
- L'image d'amplitude (figure 1.2b) représente les valeurs d'amplitude a mesurées en volts pour chaque pixel de l'image. Dans cette image, plus le pixel est blanc plus son amplitude est élevée. L'amplitude a du signal réfléchi est atténuée par rapport à l'amplitude du signal émis. Ceci est dû à plusieurs facteurs, principalement la réflectivité de l'objet : si l'indice de réflectivité est élevé alors la valeur de l'amplitude est élevée. Cette valeur donne une information sur la fiabilité de la mesure de profondeur : si l'amplitude d'un pixel est trop faible, cela veut dire que la proportion d'énergie renvoyée et le rapport signal/bruit sont faibles et par conséquent que la mesure de profondeur associée à ce pixel est peu fiable. Les autres facteurs sont la distance d'observation, l'orientation de la surface des objets observés *etc.* .
- L'image d'intensité (figure 1.2c) représente les valeurs d'intensité h mesurées en volts pour chaque pixel de l'image. Elle peut être utilisée comme une image de niveaux de gris : par exemple, cette image est généralement utilisée pour l'étalonnage de la caméra (détermination des paramètres intrinsèques détaillée dans la section B.1.3.1). Dans cette image, plus le pixel est illuminé plus sa valeur d'intensité est élevée.
- Le nuage de points 3D (figure 1.2d) représente les coordonnées 3D en mètres de chaque pixel de l'image mesurée par le capteur.

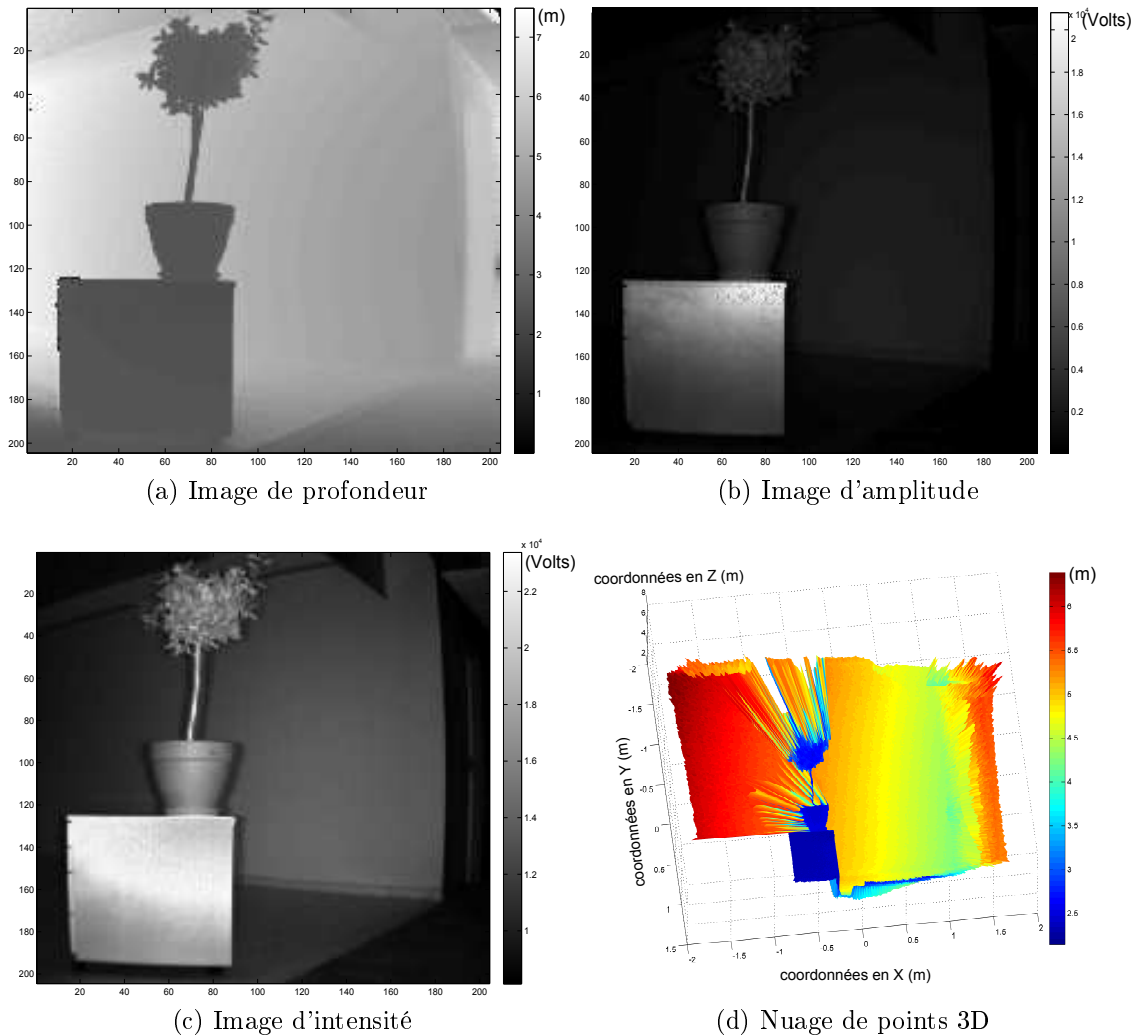


FIGURE 1.2 – Données fournies par une caméra Temps-de-Vol : (a) image de profondeur, (b) image d'amplitude, (c) image d'intensité et (d) nuage de points 3D.

1.4 Paramètres de capture

Comme toute caméra, les caméras Temps-de-Vol ont des paramètres de capture à configurer au cours des acquisitions qui vont influencer la qualité des images de profondeur. Il y a deux paramètres principaux : le Temps d'Intégration (TI ¹)

1. TI (Temps d'Intégration) désigne la durée de l'exposition : l'intervalle de temps pendant lequel le capteur laisse passer la lumière. C'est un paramètre qui peut être configuré par l'utilisateur ou ajusté automatiquement (dans le cas de quelques caméras Temps-de-Vol).

et la fréquence de modulation f_{mod} . Le paramètre TI doit être ajusté de telle sorte à maximiser la quantité de lumière qui atteint le capteur sans le saturer. La valeur optimale de TI dépend de l'indice de réflectivité de l'objet, de la quantité de lumière de la scène et de la puissance des sources de lumière de la caméra. Il faut donc trouver le bon compromis en fonction de la nature de la scène. La figure 1.3 illustre les cartes de profondeur fournies par la caméra Temps-de-Vol en faisant varier la valeur de TI . Trois valeurs de TI : 0,14 ms, 1,4 ms et 14 ms sont

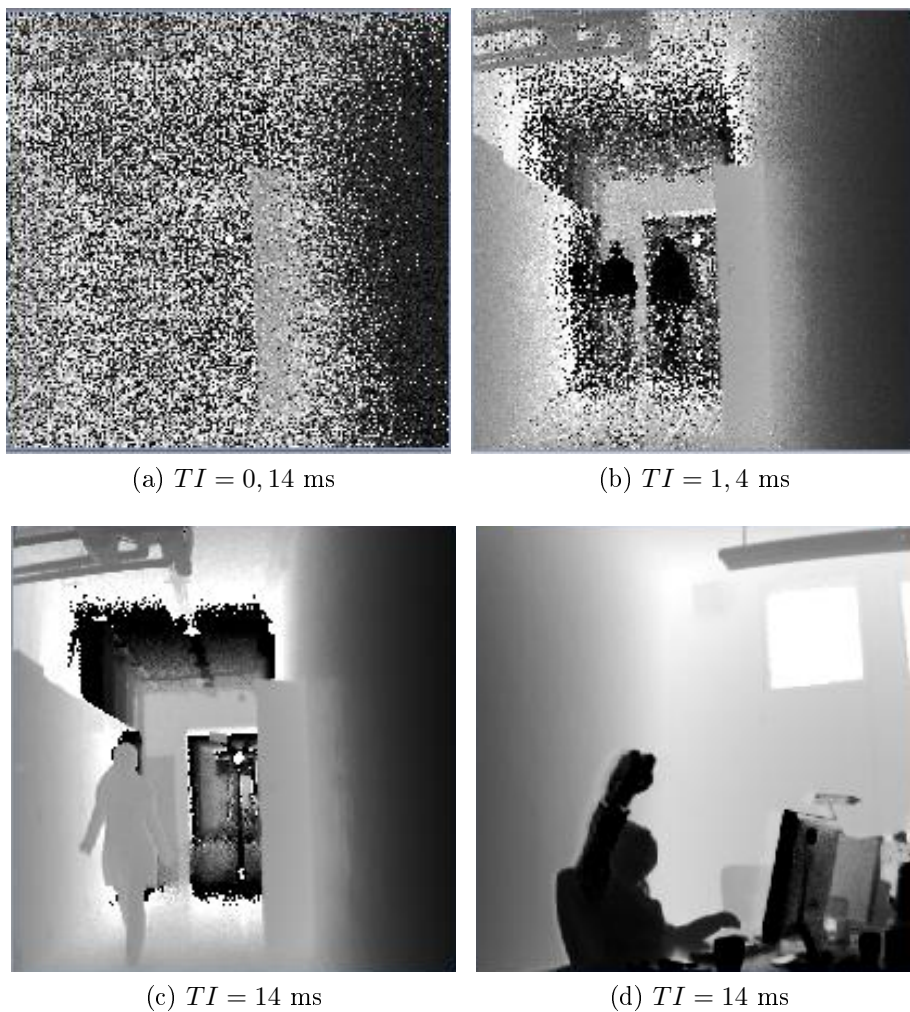


FIGURE 1.3 – Acquisitions dans différents contextes en faisant varier TI dans une scène statique (a) 0,14 ms, (b) 1,4 ms, (c) 14 ms et dans une scène dynamique (d) 14 ms (mouvement de la main). TI influence la qualité de l'image de profondeur et doit être ajusté à la nature de la scène (statique ou dynamique).

choisies. En comparant les trois premières images de la figure, nous observons que la meilleure qualité est obtenue avec le plus long TI , égal à 14 ms. En effet, un TI court implique une amplitude a faible et un bruit affectant les images de profondeur plus élevé. Par contre, un TI trop long provoque un flou de bougé comme le montre la figure 1.3d (au niveau de la main). En conclusion, nous pouvons dire que TI doit être ajusté à la nature de la scène. Dans le cas d'une scène statique, un TI égal à 14 ms est adéquat (*i.e.* donne une bonne qualité d'image). Par contre, dans le cas d'une scène dynamique, un TI moins élevé doit être utilisé au détriment du rapport signal/bruit. Des acquisitions ont aussi été faites à l'extérieur, les images obtenues sont illustrées par la figure 1.4. Elles sont très bruitées quel que soit le TI utilisé. C'est pour cette raison que la majorité des applications utilisant des capteurs Temps-de-Vol se limitent à des contextes intérieurs.

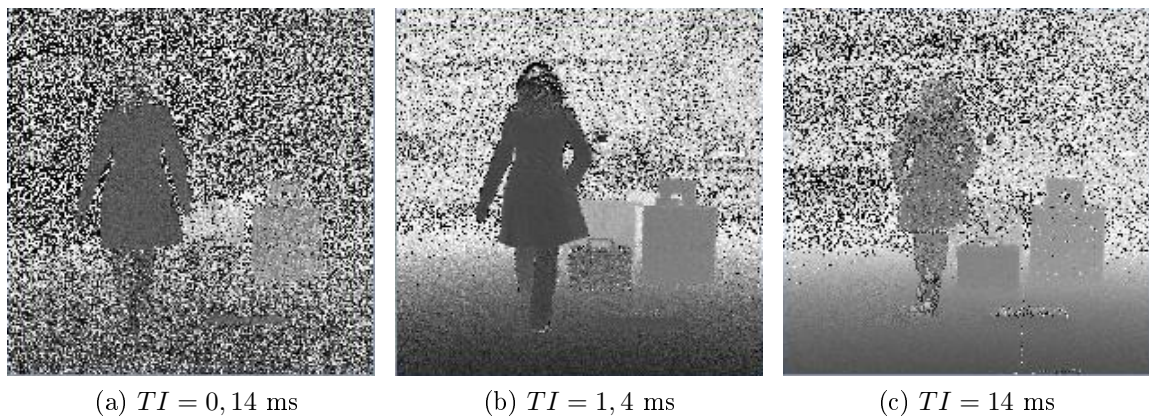


FIGURE 1.4 – Acquisitions extérieures en faisant varier TI (a) 0,14 ms, (b) 1,4 ms et (c) 14 ms. Les images sont très bruitées. Ceci est dû principalement à l'interférence avec plusieurs autres sources de lumière.

Le deuxième paramètre de capture des caméras Temps-de-Vol à configurer est la fréquence de modulation f_{mod} . La figure 1.5 illustre les cartes de profondeur d'une même scène fournies par la caméra Temps-de-Vol en variant la valeur de f_{mod} . Quatre valeurs de f_{mod} sont utilisées : 18, 19, 20 et 21 MHz (valeurs autorisées par le constructeur). Nous observons sur ces figures l'aspect périodique : les objets situés à une distance supérieure à la portée maximale (appelée distance ambiguë qui sera expliquée en détail dans la section 1.5.2) de la caméra vont être détectés comme très proches de la caméra. Par exemple, la profondeur mesurée de la personne sur la figure 1.5d est égale approximativement à 9 m, alors que la caméra mesure 2 m. En conclusion, nous pouvons dire que le paramètre f_{mod} , contrairement au TI , n'a pas d'influence sur la qualité de l'image. De plus, les valeurs autorisées

par le constructeur sont très proches (*e.g.* 18, 19, 20 et 21 MHz). Ce paramètre

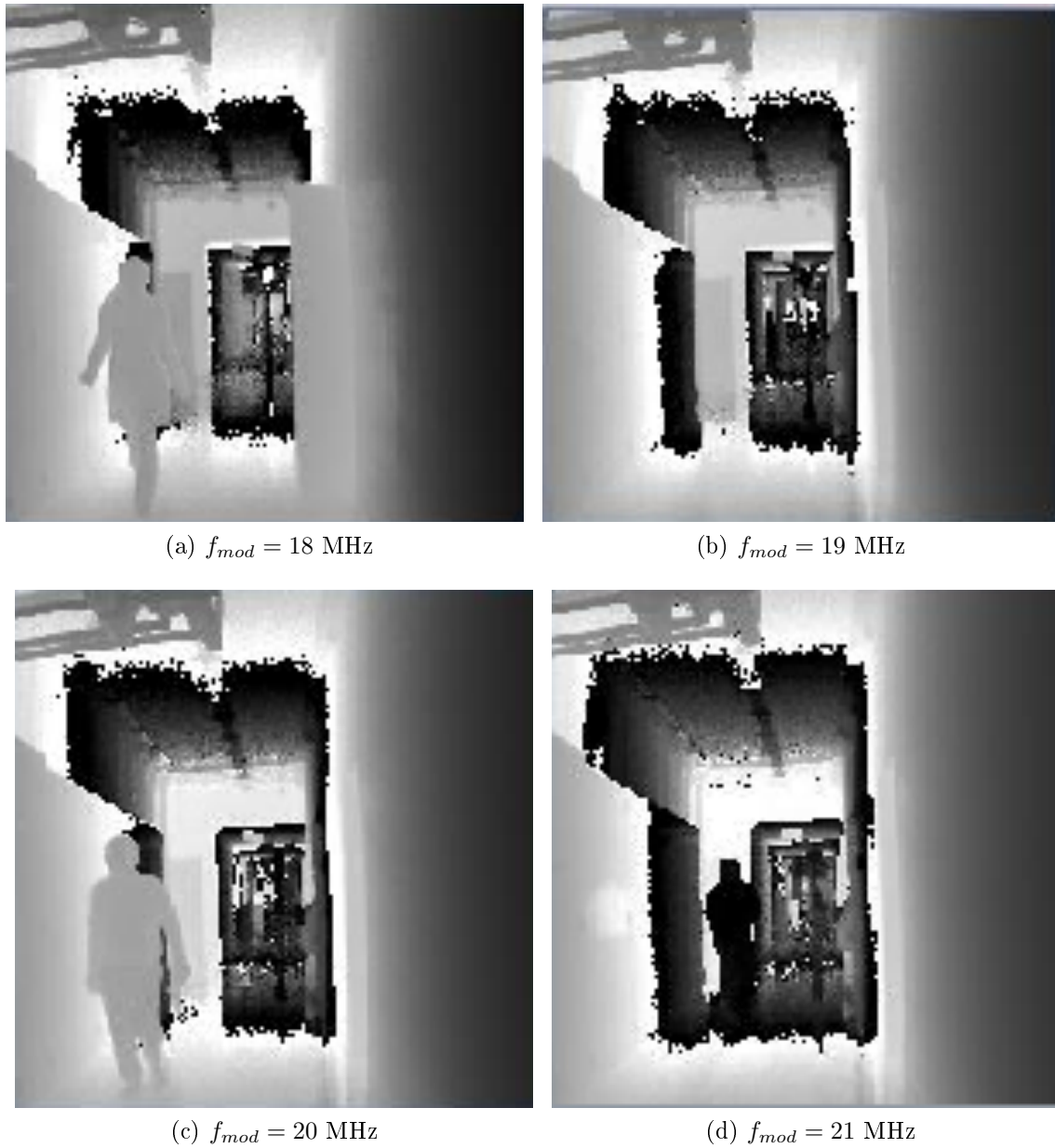


FIGURE 1.5 – Acquisitions en faisant varier f_{mod} (a) 18 MHz, (b) 19 MHz, (c) 20 MHz et (d) 21 MHz. f_{mod} n'a pas d'influence sur la qualité de l'image de profondeur. Par contre, elle change les valeurs des mesures de profondeur ainsi que la portée de la caméra.

change les valeurs des mesures de profondeur ainsi que la portée de la caméra. Il est donc important de fixer sa valeur pour effectuer les acquisitions. La valeur souvent utilisée dans la littérature est 20 MHz.

Caméra et paramètres utilisés. Dans la suite de ce mémoire, la caméra utilisée dans les différentes expérimentations est la PMD Vision CamCube3.0 de résolution 204×204 illustrée par la figure 1.6f. Elle est utilisée en intérieur, dans le cadre de scènes statiques et les valeurs réglées par défaut pour f_{mod} et TI sont respectivement 20 MHz (valeur souvent utilisée dans la littérature) et 14 ms (valeur qui offre les images de meilleure qualité).

1.5 Limites du capteur Temps-de-Vol

Les caméras Temps-de-Vol basées sur une modulation CW ont des limites propres à leur technologie. Nous énumérons dans cette section les trois principales limites : la résolution, la portée et le manque de précision des mesures de profondeur.

1.5.1 Résolution

La première limite des caméras Temps-de-Vol est la faible résolution. Nous présentons dans la figure 1.6 quelques exemples de caméras Temps-de-Vol commercialisées. La résolution des caméras disponibles aujourd'hui est faible : la plus haute résolution qui existe actuellement est de 204×204 pixels. C'est une résolution très faible comparée aux résolutions des caméras 2D utilisées dans les applications de vision. De plus, plus la résolution de la caméra est faible, plus la probabilité d'avoir des pixels couvrant des zones de profondeur non homogènes est élevée. Ceci implique un manque de précision au niveau de ces zones (*e.g.* au niveau des contours des objets). Dans la littérature, il existe plusieurs travaux qui proposent d'augmenter la résolution des images de profondeur des caméras Temps-de-Vol ("*depth super resolution*") en fusionnant les données de cette dernière avec celles d'une caméra standard de plus haute résolution [Yang *et al.* 2007a, Garro *et al.* 2009, Huhle *et al.* 2010, Park *et al.* 2011, Huhle *et al.* 2007] ou avec celles d'une paire de caméras standards (caméra stéréo) de plus haute résolution [Zhu *et al.* 2008a, Mutto *et al.* 2010, Yang *et al.* 2010, Kuhnert & Stommel 2006, Yang *et al.* 2007b, Zhu *et al.* 2010, Zhu *et al.* 2011, Guömundsson *et al.* 2008].

La faible résolution pose aussi un problème lors de l'étalonnage standard (estimation des paramètres intrinsèques de la caméra détaillée dans la section B.1.3.1

de l'Annexe B) d'une telle caméra ainsi que lors de son étalonnage stéréo (estimation des paramètres extrinsèques détaillées dans la section B.1.3.3 de l'Annexe B). Ces méthodes d'étalonnage requièrent une extraction précise des caractéristiques de l'objet d'étalonnage. Cependant, les pixels des images des caméras Temps-de-Vol peuvent couvrir des distances non homogènes (*e.g.* au niveau des contours des objets). Ce problème ne peut pas être traité par les méthodes d'augmentation de la résolution citées précédemment.



FIGURE 1.6 – Quelques exemples de caméras Temps-de-Vol basées sur la modulation de type "lock-in pixels". Leur résolution spatiale est faible : elle varie de 64×48 à 204×204 (actuellement la plus haute résolution disponible sur le marché).

1.5.2 Portée

La deuxième limite des caméras Temps-de-Vol est la courte portée (*i.e.* la plage de distance limitée). Cela est dû au fait que φ (équation (1.7)) est estimée par la fonction arctangente définie sur l'intervalle $]-\frac{\pi}{2}, \frac{\pi}{2}[$. Par conséquent, φ est incluse dans cet intervalle modulo 2π . Cependant, φ ne peut prendre que des valeurs positives (équation (1.9)). Il est donc possible de décaler et d'étendre l'intervalle de $]-\frac{\pi}{2}, \frac{\pi}{2}[$ à $]0, 2\pi[$ ce qui implique que la profondeur d soit incluse dans l'intervalle $]0, \frac{c}{2f_{mod}}[$ (équation (1.9)). Par exemple, si $f_{mod} = 20$ MHz alors $d \in]0 \text{ m}, 7,5 \text{ m}[$. La mesure de profondeur de la caméra sera alors estimée modulo 7,5 m. C'est-à-dire une mesure de profondeur de 2 m peut correspondre à un objet placé à 2 m modulo 7,5 m (2 m, 9,5 m *etc.*). En pratique, la profondeur retournée par l'équation (1.9) correspond au reste de la division de φ par 2π multiplié par $\frac{c}{2f_{mod}}$. Ce phénomène est appelé "*phase wrapping*" : il peut être considéré comme une transformation ("*wrapping*") périodique de φ avec un modulo 2π . $\frac{c}{2f_{mod}}$ est appelé distance ambiguë². Cette distance détermine donc la portée de la caméra et elle dépend de la valeur de f_{mod} (paramètre de capture à fixer présenté dans la section 1.4).

Il existe dans la littérature plusieurs travaux qui proposent des solutions pour étendre la portée des caméras Temps-de-Vol. Leur principe est de détecter les zones de dépliement de phase (dépassant la distance ambiguë) dans l'image. Une première approche présentée dans [Poppinga & Birk 2009] est basée sur la technique de seuillage, une deuxième famille d'approches présentée dans [Choi *et al.* 2010, McClure *et al.* 2010, Jutzi 2009, Droeschel *et al.* 2010b] est basée sur la segmentation d'image. Ces approches ne fonctionnent pas dans tous les cas de figure car elles font des hypothèses fortes sur la scène qui ne sont pas vérifiées en pratique, comme la continuité de la profondeur et l'invariance de la réflectivité. Une troisième famille utilise une méthode multi-fréquentielle (différentes fréquences de modulation) avec une seule caméra Temps-de-Vol [Droeschel *et al.* 2010a, Payne *et al.* 2009] ou avec deux caméras Temps-de-Vol [Choi & Lee 2012]. Cette dernière famille d'approches donne de meilleurs résultats comparée aux deux premières. Cependant, aujourd'hui les technologies disponibles ne permettent pas un contrôle dynamique de f_{mod} sur une même caméra. Cette famille d'approches requiert donc l'utilisation d'une paire de caméras Temps-de-Vol (avec une seule caméra, le traitement ne peut pas être fait automatiquement³) ce qui constitue un système à coût très élevé. Un exemple comparant

2. La distance ambiguë est la distance maximale à partir de laquelle le phénomène de dépliement de phase "*wrapping phase*" commence à être observé.

3. La caméra Temps-de-Vol réagit mal au changement automatique de f_{mod} . Avec la famille

des images de profondeur avant et après correction de dépliement de phase est présenté en figure 1.7. L'image de gauche représente l'image de profondeur mesurée par la caméra Temps-de-Vol. Le phénomène de dépliement de phase est observé sur les bords de l'image (les zones en noir) où la profondeur mesurée par la caméra est incorrecte : les objets dans cette zone sont estimés très proches. L'image de droite est le résultat obtenu après correction de dépliement de phase basée sur une méthode de segmentation par *graph-cuts*.

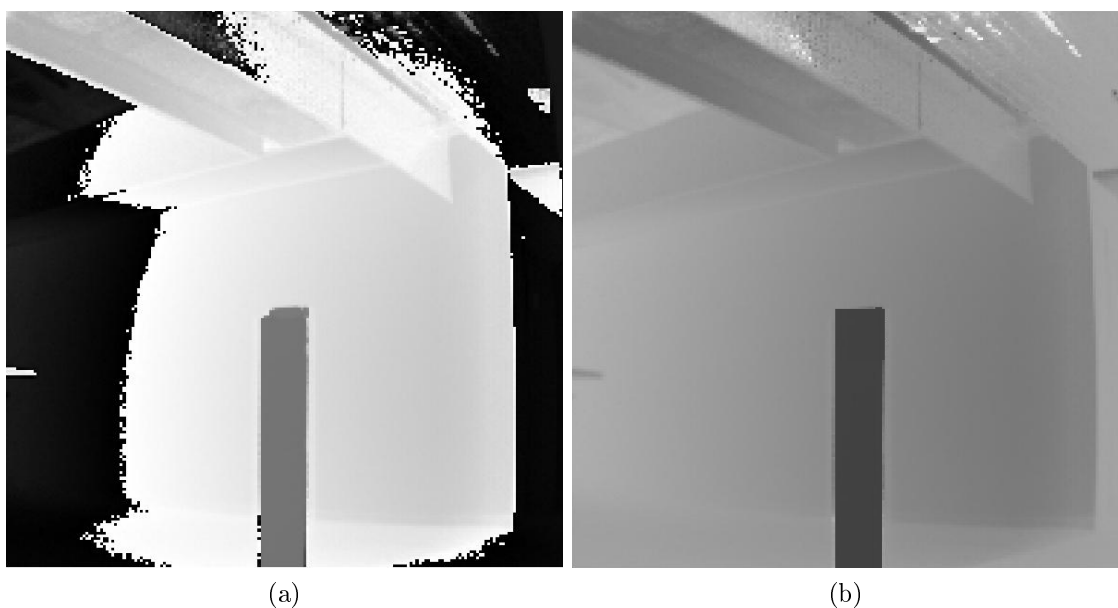


FIGURE 1.7 – Exemple d'image de profondeur (a) avec des objets situés à des distances supérieures à la distance ambiguë $\frac{c}{2f_{mod}}$. Le phénomène de dépliement de phase ("*unwrapping phase*") est observé aux bords de l'image (les zones noires). La même image est présentée dans (b) après correction de dépliement de phase.

1.5.3 Imprécision des mesures de profondeur

La troisième limite des caméras Temps-de-Vol est l'imprécision des mesures de profondeur. Deux sources d'imprécision peuvent être distinguées. La première est l'incertitude de mesure due au bruit du capteur. La deuxième est la distorsion affectant les images de profondeur. Notons que le bruit est stochastique alors que le facteur principal de la distorsion est systématique.

d'approche multi-fréquentielle utilisant une seule caméra Temps-de-Vol, le changement de f_{mod} doit être fait manuellement. Par conséquent, la correction de dépliement de phase ne peut pas se faire en ligne avec ces méthodes.

1.5.3.1 Incertitude

L'incertitude des mesures de profondeur est due au bruit affectant les images de profondeur. Le bruit peut avoir différentes sources. Les principales sources externes sont l'interférence avec d'autres sources de lumière et la faible réflectivité des objets de la scène. Le bruit causé par l'interférence avec d'autres sources de lumière est prédominant dans un contexte extérieur (voir figure 1.4), mais devient relativement négligeable en intérieur. C'est pour cette raison que les applications utilisant des caméras Temps-de-Vol se limitent actuellement à des contextes intérieurs. Le niveau du bruit dû à la faible réflectivité des objets peut être détecté à partir des mesures d'amplitude du signal réfléchi : le signal réfléchi par les objets de faible réflectivité est d'amplitude faible.

La deuxième source du bruit est liée au fonctionnement interne du capteur. C'est un bruit électronique causé par le fait que le courant électrique n'est pas continu mais constitué de porteurs de charge élémentaires (les électrons). Comme déjà expliqué dans la section 1.4, il dépend du TI (paramètre de capture). Plus le TI est court plus le bruit est élevé. Il est prédominant dans un environnement avec peu de sources de perturbation (d'autres sources de lumière) *i.e.* en contexte intérieur.

Les travaux existants présentent des solutions pour diminuer le bruit affectant les images de profondeur. Ces travaux sont principalement des méthodes de filtrage [Kim *et al.* 2010, Kim *et al.* 2006, Cho *et al.* 2007, Zhu *et al.* 2008b]. Cependant, il n'y a pas eu pour le moment de travaux qui caractérisent le bruit interne des caméras Temps-de-Vol. Les deux modèles de bruit présentés dans la littérature font l'hypothèse que le bruit suit une distribution normale sans démonstration ni validation empirique. De plus, ces modèles sont basés uniquement sur l'information d'amplitude. Il n'existe pas d'étude présentant les facteurs de variation du bruit du capteur. Ces modèles ainsi que leurs limites seront présentés en détail dans la section 3.2 du chapitre 2.

Il serait donc intéressant de caractériser et analyser les facteurs de variation du bruit interne du capteur Temps-de-Vol et de pouvoir modéliser et estimer l'incertitude pour chaque point de l'espace, pour chaque TI (un facteur de variation du bruit présenté précédemment dans la section 1.4) et pour chaque réflectivité. Cela constitue une partie de notre étude que nous détaillons dans le chapitre 3 de ce mémoire.

1.5.3.2 Distorsion

En plus de l'incertitude de mesure due au bruit interne du capteur, les caméras Temps-de-Vol présentent une distorsion de la mesure de profondeur. Le bruit pré-

senté précédemment est stochastique alors que le facteur principal de la distorsion est systématique. Cette distorsion est illustrée par la figure 1.8 qui présente un exemple d'une image de profondeur correspondant à un cube. Le pixel de coordonnées (x, y) représente la mesure de profondeur du point 3D noté \mathbf{Q} (la distance entre \mathbf{Q} et la caméra). Cette mesure, notée d_{TOF} , est à comparer à la profondeur réelle (vérité terrain notée d_{GT}). d_{TOF} correspond à une mesure distordue de d_{GT} à laquelle vient s'ajouter un bruit stochastique (dû au bruit interne du capteur). De ce fait, pour un point 2.5D donné, nous obtenons la relation suivante :

$$d_{TOF} = Dist(d_{GT}) + bruit \quad (1.11)$$

où $Dist$ est la fonction de la distorsion de la mesure de profondeur. Dans cette section, nous énumérons, tout d'abord, les facteurs de distorsion de la mesure de profondeur. Ensuite, nous nous intéressons aux caractéristiques de cette distorsion.

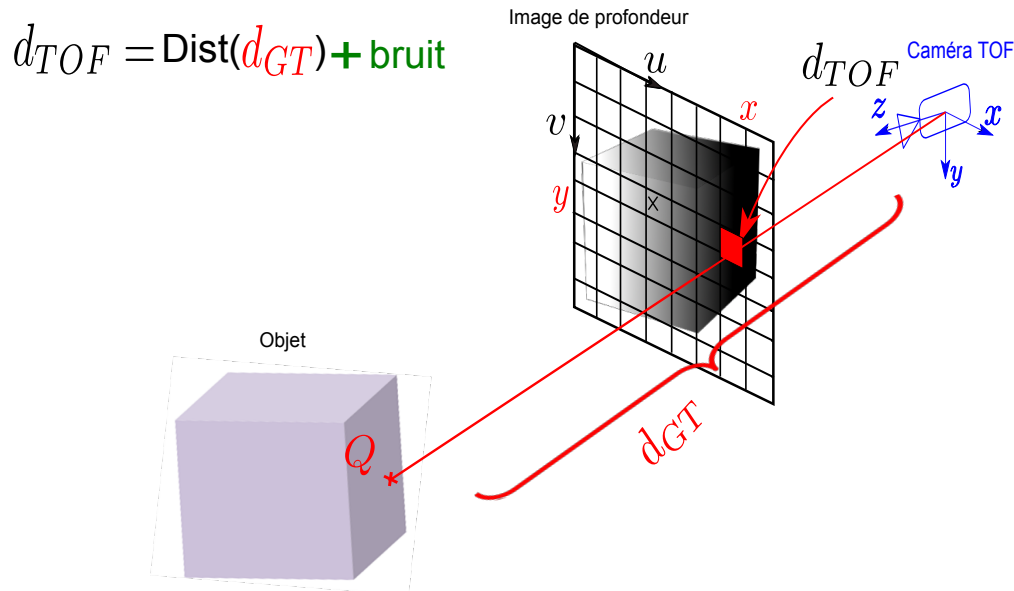


FIGURE 1.8 – La profondeur du point de la surface du cube d_{TOF} issue du capteur Temps-de-Vol correspond à une mesure distordue de la profondeur réelle d_{GT} à laquelle s'ajoute un bruit stochastique.

Facteurs de distorsion. Dans la littérature, de nombreux travaux (*e.g.* ceux présentés dans [Karel *et al.* 2007, Guömundsson *et al.* 2007, Weyer *et al.* 2008]) se sont intéressés à l'étude des différents facteurs de distorsion. La distorsion n'a pas un facteur unique mais est issue d'une multitude de facteurs, chacun d'entre eux ayant une influence plus ou moins importante. Le principal facteur est l'*erreur*

systematique inhérente à la technologie Temps-de-Vol (présentée en détail dans la section 1.2) qui emploie une démodulation par un signal sinusoïdal qui est en pratique un signal sinusoïdal imparfait. Le deuxième facteur est la variation du TI . La figure 1.9 montre un exemple, pour un pixel donné (centre de l'image), de la variation de la distorsion de la mesure de profondeur selon TI . Ceci est dû à la variation des mesures de profondeur en fonction du TI . En effet, pour une même scène, des cartes de profondeur différentes sont obtenues en faisant varier les valeurs du TI . La cause est la variation du nombre d'électrons collectés en fonction du TI (pour un TI plus long, le nombre d'électrons est plus élevé).

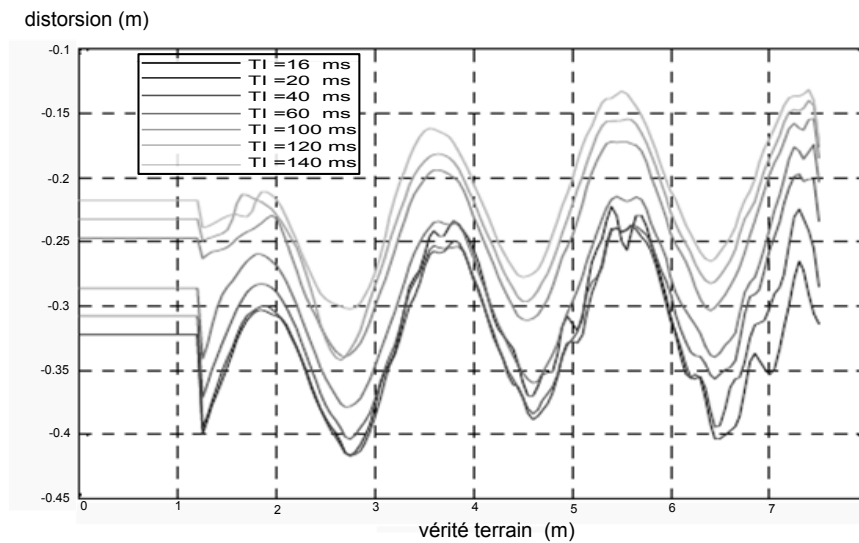


FIGURE 1.9 – Distorsion de la mesure de profondeur calculée pour différents TI . Chaque courbe correspond à un TI donné (*extrait de [Kahlmann et al. 2006]*).

Le troisième facteur est la variation de la réflectivité des objets de la scène. Dans la figure 1.10, la carte de profondeur d'un damier plan d'étalonnage est présentée. Des objets situés à la même distance de la caméra et d'indices de réflectivité différents (des carrés noirs/blancs) n'ont pas la même mesure de profondeur. Cependant, d'après les travaux présentés dans [Kahlmann et al. 2006], la distorsion due à ce facteur est négligeable comparée aux facteurs cités précédemment. Le quatrième facteur est la variation de l'angle d'incidence de l'illumination par rapport à la surface de l'objet. La distorsion due à ce facteur est difficilement détectable, seulement les trois premiers facteurs (*erreur systématique*, la variation du TI et la variation de la réflectivité) ont été pris en compte dans les méthodes d'étalonnage de la littérature (présentées en détail dans le chapitre 2).

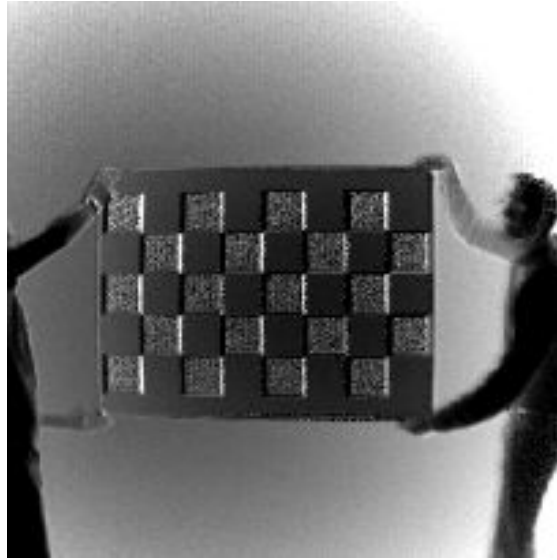


FIGURE 1.10 – Carte de profondeur d'un damier de calibrage. La profondeur mesurée par la caméra varie selon l'indice de réflectivité de l'objet. Des objets (les carrés du damier) situés à la même distance de la caméra n'ont pas les mêmes mesures de profondeur.

Caractéristique de la distorsion. Le modèle exact de la distorsion de la mesure de profondeur n'a pas été établi. En effet, le modèle physique régissant la technologie Temps-de-Vol est très complexe. Dans la littérature, quelques caractéristiques de la distorsion ont été établies de manière empirique. Pour un TI donné, la distorsion est continue et elle varie en fonction de deux paramètres : d (la profondeur mesurée) et (u,v) (la position du pixel dans l'image). Le graphe représentant la distorsion pour différents pixels en fonction de la vérité terrain (les mesures de profondeur dépourvues de toutes distorsions, les méthodes de calcul de la vérité terrain seront présentées en détail dans la section 2.2.4 du chapitre 2) a généralement une forme "sinusoïdale" ou polynomiale illustrées par la figure 1.11. La figure 1.12 (extraite de [Kahlmann *et al.* 2006]) présente un exemple de la variation en fonction de la position du pixel dans l'image. Elle représente la distorsion calculée pour tous les pixels d'une image d'un mur situé à une distance égale à 2,45 m. Ce graphe montre bien que la distorsion liée à la position du pixel dans l'image n'est pas négligeable : elle varie entre -30 cm et +15 cm. Ce phénomène reste valable pour toutes les caméras Temps-de-Vol basées sur la démodulation de type "lock-in pixels". Cependant l'amplitude de la distorsion varie d'une caméra à une autre. La variation de la distorsion en fonction de la position du pixel est

causée par les différentes propriétés matérielles du circuit CMOS⁴ de la caméra et la répartition non uniforme sur toute l'image de l'illumination proche IR.

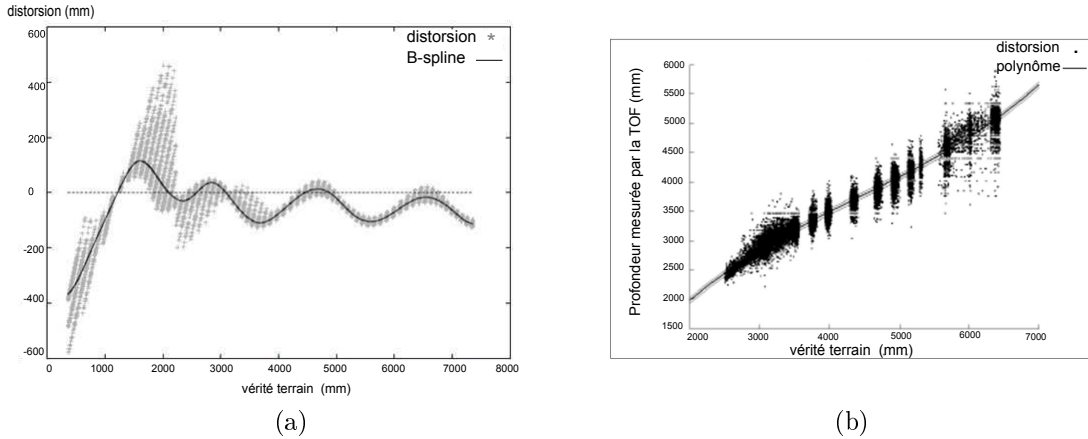


FIGURE 1.11 – Un exemple de graphe (a) représentant la distorsion de la mesure de profondeur d'une caméra Temps-de-Vol en fonction de la vérité terrain (en gris). La courbe qui approxime au mieux ces données oscille autour de zéro (fonction *B-spline* tracée en noir) (*extrait de [Lindner & Kolb 2006]*). (b) représente la mesure de profondeur d'une caméra Temps-de-Vol en fonction de la vérité terrain (en noir). La ligne dans le graphe représente la fonction qui approxime au mieux les points 2.5D issues du capteur Temps-de-Vol (*extrait de [Schiller et al. 2008]*).

1.6 Conclusion

Dans ce chapitre, nous avons, tout d'abord, expliqué le principe de fonctionnement des caméras Temps-de-Vol et plus précisément celles qui sont basées sur la démodulation de type "*lock-in pixels*". Ensuite, nous avons présenté leurs données de sortie ainsi que les paramètres de capture (TI et f_{mod}) à configurer avant de faire des acquisitions. Enfin, nous avons présenté les limites principales de cette technologie. Les deux premières étant la faible résolution et la courte portée. Nous avons vu que ces limites ne peuvent pas être palliées sans modifier le mode de fonctionnement de la caméra Temps-de-Vol (*e.g.* utiliser deux images de profondeur acquises à des f_{mod} différentes, fusionner avec d'autres capteurs de plus haute résolution). En revanche, la troisième limite qui est l'incertitude (due au bruit du

4. C'est un circuit intégré logique CMOS (Complementary Metal Oxyde Semiconductor), réalisé à partir de transistors à effet de champ de type MOS.

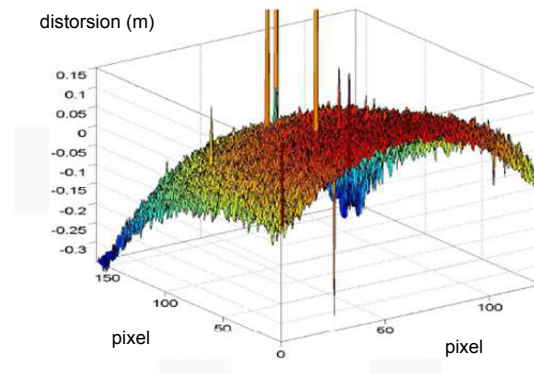


FIGURE 1.12 – Exemple de la distorsion de la mesure de profondeur calculée pour tous les pixels de l'image d'un mur situé à une distance égale à 2,45 m de la caméra (extrait de [Kahlmann et al. 2006]).

capteur) et la distorsion de la mesure de profondeur peut être traitée sans modifier le mode de fonctionnement de la caméra. Nous nous sommes concentrés dans nos travaux sur la troisième limite et le chapitre suivant en présente l'état de l'art.

CHAPITRE 2

Etat de l'art

Ce chapitre a pour but, dans un premier temps, de faire un tour d'horizon des approches existantes pour la modélisation du bruit interne des caméras Temps-de-Vol. Dans un second temps, les méthodes existantes pour l'étalonnage de la mesure de profondeur seront présentées. Les limites liées à ces méthodes seront ensuite discutées.

2.1 Modélisation du bruit des caméras Temps-de-Vol

2.1.1 Méthodes de la littérature

Comme présenté dans la section 1.5.3.1 du chapitre 1, nous pouvons distinguer deux types de bruit : un bruit externe et un bruit interne. Nous nous intéressons dans nos travaux au bruit interne qui ne dépend que du fonctionnement du capteur : ce dernier n'a été que peu étudié dans la littérature. Il introduit une incertitude sur la mesure de profondeur. Cela veut dire que différentes mesures effectuées d'un même point en maintenant la position de la caméra fixe, ne seront pas identiques. La figure 2.1 présente un exemple d'image représentant l'écart-type de la mesure de profondeur d'une vue fronto-parallèle d'un mur blanc situé à une distance approximativement égale à 4 m, calculé pour 100 mesures de profondeur par pixel (pour TI égale à 14 ms). Le bruit est moins élevé au centre du capteur où il y a une meilleure illumination de la scène : l'écart-type varie de 5 mm au centre de l'image et atteint jusqu'à 12 mm sur les bords.

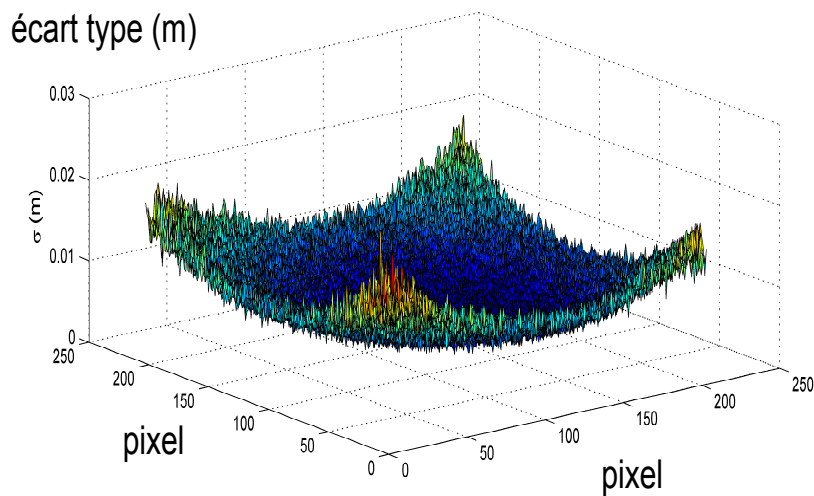


FIGURE 2.1 – L'écart-type σ de la mesure de profondeur pour une image d'un mur blanc situé à une distance approximativement égale à 4 m, calculé pour 100 mesures de profondeur par pixel et un TI égal à 14 ms.

Nous avons trouvé dans la littérature deux travaux sur le bruit affectant les mesures de profondeur des caméras Temps-de-Vol. Le premier est présenté dans [Falie & Buzuloiu 2007] qui expliquent principalement l'origine physique du bruit : la non continuité du courant électrique. Les auteurs présentent aussi une méthode

de modélisation implicite du bruit. En effet, la profondeur d du capteur est calculée à partir des différentes valeurs de la fonction de corrélation notées A_0, A_1, A_2, A_3 (pour plus de détails, voir la section 1.2 du chapitre 1). Une modélisation des $\{A_i\}, i = 1 \dots 4$ est proposée et est donnée par :

$$A_i = g_i(e_i + n_i) \quad (2.1)$$

où g_i représente le gain, e_i le nombre d'électrons et n_i un bruit additionnel introduit par le capteur et indépendant du bruit électronique présenté précédemment. À partir de ce modèle, les écarts-types des $\{A_i\}, i = 1 \dots 4$ sont calculés. Ensuite, ces derniers sont utilisés pour déduire l'écart-type de la mesure de profondeur d à partir des équations (2.1), (1.7) et (1.9). Notons que l'écart-type de la mesure de profondeur d représente le bruit du capteur. Cependant, nous ne pouvons pas obtenir, à partir de la modélisation des A_i , une expression analytique du bruit qui serait facilement utilisable en pratique. Ce modèle reste théorique et n'est pas facilement applicable. Une solution à ce problème est présentée par la suite dans [Edeler *et al.* 2010] où un modèle de bruit qui admet une expression analytique est proposé. Le bruit est modélisé par son écart-type noté σ_d (écart-type de la profondeur d). σ_d est une fonction de l'amplitude a du signal et de l'intensité h et est donnée par :

$$\sigma_d \propto \frac{\sqrt{h}}{a} \quad (2.2)$$

Ce modèle donne une approximation grossière du bruit et non pas sa valeur exacte. De plus, nous allons montrer dans le paragraphe suivant la limite de ce modèle par une simple expérience comparant la variation de σ_d en fonction de la distance avec celle d'un σ_d empirique.

Par ailleurs, notons qu'il existe des travaux sur la modélisation du bruit d'autres capteurs 3D. Par exemple, pour les caméras Kinect, un modèle de bruit a été présenté dans [Nguyen *et al.* 2012]. Ce modèle est déterminé empiriquement en observant que le bruit de la caméra Kinect dépend non seulement de la coordonnée en Z mais aussi de l'angle entre l'axe de vue de la caméra et la surface de l'objet observé (noté θ) (une évaluation du bruit des caméras Kinect est présentée dans l'annexe A). De ce fait, le bruit est défini comme étant composé d'un bruit latéral et d'un bruit axial. Pour chaque type de bruit, un modèle paramétrique basé sur l'écart-type est présenté. Le premier noté σ_L est donné par l'équation (2.3). Le

deuxième noté σ_Z est défini par l'équation (2.4).

$$\sigma_L(\theta) = 0,8 + 0,035 \times \frac{\theta}{\frac{\pi}{2} - \theta} \quad (2.3)$$

$$\sigma_Z(Z, \theta) = 0,0012 + 0,0019(Z - 0,4)^2 + \frac{0,0001}{\sqrt{Z}} \frac{\theta^2}{(\frac{\pi}{2} - \theta)^2} \quad (2.4)$$

2.1.2 Discussion et positionnement

Le premier modèle présenté ne permet pas d'estimer l'incertitude des mesures de profondeur en chaque point (aucune expression analytique n'a été donnée). Quant au deuxième modèle, il permet d'estimer le bruit en chaque point. Cependant, il est basé uniquement sur l'information de l'amplitude. Nous avons effectué une simple expérience pour évaluer ce modèle dans une scène où l'indice de réflectivité est invariant. Pour ce faire nous avons effectué des acquisitions d'un mur blanc en couvrant différentes distances (d allant de 1 m à 7 m). Ensuite, nous calculons l'écart-type σ de la mesure de profondeur calculée pour 100 mesures par pixel. Sur un graphe, nous traçons ces valeurs en fonction de la distance d et sur un autre les valeurs de $\frac{\sqrt{h}}{a}$ (expression du deuxième modèle de bruit) en fonction de la distance d (voir figure 2.2). En comparant les deux graphes, nous observons

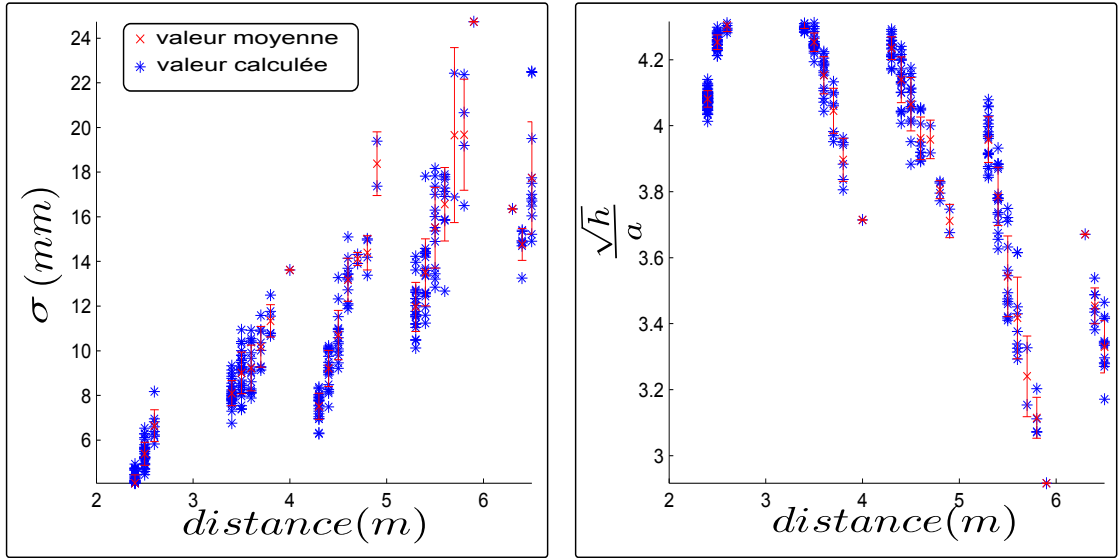


FIGURE 2.2 – Comparaison de la variation en fonction de la profondeur de (a) l'écart-type σ de la mesure de profondeur calculé pour 100 mesures par pixel et (b) la fonction de modélisation du bruit $\frac{\sqrt{h}}{a}$ proposée dans [Edeler *et al.* 2010]. Les données utilisées correspondent à un mur blanc.

que $\frac{\sqrt{h}}{a}$ ne varie pas de la même façon que σ . Cela limite la validité de ce modèle.

Une autre limite générale des deux premiers modèles est que le bruit du capteur n'a pas été caractérisé. L'hypothèse que le bruit suit une distribution normale est posée sans présenter de démonstration. De plus, les facteurs de variation du bruit n'ont pas été étudiés. Quant au troisième modèle de bruit présenté, il est propre aux caméras Kinect. Nous nous sommes inspirés du modèle de bruit du capteur Kinect pour trouver un modèle adéquat aux caméras Temps-de-Vol. Nous proposons, tout d'abord, d'identifier les différents facteurs de variation du bruit des caméras Temps-de-Vol, ensuite, nous présentons un modèle paramétrique dont les paramètres sont les facteurs de variation du bruit.

En conclusion, nous pouvons dire qu'il n'existe pas de travaux qui caractérisent et qui présentent un modèle de bruit précis pour les caméras Temps-de-Vol. Dans ce mémoire, nous proposons, tout d'abord, une caractérisation du bruit interne des caméras Temps-de-Vol : nous démontrons empiriquement que ce bruit suit une distribution normale. Nous proposons de vérifier empiriquement l'hypothèse des travaux de la littérature qui considèrent que le bruit ne dépend que de l'amplitude du signal reçu en intégrant d'autres facteurs tels que la position du pixel dans l'image (u, v) , la profondeur d , le Temps d'Intégration TI et l'indice de réflectivité. Nous proposons deux modèles de bruit qui intègrent ces facteurs. Nous choisissons des modèles de bruit plus précis qui dépendent de plusieurs facteurs, car l'amplitude seule n'est pas suffisante (voir section 3.4).

Pour prendre seulement en compte le bruit interne du capteur et éviter d'introduire le facteur de la réflectivité, dans notre modélisation, les cartes de profondeur utilisées correspondent à des vues d'un mur blanc. Nos travaux sur la caractérisation et modélisation du bruit interne des caméras Temps-de-Vol seront présentés dans le chapitre 3.

2.2 Étalonnage de la distorsion de la mesure de profondeur

Outre le bruit, la mesure de profondeur des caméras Temps-de-Vol est affectée par une distorsion, *i.e.* une déformation affectant la profondeur mesurée. Nous avons vu, dans la section 1.5.3.2 du chapitre 1, que cette distorsion n'est pas une simple constante. Elle varie en fonction de la position du pixel dans l'image et de la profondeur. Les facteurs de la distorsion ont déjà été introduits dans la même section. Le facteur principal est *l'erreur systématique*.

2.2.1 Introduction

La correction de la distorsion de la mesure de profondeur se fait généralement en deux étapes comme le présente la figure 2.3. La première étape, se déroulant hors ligne, est l'étalonnage de la mesure de profondeur. Les données d'entrée sont les cartes de profondeur fournies par la caméra Temps-de-Vol et la vérité terrain correspondante. La vérité terrain représente les cartes de profondeur dépourvues de tout bruit et de toute distorsion. Les méthodes d'étalonnage actuelles reposent toutes sur le même principe qui consiste à modéliser et estimer la fonction de correction de la distorsion (que nous notons FCD) à partir de la déviation entre la vérité terrain et les cartes de profondeur associées. Différentes méthodes d'acquisition de la vérité terrain existent dans la littérature, qui fournissent le même type de données et sont détaillées dans la section 2.2.4. Notons que nous pouvons choisir le modèle de la distorsion indépendamment de la méthode utilisée pour le calcul de la vérité terrain. Nous pouvons dire alors que la FCD est définie par le modèle de distorsion choisi. Une fois le modèle choisi, il ne reste qu'à estimer la FCD. La deuxième étape, se déroulant en ligne, est la correction des cartes de profondeur en temps réel. La correction en ligne est automatique une fois la FCD estimée. La difficulté réside donc dans la première étape : l'étalonnage de la mesure

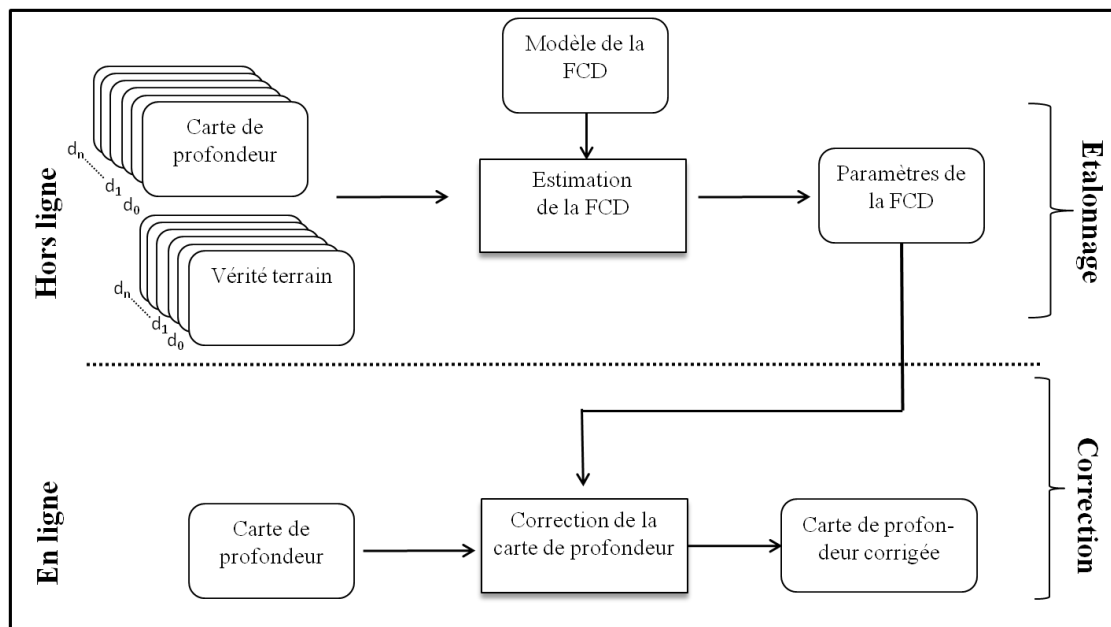


FIGURE 2.3 – Principe général des méthodes de correction de la distorsion de la mesure de profondeur de la littérature.

de profondeur.

Les méthodes d'étalonnage de la mesure de profondeur de la littérature reposent principalement sur une modélisation paramétrique de la FCD. Elles se distinguent par le modèle qu'elles emploient. Nous détaillons dans ce qui suit, tout d'abord, les différents modèles existants de la FCD. Ensuite, nous présentons les méthodes d'estimation de la FCD, et les méthodes d'acquisition de la vérité terrain. Enfin, les limites liées à ces méthodes seront mises en avant.

2.2.2 Modélisation de la FCD

Nous présentons, dans cette section, les différentes méthodes de modélisation de la FCD. Nous nous intéressons tout d'abord aux modèles prenant en compte le facteur principal de la distorsion, *l'erreur systématique*. Ensuite, nous montrons comment ces modèles ont été étendus pour prendre en compte d'autres facteurs de distorsion, *i.e.* la réflectivité et le *TI*.

Comme présenté précédemment dans la section 1.5.3.2 du chapitre 1, la distorsion de la mesure de profondeur due à *l'erreur systématique* varie en fonction de deux paramètres : la profondeur et la position du pixel dans l'image. Les méthodes d'étalonnage de la distorsion de la mesure de profondeur de la littérature peuvent être classées en deux familles selon le nombre de modèles utilisés. La première famille utilise un seul modèle exprimant ainsi sous le même formalisme les deux variations : la FCD calculée est une fonction de (u, v) , position du pixel dans l'image, et de d , la profondeur. Cependant, il y a dans la littérature un seul modèle dans cette famille. Il est très simple, ne permettant pas de bien modéliser les variations complexes de la distorsion. La deuxième famille de méthodes présente une modélisation plus complexe. Par contre, elles utilisent deux modèles. Un premier modélisant l'*"étalonnage global"* et un deuxième additif modélisant l'*"étalonnage par pixel"*. Nous appelons *"étalonnage global"*, l'étalonnage prenant seulement en compte la variation de la distorsion en fonction de la profondeur et *"étalonnage par pixel"* prenant en compte la position du pixel dans l'image.

2.2.2.1 Un seul modèle

La seule méthode de la littérature qui utilise un seul modèle est présenté dans la publication [Schiller *et al.* 2008]. Cette méthode corrige l'erreur principale : *l'erreur systématique*. La FCD proposée dans cette méthode est modélisée par une fonction polynomiale de degré 3 qui dépend de (u, v) (la position du pixel dans l'image) et de $d(u, v)$ (la profondeur mesurée par la caméra associée à (u, v)). Le

polynôme est défini par l'équation suivante :

$$d(u, v, d)_{corr} = a_0 + (1, 0 + a_1)d(u, v) + a_2u + a_3v + a_5d(u, v)^2 + a_6d(u, v)^3. \quad (2.5)$$

où $d(u, v)_{corr}$ représente la profondeur corrigée et a_l $l = 0, \dots, 6$ les paramètres du polynôme de la FCD à estimer.

La limite de ce modèle est la simplicité de la fonction choisie pour représenter les variations de la distorsion de la mesure de profondeur. Cette fonction n'est pas très flexible : elle ne permet pas de modéliser des variations complexes.

2.2.2.2 Deux modèles additifs

Une première méthode dans cette famille d'approches est présentée dans [Kahlmann *et al.* 2006]. Une fonction qui a l'avantage d'être plus flexible que la fonction polynômiale présentée précédemment est utilisée dans cette méthode. Une fonction discrète est utilisée pour modéliser l'"*étalonnage global*". Cependant, cette fonction est estimée en utilisant seulement les corrections calculées pour le pixel central de l'image. Se limiter aux pixels centraux de l'image met en cause la précision de la FCD estimée. L'"*étalonnage par pixel*" est modélisé par une simple matrice, de même taille que l'image, représentant la correction à appliquer à chaque pixel. Cette matrice est notée "*FPN*" ("*Fixed Pattern Noise*") et elle est calculée pour une distance donnée (égale à 2,45 m). C'est la seule méthode de la littérature qui présente un modèle non paramétrique. Cependant, la limite de ce modèle est le manque de précision dû à l'utilisation du pixel central seulement pour l'"*étalonnage global*". De plus, une hypothèse forte est faite sur la variation de la distorsion en fonction de la position du pixel dans l'image. Elle est supposée constante pour toutes les profondeurs. Cependant, cette hypothèse n'a jamais été démontrée.

Une deuxième méthode est introduite dans [Lindner & Kolb 2006] qui propose un modèle plus complexe et plus flexible qu'une fonction discrète pour l'"*étalonnage global*". Une fonction de type *B-spline* (les fonctions *B-spline* sont présentées dans l'Annexe B section B.2.2) est utilisée. L'"*étalonnage global*" est donc modélisé par l'équation suivante :

$$d(u, v, k)_{glob} = d(u, v, k) - b(d), \quad (2.6)$$

où $d(u, v, k)_{glob}$ représente la profondeur corrigée du pixel (u, v) de l'image k par le processus de l'"*étalonnage global*" et $b(d)$ la valeur de la *B-spline* correspondante à la profondeur $d(u, v, k)$ (issue du capteur) définie par :

$$b(d) = \sum_{i=1}^p b_i(d)P_i, \quad (2.7)$$

où P_i représente les points de contrôle de la B -spline. Quant à l'"étalonage par pixel", un modèle plus précis qu'une "FPN" est proposé dans cette méthode. En effet, il n'est pas fait d'hypothèse forte sur la variation de la distorsion. Contrairement, au modèle "FPN", tous les pixels sont pris en considération, et pas seulement ceux du centre de l'image. Pour chaque pixel (u, v) , une droite $l_{u,v}(d)$ est estimée en minimisant l'équation suivante :

$$\sum_k \|d(u, v, k) - d_k\|^2. \quad (2.8)$$

où d_k représente la moyenne de la mesure de profondeur de l'image k et $d(u, v, k)$ représente la profondeur du pixel (u, v) dans l'image k . En prenant en considération les deux processus en même temps l'"étalonage par pixel" et l'"étalonage global", la FCD proposée dans cette méthode est modélisée par l'équation suivante :

$$d(u, v, k)_{corr} = d(u, v, k) - b[(d - l_{u,v}(d))]. \quad (2.9)$$

Le modèle proposé dans cette méthode est plus adéquat que le modèle précédent pour modéliser les variations complexes de la distorsion de la mesure de profondeur.

2.2.2.3 Extension des modèles

Les modèles présentés précédemment corrigent la distorsion due à l'*erreur systématique*, facteur principal de la distorsion. D'autres travaux prenant en compte d'autres facteurs, *i.e.* variation de la réflectivité et TI , sont présentés dans la littérature. Cependant, ces travaux ne proposent pas de nouveaux modèles : seulement des extensions aux modèles existants.

Pour prendre en compte le facteur de la variation de la réflectivité, les travaux de la littérature proposent d'étendre le modèle B -spline. Puisque, la réflectivité est liée à l'intensité du signal réfléchi, les images d'intensité (voir la section 1.2 du chapitre 1 pour plus de détails) sont donc utilisées (les mesures d'intensité sont notées h). Par exemple, dans les travaux présentés dans [Lindner & Kolb 2007], une B -spline qui dépend de l'intensité de chaque pixel, notée $b(h)$, est couplée avec la B -spline $b(d)$ dépendante de la profondeur mesurée ($b(d)$ est donnée par l'équation (2.7)). La fonction B -spline améliorée est décrite dans l'équation ci-dessous :

$$b(d, h) = \sum_{i=1}^p \sum_{j=1}^q b_i(d)b_j(h)P_{ij}, \quad (2.10)$$

où d représente la mesure de profondeur, h la mesure de l'intensité et P_{ij} les points de contrôle de la B -spline. Le modèle d'"étalonage global" amélioré est alors donné

par :

$$d(u, v, k)_{corr} = d(u, v, k) - b(d, h). \quad (2.11)$$

Dans ce modèle, un grand nombre de données de vérité terrain est nécessaire. En effet, ces données doivent couvrir un très large intervalle de profondeurs et d'intensités. Cependant, en pratique, c'est une tâche difficile à mettre en œuvre. Une solution à ce problème est proposée par la suite dans [Lindner *et al.* 2010]. Une correction additive aux cartes de profondeur déjà corrigées par le modèle *B-spline* (l'équation (2.9)) est proposée. Cette correction additive est modélisée par un polynôme de degré 3. Elle est basée sur le fait que des mesures de profondeur proches doivent avoir des mesures d'intensité proches. La figure 2.4 montre un exemple des cartes de profondeur obtenues après cette étape. Elles sont plus lisses et plus homogènes principalement sur les zones de transition.

Pour prendre en compte le facteur de variation du *TI*, des FCDs avec différents *TI* sont modélisées. Par la suite, la FCD pour un *TI* donné est obtenue par simple interpolation *2D*.



FIGURE 2.4 – Damier d'étalonnage avant (image à gauche) et après (image à droite) correction de la distorsion liée à la variation de la réflectivité (*extrait de [Lindner et al. 2010]*).

2.2.3 Estimation de la FCD

Une fois la FCD modélisée, il ne reste qu'à estimer ses paramètres. Dans la littérature, il existe deux approches différentes. La première approche, estime les paramètres de la FCD que nous notons a_l indépendamment des autres paramètres de l'étalonnage standard de la caméra (les paramètres intrinsèques de la caméra que nous notons b_k). L'étalonnage standard est préalablement réalisé. Dans ce cas, l'estimation des a_l est faite selon le modèle de la FCD choisi, *i.e.* polynômial ou de type *B-spline* (des méthodes d'estimation standards sont utilisées). Cependant, la

faible résolution de la caméra Temps-de-Vol est un handicap pour un étalonnage standard précis. Cela peut diminuer la précision de la fonction de correction (la FCD). Pour garantir une meilleure précision, une deuxième approche est proposée. Tous les paramètres (a_l et b_k) sont estimés dans un même processus d'optimisation. Pour ces méthodes, l'image d'intensité est utilisée en plus de l'image de profondeur. En intégrant les deux sources d'information, l'intensité et la profondeur, une meilleure précision d'étalonnage peut être atteinte. Pour expliquer cette approche, nous notons i l'index d'un pixel dans l'image, j le numéro de l'image et k l'index de la caméra ($k \geq 1$: plusieurs caméras peuvent être couplées avec la caméra Temps-de-Vol). Le vecteur \mathbf{p}_{ijk} représente un pixel de l'image ($\mathbf{p}_{ijk} = (u, v)$), $d(i, j, k)_{corr}$ représente la profondeur corrigée et a_l les paramètres de la FCD. Les coordonnées du point 3D correspondant exprimées dans le repère monde sont notées \mathbf{P}_i et peuvent être exprimé, comme expliqué dans l'équation (B.16), par :

$$\mathbf{P}_i = d(i, j, k)_{corr} \frac{\mathbf{R}_{jk}^T \mathbf{K}_k^{-1} \mathbf{p}_{ijk}}{\sqrt{\mathbf{p}_{ijk}^T \mathbf{K}_k^{-T} \mathbf{K}_k^{-1} \mathbf{p}_{ijk}}} + \mathbf{C}_{jk}, \quad (2.12)$$

où \mathbf{C}_{jk} est la position de la caméra, \mathbf{R}_{jk} est la matrice de rotation de la caméra et \mathbf{K}_k est la matrice des paramètres intrinsèques de la caméra (ces notions sont détaillées dans la section B.1.1.1). Le repère monde est choisi de telle sorte qu'il soit attaché au damier d'étalonnage et que les coordonnées Z des points soient égales à 0 (puisque le damier est plan). Ce choix est fait pour simplifier les calculs. L'équation (2.12) exprimée pour la troisième composante de \mathbf{P}_i donne :

$$d(i, j, k)_{corr} = \frac{-\mathbf{C}_{z,jk} \sqrt{\mathbf{p}_{ijk}^T \mathbf{K}_k^{-T} \mathbf{K}_k^{-1} \mathbf{p}_{ijk}}}{\mathbf{R}_{z,jk}^T \mathbf{K}_k^{-1} \mathbf{p}_{ijk}} = f_{ijk}^1(a_l), \quad (2.13)$$

où $\mathbf{C}_{z,jk}$ représente la composante Z du vecteur \mathbf{C}_{jk} , $\mathbf{R}_{z,jk}$ représente la dernière colonne de la matrice de rotation \mathbf{R}_{jk} et f_{ijk}^1 est la FCD à estimer. En supposant que l'intensité d'un point 2D \mathbf{p}_{ijk} est égale à l'intensité de référence du point 3D associé \mathbf{P}_i (du damier d'étalonnage), l'égalité suivante est obtenue :

$$I(\mathbf{p}_{ijk}) = I_{ref}(\mathbf{P}_i) \quad (2.14)$$

où I_{ref} représente l'intensité de référence (l'intensité du damier d'étalonnage) et $I(\mathbf{p}_{ijk})$ est l'image d'intensité. En fusionnant les équations (2.12) et (2.13), l'équation suivante est obtenue :

$$I(\mathbf{p}_{ijk}) = I_{ref} \left(\mathbf{C}_{jk} - \frac{\mathbf{C}_{z,jk} \mathbf{R}_{jk}^T \mathbf{K}_k^{-1} \times \mathbf{p}_{ijk}}{\mathbf{R}_{z,jk}^T \mathbf{K}_k^{-1} \times \mathbf{p}_{ijk}} \right) = f_{ijk}^2(a_l, b_k) \quad (2.15)$$

La fonction $f_{ijk}^1(a_l)$ est remplacée par l'expression du modèle de la FCD choisi. Les paramètres a_l sont estimés en même temps que b_k (paramètres intrinsèques), sous un même processus d'optimisation en se basant sur les équations (2.15) et (2.13), en minimisant les erreurs Δ_1 et Δ_2 données par :

$$\Delta_1 = d(i, j, k)_{corr} - f_{ijk}^1(a_l) \quad (2.16)$$

$$\Delta_2 = I(\mathbf{p}_{ijk}) - f_{ijk}^2(a_l, b_k) \quad (2.17)$$

Cette méthode d'optimisation est utilisée dans [Schiller *et al.* 2008, Lindner *et al.* 2010].

En conclusion, nous pouvons dire que cette deuxième méthode d'estimation simultanée de tous les paramètres est théoriquement meilleure. Cependant, en pratique, il est difficile de pondérer les erreurs d'ordres de grandeur différents : l'erreur millimétrique Δ_1 et l'erreur pixélique Δ_2 . De plus, la mise en œuvre est difficile puisqu'il est préférable d'utiliser des images de types différents pour chaque type de paramètre : un mur uniforme pour l'estimation des a_l et un mur avec motif pour l'estimation des b_l . C'est pour cette raison que nous avons choisi d'estimer séparément ces paramètres.

2.2.4 Acquisition d'une vérité terrain

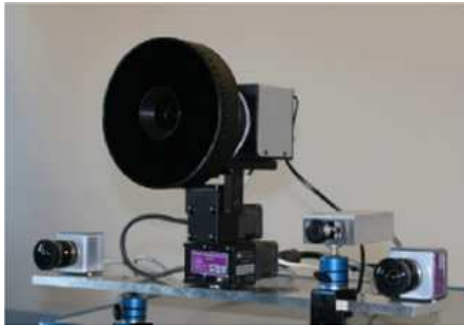
L'acquisition de la vérité terrain est une étape très importante dans les méthodes d'étalonnage existantes. En effet, ces dernières se basent sur le calcul de la déviation entre la vérité terrain et les mesures de profondeur associées (issues du capteur). Il est donc important que les mesures de la vérité terrain soient précises.

Il existe deux méthodes différentes pour calculer la vérité terrain. La première est l'utilisation d'un outil de mesure indépendant : un rail de mesure "*distance measurement track line*" ou une "*total station*"¹. Cependant, ces systèmes sont coûteux. De plus, la mise en œuvre n'est pas explicitée dans les travaux de la littérature. Nous pensons qu'une étape supplémentaire est nécessaire : soit un étalonnage des deux systèmes (l'outil de mesure et le capteur Temps-de-Vol) soit un recalage des images fournies par les deux systèmes. Notons que cette étape peut affecter la précision de la vérité terrain obtenue.

La deuxième méthode de calcul de la vérité terrain ne nécessite pas l'utilisation d'un matériel coûteux. Elle utilise un objet d'étalonnage dont la géométrie doit être parfaitement connue (en général un damier plan d'étalonnage). Avant d'expliquer le principe de cette méthode, notons que cette méthode est basée sur la détection

1. La "*total station*" est un instrument électronique/optique qui permet de mesurer la distance d'un point particulier par rapport l'instrument.

de points ou droites clés de l'objet d'étalonnage. Cependant, la faible résolution de la caméra Temps-de-Vol est un handicap pour une détection précise de ces caractéristiques (plus de détail dans la section 1.5.1). Une ou plusieurs caméras standards de haute résolution sont alors couplées avec la caméra Temps-de-Vol (un exemple est illustré par la figure 2.5). Le capteur Temps-de-Vol est couplé avec trois autres caméras sur la figure 2.5a. Tout d'abord, les paramètres intrinsèques des



(a) extrait de [Schiller et al. 2008]



(b) extrait de [Lindner & Kolb 2007]

FIGURE 2.5 – Exemple de caméras utilisées pour l'étalonnage de la mesure de profondeur : une caméra Temps-de-Vol couplée (a) avec trois caméras de haute résolution et (b) avec une caméra de haute résolution.

caméras ainsi que les paramètres extrinsèques doivent être déterminés. Ensuite, la vérité terrain est obtenue, pour chaque pixel de l'image, par l'estimation de pose. Cette approche ainsi que sa limite, qui est principalement le manque de précision de la vérité terrain obtenue, seront présentées en détails dans la section 4.3.1.1 du chapitre 4. Le manque de précision est causé par l'accumulation des erreurs à chaque étape d'étalonnage (estimation des paramètres intrinsèques de chaque capteur et des paramètres extrinsèques).

2.2.5 Discussion et positionnement

2.2.5.1 Limites des méthodes de la littérature

La première limite des méthodes de la littérature est le choix arbitraire du modèle de la distorsion (polynômial ou de type *B-spline*), alors qu'aucune validation empirique ou théorique de ces modèles n'a été effectuée. La deuxième limite est le manque de précision. En effet, le premier modèle présenté (polynômial) est très simple, il ne permet pas de modéliser avec précision la distorsion.

C'est aussi le cas de la deuxième famille d'approches puisqu'elle est basée sur deux modèles additifs : un modélisant la correction globale et un autre modélisant la correction par pixel ce qui peut diminuer la précision de l'étalonnage. La troisième limite de ces méthodes est leur mise en œuvre compliquée puisqu'elles nécessitent de disposer d'une vérité terrain pour chaque pixel (*e.g.* de 0,75 m à 7,5 m avec un pas de 10 cm pour chaque pixel [Lindner & Kolb 2006] et de 1,25 m à 7,5 m avec un pas de 50 cm [Kahlmann *et al.* 2006]). En effet, comme déjà expliqué dans la section 2.2.4, un système additionnel est nécessaire, *i.e.* un rail de mesure [Lindner & Kolb 2006] et [Kahlmann *et al.* 2006] ou un objet d'étalonnage [Schiller *et al.* 2008] et [Lindner & Kolb 2007]. Le premier système est coûteux. Le second système nécessite un étalonnage stéréo très précis qui est très difficile à réaliser à cause de l'accumulation des erreurs lors de l'étalonnage des caméras utilisées, principalement, celui de la caméra Temps-de-Vol. .

2.2.5.2 Méthodes proposées

Nous proposons deux méthodes d'étalonnage de la mesure de profondeur. La première présente une solution pour les deux premières limites et la deuxième une solution pour la troisième limite :

- La première méthode que nous présentons, contrairement aux méthodes de la littérature, est basée sur un modèle non paramétrique. Ce type de modèle est choisi pour différentes raisons. Premièrement parce qu'il peut être appliqué aux différents types de caméras 3D puisqu'il n'impose aucun modèle particulier. Deuxièmement, pour comparer la précision d'un tel modèle par rapport aux modèles paramétriques de la littérature. Un autre avantage de notre méthode est qu'elle modélise sous un même formalisme à la fois la variation de la distorsion selon la mesure de profondeur et selon la position du pixel dans l'image : il ne s'agit pas de calculer deux corrections additionnelles comme dans la plupart des travaux de la littérature, ce qui permet d'augmenter la précision de l'étalonnage. Nous proposons un modèle plus flexible qu'un simple polynôme.
- La deuxième méthode que nous présentons répond à un autre problème qui est la mise en œuvre difficile des méthodes de la littérature qui nécessitent une mesure de vérité terrain pour chaque pixel. Dans la méthode proposée 4 correspondances de points entre la vérité terrain et les images du capteur Temps-de-Vol sont suffisantes. Notre méthode est composée de deux étapes. La première étape étant une correction de la non planarité qui ne nécessite pas de disposer d'une vérité terrain, seules des cartes de profondeur correspondantes à différentes vues d'un mur blanc sont nécessaires. La deuxième

étape est une correction affine qui nécessite de disposer de 4 correspondances de points entre la vérité terrain et les images du capteur Temps-de-Vol. Par conséquent, l'avantage de la méthode proposée réside dans sa facilité de mise en œuvre et la possibilité de faire l'étalonnage en temps réel.

Chacune des méthodes que nous proposons présente une solution à un problème différent. La première est basée entièrement sur des mesures de vérité terrain (plusieurs mesures sont utilisées), tandis que la deuxième est basée sur la contrainte de planarité et très peu de correspondances de points entre la vérité terrain et les images du capteur sont utilisées. La première peut être plus précise (si la vérité terrain est de grande précision) que la deuxième qui a l'avantage d'être plus pratique à mettre en œuvre. Elles seront donc testées sur un même jeu de données pour comparer leur précision.

Les méthodes que nous proposons corrigent les distorsions de la mesure de profondeur liées à *l'erreur systématique*. En effet, c'est le facteur qui engendre les distorsions les plus grandes et qui doit être corrigé en priorité. Pour cette raison, les cartes de profondeur utilisées pour l'étalonnage correspondent à des vues d'un mur blanc, pour éviter les distorsions liées aux autres facteurs de distorsion. Néanmoins, les méthodes proposées peuvent être étendues pour prendre en compte d'autres facteurs de distorsion. Par exemple, une simple interpolation $2D$ (voir section 2.2.2) permet de prendre en considération le facteur de variation du TI .

2.3 Conclusion

Dans ce chapitre, nous avons tout d'abord présenté l'état de l'art de la modélisation du bruit des capteurs Temps-de-Vol. Nous avons vu qu'il n'existe pas, à notre connaissance, de travaux qui caractérisent et qui présentent un modèle de bruit précis. Cela constitue l'objet de nos travaux présentés dans le chapitre suivant. Ensuite, nous avons présenté les méthodes de l'étalonnage de la mesure de profondeur des caméras Temps-de-Vol de la littérature. Elles présentent trois limites : le choix arbitraire du modèle d'étalonnage, le manque de précision et la mise en œuvre délicate. Les méthodes que nous proposons, présentent des solutions à ces limites et seront présentées dans les chapitres suivants.

CHAPITRE 3

Modélisation du bruit

Une des limite des caméras Temps-de-Vol est l'incertitude des mesures de profondeur due au bruit du capteur. Cela veut dire que différentes mesures d'un même point effectuées en maintenant la position de la caméra fixe ne seront pas identiques. Il est important de connaître pour chaque mesure de profondeur l'incertitude associée afin de pouvoir quantifier la qualité de la mesure. Dans ce chapitre, nous commençons par caractériser le bruit des caméras Temps-de-Vol et déterminer ses facteurs de variation. Cette étude est réalisée, à notre connaissance, pour la première fois. Les travaux sur le bruit des caméras Temps-de-Vol présentés dans la littérature considèrent que celui-ci ne dépend que de l'amplitude du signal reçu. Dans ce chapitre, nous proposons de vérifier cette hypothèse en intégrant d'autres paramètres, tels que la position dans l'image, la profondeur, et le Temps d'Intégration TI pour modéliser le bruit. Nous proposons donc deux modèles de bruit qui dépendent des facteurs de sa variation. Le premier est une fonction de la position du pixel dans l'image et de l'amplitude. Le deuxième est plus complexe : les facteurs de variation du bruit sont dissociés. Le modèle est une fonction de la position du pixel dans l'image, de la profondeur et du TI. Une comparaison de la précision des deux modèles sera présentée. Les travaux décrits dans ce chapitre ont donné lieu à la publication [Belhedi et al. 2012a].

3.1 Introduction

Comme cela a été présenté dans la section 2.1, le bruit des caméras Temps-de-Vol est un sujet peu étudié dans la littérature et il est généralement considéré dans les travaux qu'il suit une distribution normale et qu'il varie de manière proportionnelle à $\frac{1}{a}$ (a étant l'amplitude du signal reçu). Cependant, il n'existe pas de travaux qui proposent de caractériser le bruit dû au capteur : la normalité du bruit n'a jamais été démontré. De même, il n'existe pas de travaux comparant le modèle $\frac{1}{a}$ avec des modèles de variation du bruit intégrant d'autres paramètres. Nous avons effectué une simple expérience pour vérifier si l'amplitude est suffisante pour modéliser le bruit. Il s'agit d'évaluer un modèle linéaire en fonction de $\frac{1}{a}$. Pour ce faire nous avons effectué des acquisitions d'un mur blanc en couvrant différentes distances (d allant de 1 m à 7 m). L'écart-type (σ empirique) de la mesure de profondeur pour 100 mesures par pixel est ensuite calculé. Les paramètres du modèle linéaire sont alors estimés à partir de ces données. Nous observons ensuite la réponse de ce modèle en fonction de σ empirique pour différents points. Pour un modèle idéal, la réponse doit correspondre à une droite de pente 1. Cependant, la figure 3.1 montre que pour des grandes valeurs de σ , les points s'écartent de la

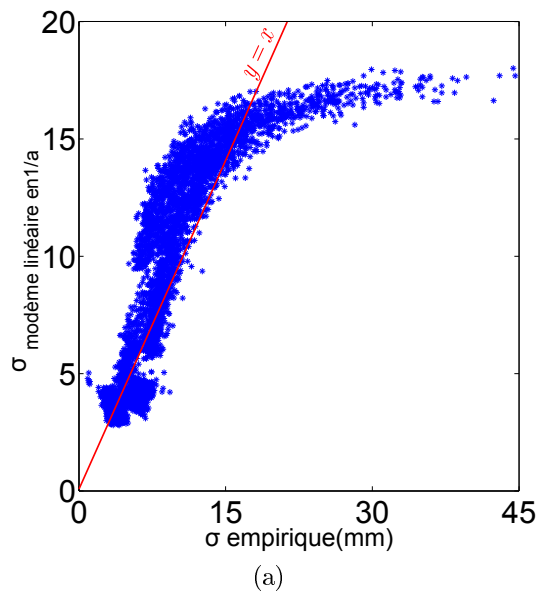


FIGURE 3.1 – Réponse du modèle linéaire en fonction de $\frac{1}{a}$. σ prédit par ce modèle est tracé en fonction de σ empirique pour différents points. Les données correspondent à un mur blanc. Pour des grandes valeurs de σ , les points s'écartent de la droite $y = x$.

droite. Cela veut dire que l'amplitude seule n'est pas suffisante pour modéliser le bruit.

Il serait donc intéressant de déterminer les facteurs de variation du bruit et de proposer un modèle qui en dépend, comme le modèle proposé pour les caméras Kinect (équation (2.3) et (2.4)). L'avantage d'un tel modèle est qu'il est plus flexible et peut être facilement amélioré pour prendre en compte d'autres facteurs.

Nous commençons alors par démontrer empiriquement que le bruit du capteur suit une distribution normale. Nous proposons aussi de modéliser la variation du bruit en considérant d'autres paramètres que la seule amplitude, comme la position du pixel dans l'image, la profondeur, le Temps d'Intégration (TI) et la réflectivité des objets. Deux modèles continus du bruit sont proposés. Le premier dépend de la position du pixel dans l'image et de l'amplitude car cette dernière modélise les autres facteurs de variation du bruit cités ci-dessus (profondeur, TI et réflectivité). La fonction utilisée dans ce premier modèle est une Spline lisse, connue sous le nom de Spline en plaque mince TPS ("*Thin-Plate Spline*"), qui est une fonction adéquate pour modéliser des variations complexes mais lisses. Le deuxième modèle est plus complexe : les facteurs de variation du bruit sont dissociés. Il dépend de la position du pixel dans l'image, de la profondeur et du TI . Le facteur de la réflectivité n'est pas pris en compte dans le deuxième modèle. Cependant, il peut être facilement amélioré pour prendre en compte ce nouveau facteur. Cela constitue une perspective de nos travaux. Le deuxième modèle donne alors le bruit pour chaque point de l'espace 2.5D et pour chaque TI . Pour éviter de modéliser le bruit pour plusieurs valeurs de TI , ce qui n'est pas pratique, nous proposons un modèle permettant de passer d'un TI à un autre. De ce fait, notre deuxième modèle est composé de deux transformations. La première transformation donne l'écart-type du bruit pour chaque pixel et chaque profondeur pour un TI donné (que nous notons TI_1). Elle utilise aussi la fonction de type TPS 3D. La deuxième transformation est une transformation linéaire entre la distribution du bruit à TI_1 et les autres TIs . Enfin, nous comparons nos deux modèles et présentons une simple application du modèle de bruit. L'application est le filtrage des images de profondeur qui sera comparée au filtre bilatéral. Notre premier modèle est utilisé car les images contiennent des objets ayant des indices de réflectivité différents.

Notons que l'application principale de notre modèle de bruit est l'estimation de l'incertitude des mesures 3D basées sur des données de la caméra Temps-de-Vol. Un exemple de mesure est le calcul des dimensions d'un objet à partir de son image de profondeur. Cet exemple d'application sera présenté dans le chapitre 6 et se base sur notre deuxième modèle de bruit car la réflectivité des mesures utilisées est constante. L'incertitude de la mesure est calculée par propagation de l'erreur à travers les différentes transformations permettant de passer des points 2D à leur

distance 3D.

3.2 Caractérisation du bruit

Notre caractérisation du bruit est basée sur l'étude de la variation des mesures de profondeur de différentes images de profondeur prises dans les mêmes conditions. Des images d'un mur blanc (pour éviter le bruit dû à d'autres facteurs comme la réflectivité) sont acquises pour des distances allant de 0,5 m à 7,5 m et pour différents TIs . A chaque distance, 100 images de profondeur sont acquises. Ensuite, une scène avec des objets de différentes réflectivité est acquise pour montrer la variation du bruit en fonction de la réflectivité. Nous commençons par caractériser le bruit pour un TI fixé, ensuite nous nous concentrons sur la dépendance du bruit aux TIs .

3.2.1 Caractérisation du bruit pour un TI fixé

Normalité du bruit. Pour caractériser le bruit, nous commençons par tester la normalité de la distribution du bruit. Il existe principalement deux familles d'approches de tests de normalité. La première est la famille des approches empiriques et graphiques et la deuxième est la famille des approches probabilistes. Pour plus de certitude, nous avons choisi d'utiliser deux tests de la première famille qui sont les diagrammes de probabilité normale "*normal probability plot*" et les histogrammes de la distribution et un test de la deuxième famille d'approches qui est le test de Lilliefors. Avant de présenter ces tests et les résultats obtenus, notons que les données d'entrée utilisées pour ces tests sont des points des images de profondeur acquises avec un TI égale à 14 ms.

Pour le premier test qui est les diagrammes de probabilité normale, les données sont tracées en fonction d'une distribution théorique normale de telle sorte que les 100 observations associées a un point de l'image pour une distance donnée doivent approximativement former une ligne droite (plus de détails sur ce type de diagrammes sont présentés dans [Chambers *et al.* 1983]). L'éloignement des données par rapport à cette ligne droite indique que la distribution de ces données s'éloigne significativement de la distribution normale. La figure 3.2 montre un exemple correspondant à un point de l'image : nous observons que les 100 observations suivent approximativement une ligne droite. Cela prouve que le bruit suit bien une distribution normale.

Pour réaliser le deuxième test de normalité graphique, l'ensemble des points 2.5D des données d'entrée est considéré. Les histogrammes associés à quelques

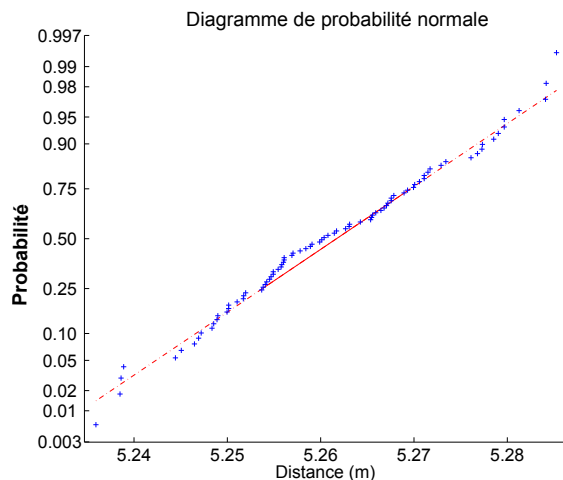


FIGURE 3.2 – Diagramme de probabilité normale correspondant à un point 2.5D des données d’entrée. Les mesures (tracées en bleu) suivent approximativement la ligne droite de la distribution normale théorique (tracée en rouge).

points montrant la distribution de ces mesures de profondeur sont tracés sur la figure 3.3. Ces histogrammes, comme nous pouvons le voir graphiquement, suivent des distributions normales.

Après la vérification empirique et graphique, un test de normalité probabiliste (plus robuste comparé aux approches empiriques et graphiques) est maintenant utilisé. Le test de Lilliefors (qui est une adaptation du test de Kolmogorov-Smirnov¹) permet de tester l’hypothèse nulle sur des mesures de la loi normale. Le test retourne la valeur logique $h = 1$ si il rejette l’hypothèse nulle, sinon il retourne $h = 0$. Ce test est effectué pour chaque point 2.5D des données d’entrée, ensuite la valeur médiane et la valeur moyenne de h pour toutes les données sont calculées. La valeur médiane est égale à 0 et la valeur moyenne est aussi très proche de zéro (égale à 0,15). Ces valeurs confirment que le bruit des caméras Temps-de-Vol suit bien la distribution normale.

Variation du bruit. Après la démonstration de la normalité de la distribution du bruit, nous nous intéressons maintenant à la variation du bruit en fonction de deux facteurs : la position du pixel dans l’image et la profondeur. Comme on a validé que le bruit affectant un point 2.5D suit une distribution normale centrée en 0, nous proposons de représenter la quantité du bruit par l’écart-type σ de la distribution. σ est calculé pour chaque point 2.5D des images de profondeur

1. Le test de Kolmogorov-Smirnov est utilisé pour déterminer si un échantillon suit bien une loi donnée.

d'entrée. Les valeurs de σ varient approximativement de 0 jusqu'à 30 mm (la valeur moyenne de σ est égale à 12,5 mm) comme le montre la figure 3.4.

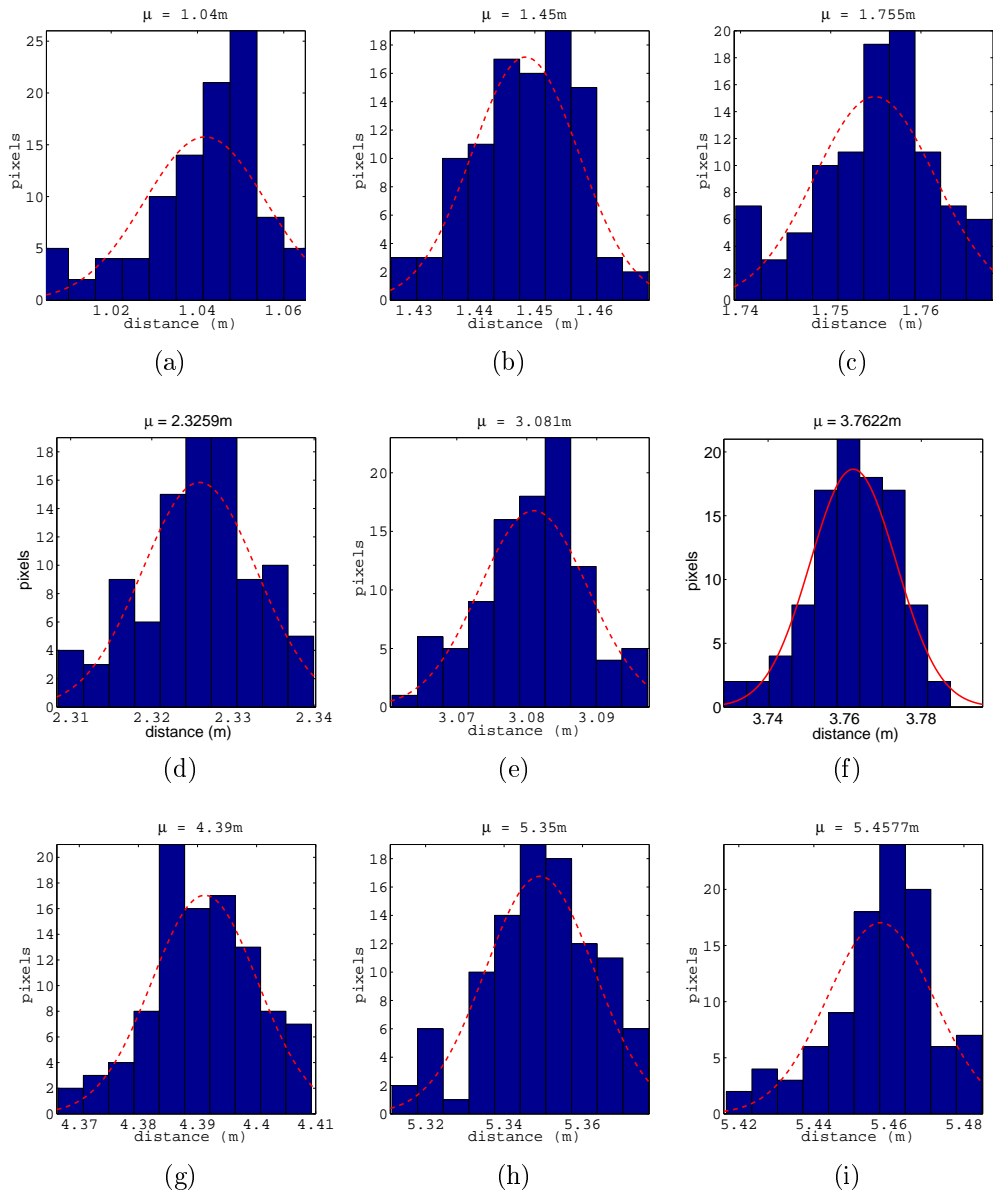


FIGURE 3.3 – Histogrammes de la distribution de quelques points 2.5D (d'un mur blanc). Chaque graphe représente la distribution de 100 mesures de profondeur. Graphiquement, ces histogrammes correspondent à des distributions normales, comme le confirme le test de Kolmogorov-Smirnov.

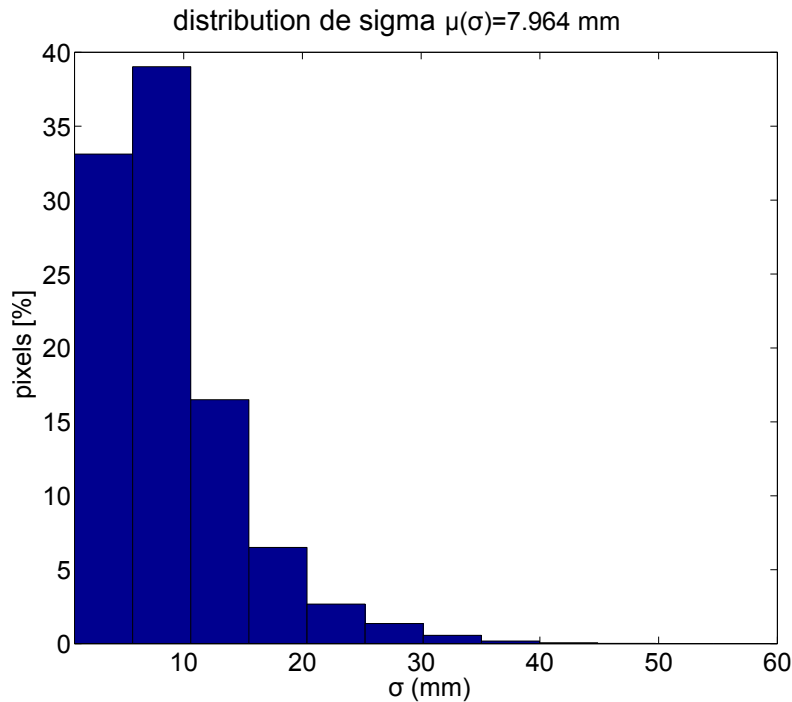


FIGURE 3.4 – Distribution de l'écart-type à chaque pixel des images de profondeur (d'un mur blanc). σ est calculé pour chaque pixel à partir de 100 mesures de profondeur. σ varie approximativement de 0 à 30 mm avec une valeur moyenne égale à 12,5 mm.

La figure 3.6e présente les valeurs de σ pour chaque point 2.5D à une distance approximative $d=3$ m. Comme nous pouvons le constater, σ est plus élevé au niveau des bords de l'image (de 11 à 22 mm) qu'au centre de l'image (6 to 7 mm). Une plus grande précision est observée au centre de l'image où l'illumination de la scène est plus élevée. Le même phénomène est observé pour toutes les autres images de profondeur d'entrée. Nous observons maintenant la variation de l'écart-type en fonction de la profondeur (illustrée par la figure 3.7a). La valeur de σ à chaque pixel des images de profondeur est calculée et est tracée en fonction des valeurs de profondeur.

Nous montrons brièvement que le bruit varie également en fonction de la réflectivité. Nous utilisons une image d'un damier d'étalonnage avec des carrés noirs et blancs (voir figure 3.5). L'écart type calculé pour chaque pixel à partir de chaque mesure de profondeur montre que des objets à la même distance de la caméra n'ont pas le même écart type. Nous observons que le bruit est plus élevé sur les carrés noirs.

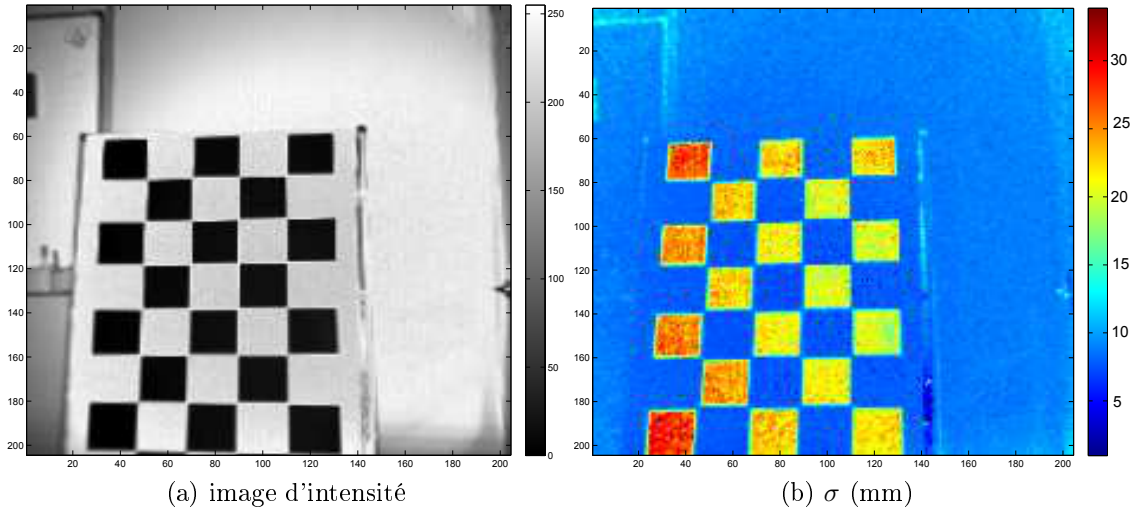


FIGURE 3.5 – Exemple montrant la dépendance du bruit de la réflectivité. (a) Image d'intensité d'un damier d'étalonnage noir et blanc. (b) L'écart type σ calculé pour chaque pixel à partir de 100 mesures de profondeur. Les carrés du damier de différentes couleurs situés à la même distance de la caméra n'ont pas le même σ (bruit plus élevé sur les carrés noirs).

3.2.2 Dépendance du bruit au TI

Nous nous intéressons maintenant à la variation de l'écart-type du bruit σ en fonction des différents TIs . Il est connu que le TI a une influence sur la précision des mesures de profondeur. Un TI plus long donne des images de profondeur de plus grande précision (bruit moins élevé). Cependant, la probabilité du flou de bougé est plus élevée. Dans cette section, nous utilisons des acquisitions effectuées à différents TIs : 1 ms, 2 ms, 3,5 ms, 7 ms et 14 ms. Nous commençons tout d'abord par comparer la variation de σ en fonction de la position pixel dans l'image pour différents TIs . La figure 3.6 présente l'écart-type calculé correspondant à une image de profondeur à une distance approximative de $d = 3$ m. Nous constatons que pour les différents TIs les figures ont la même forme. σ augmente du centre de l'image vers les bords de l'image. Cependant, l'amplitude de σ varie en fonction du TI . Par exemple, $\sigma_{TI=14}$ (σ calculé pour un $TI = 14$ ms) varie de 6 mm jusqu'à 12 mm (figure 3.6e), comparé à $\sigma_{TI=1}$ qui varie de 30 mm à 100 mm (figure 3.6a).

Nous comparons maintenant la variation de σ en fonction de la profondeur, pour les différents TIs . Cela est présenté par les graphes de la figure 3.7. Les cinq premiers graphes de cette figure présentent un aperçu de la variation de σ en fonction de la profondeur pour une plage de distance allant de 2 m jusqu'à 6,5 m.

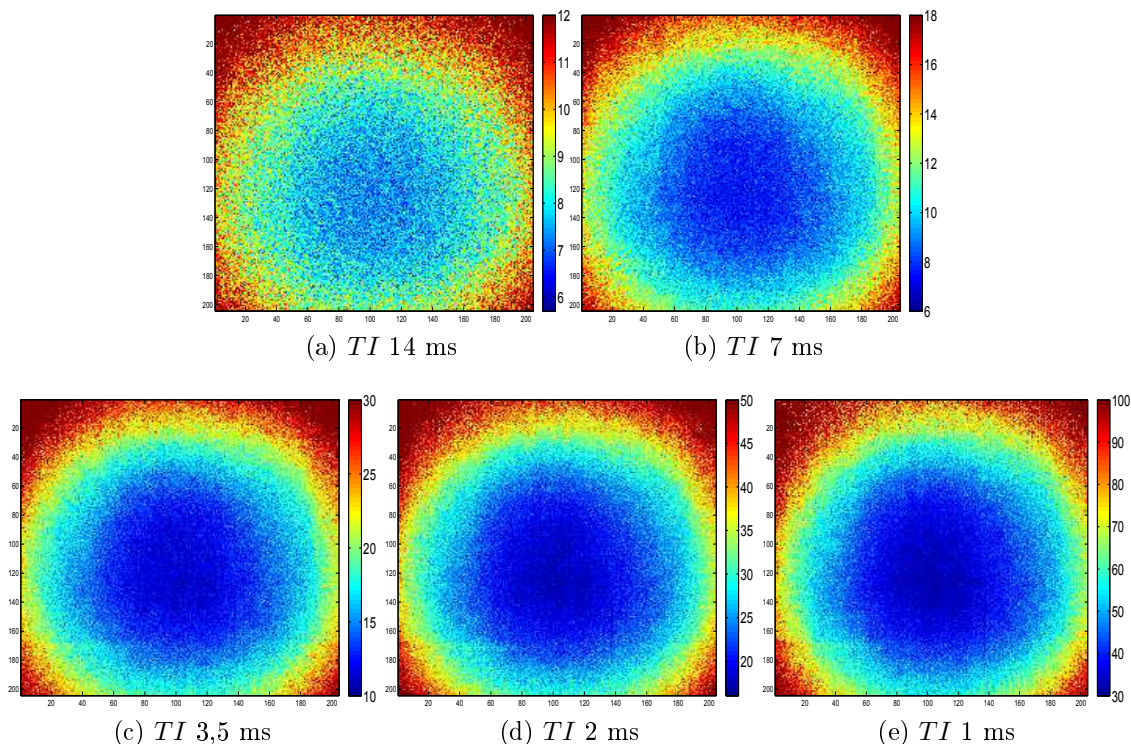


FIGURE 3.6 – Écart-type σ (mm) (d'une image de profondeur à une distance approximative de 3 m) pour différents TIs : (a) 14 ms, (b) 7 ms, (c) 3,5 ms, (d) 2 ms et (e) 1 ms. σ augmente de façon significative du centre de l'image, où l'illumination des objets observées est plus élevée, vers les bords de l'image.

Nous observons que σ augmente en fonction de la profondeur. Cela est vérifié pour les différentes valeurs de TIs . Nous nous limitons à cet intervalle de profondeur pour éviter de représenter les mesures de profondeurs aux bords de cet intervalle qui sont invalides. Le sixième graphe 3.7f superpose les différentes distributions du bruit pour les différentes valeurs de TIs . Ils montrent que la valeur de σ est inversement proportionnelle au TI .

L'écart-type du bruit donne une information sur la précision des mesures de profondeur. Cette information est essentielle pour différentes applications, puisqu'elle mesure la proximité des résultats aux données réelles. Cependant les valeurs des écarts-types calculées à partir des données d'entrée ne sont pas suffisantes. En effet, elles ne couvrent pas l'ensemble de l'espace 2.5D. Il est important d'avoir le bruit du capteur Temps-de-Vol pour tout l'espace 2.5D. Nous proposons alors un modèle continu du bruit.

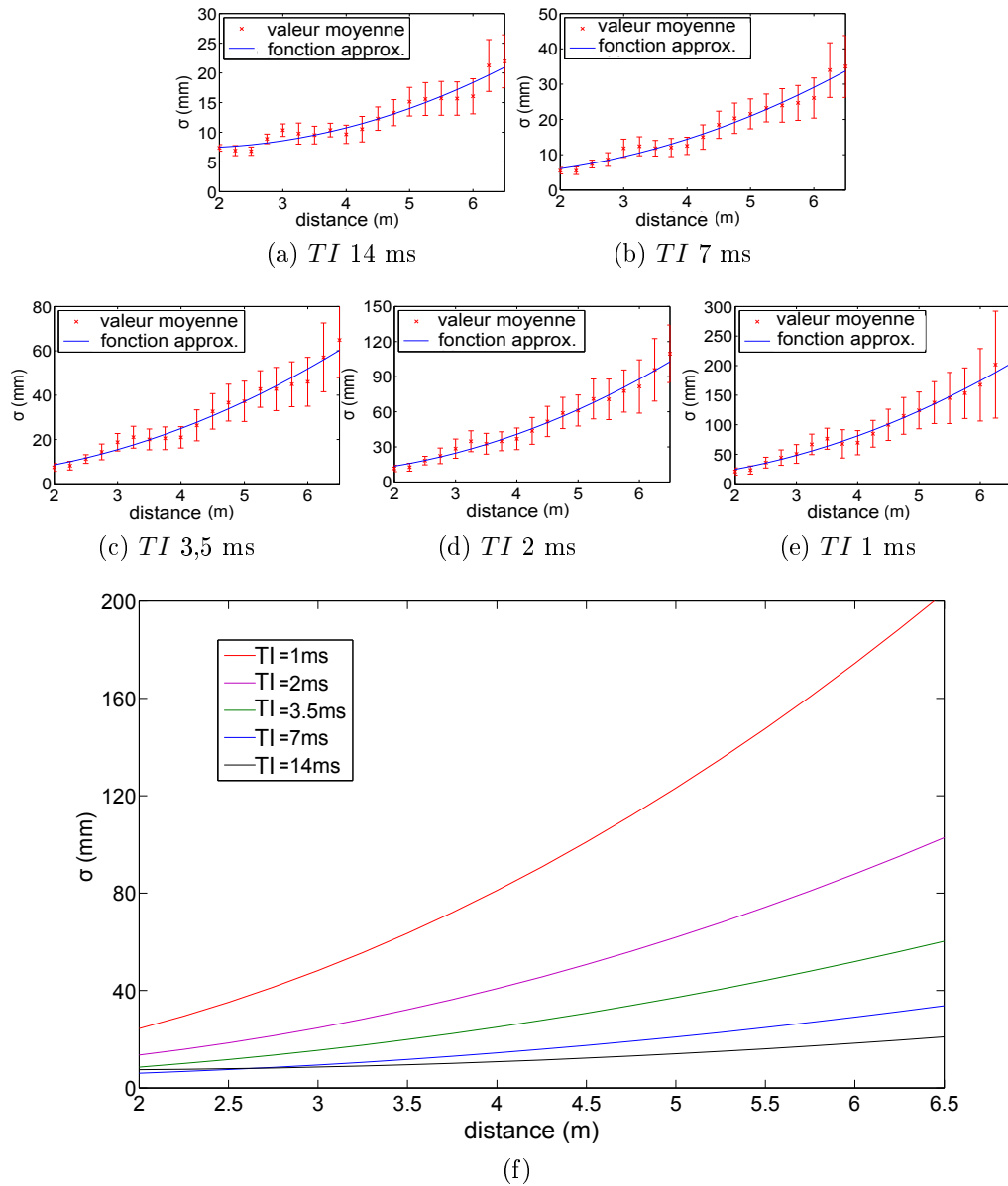


FIGURE 3.7 – Écart-type σ (mm) correspondant à des images de profondeur d'un mur blanc tracé en fonction de la profondeur, pour différents TI s : (a) 14 ms, (b) 7 ms, (c) 3,5 ms, (d) 2 ms et (e) 1 ms. La valeur moyenne de σ est tracée en points rouges et la fonction approx. en ligne bleue. La valeur moyenne de σ varie approximativement de (a) 6 mm jusqu'à 22 mm, (b) 6 mm jusqu'à 35 mm, (c) 7 mm jusqu'à 64 mm, (d) 11 mm jusqu'à 110 mm et (e) 20 mm jusqu'à 200 mm respectivement. Les fonctions approx. les données, pour les différents TI s, sont tracées dans un même graphe dans (f).

3.3 Modélisation continue du bruit

Comme cela a été détaillé dans la section 3.2.1, le bruit des caméras Temps-de-Vol suit une distribution normale, il peut donc être représenté par l'écart-type de sa distribution. Notre objectif est de fournir un modèle continu du bruit c'est-à-dire sa valeur pour chaque point 2.5D. Dans les travaux de la littérature le bruit est considéré dépendant uniquement de l'amplitude du signal et qu'ils sont linéairement liés. Ce type de modèle a l'avantage d'être simple à estimer et peut être utilisé sans a priori sur la scène (*e.g.* réflectivité des objets, TI , profondeur, position du pixel dans l'image) car l'amplitude est une donnée fournie avec toutes les caméras Temps-de-Vol dont la valeur dépend de l'ensemble de ces paramètres. Cependant une expérience simple permet de voir que ce modèle semble imparfait, en particulier pour les grandes valeurs de σ (figure 3.1).

Dans cette section nous souhaitons vérifier si un modèle plus complexe (que le modèle linéaire) intégrant d'autres paramètres (*e.g.* position du pixel dans l'image) permettrait d'améliorer la modélisation du bruit. Dans un premier temps, nous proposons un modèle intégrant à la fois l'amplitude et la position du pixel dans l'image noté modèle position-amplitude. Dans un second temps, nous proposons un modèle où nous considérons à la fois la position du pixel dans l'image, la profondeur, le TI et la réflectivité (ce sont les paramètres sous-jacent à l'amplitude) noté modèle position-profondeur- TI . Ces deux modèles du bruit seront estimés à partir d'un ensemble de valeurs discrètes des écart-types. Ces valeurs correspondent aux écart-types calculés à partir des acquisitions d'un mur blanc en couvrant différentes distances (d allant de 1 m à 7 m).

3.3.1 Modèle position-amplitude

Un premier modèle de bruit qui prend en considération en plus de l'amplitude a , la variation en fonction de la position du pixel dans l'image (u, v) (modèle position-amplitude) est proposé. L'amplitude étant fonction de la profondeur, du TI et de la réflectivité. Pour mieux modéliser la variation du bruit en fonction de ces deux paramètres, nous proposons d'utiliser un modèle non linéaire de type plaque mince de dimension 3 "*3D Thin-Plate Spline*" (TPS 3D) (détaillée dans le paragraphe suivant). Cette fonction est choisie car elle réduit l'empreinte mémoire et permet d'avoir un modèle continu pour l'ensemble des points $\mathbf{r} = (u \ v \ a)^T$ sans nécessiter d'avoir un grand nombre d'observations.

Modèle de type TPS-3D. La fonction TPS 3D est contrôlée par l centres 3D \mathbf{c}_k basée sur l'association de cibles α_k aux l centres 3D \mathbf{c}_k et est paramétrée par un

vecteur de $l + 4$ coefficients $\mathbf{h}^\top = (\mathbf{w}^\top \ \mathbf{a}^\top)$ (pour plus de détail voir la section B.2.3 de l'Annexe B). Soit un point noté $\mathbf{r} = (u \ v \ a)^\top$ et $\boldsymbol{\ell}_\mathbf{r}^\top$ le vecteur défini par :

$$\boldsymbol{\ell}_\mathbf{r}^\top = ((d(\mathbf{r}, \mathbf{c}_1)) \ \cdots \ (d(\mathbf{r}, \mathbf{c}_l)) \ \mathbf{r}^\top \ 1), \quad (3.1)$$

où d est la distance euclidienne. La fonction TPS 3D à un point \mathbf{r} est donnée par :

$$\begin{aligned} \omega(\mathbf{r}, \mathbf{h}) &= \boldsymbol{\ell}_\mathbf{r}^\top \mathbf{h} \\ &= \left(\sum_{k=1}^l \mathbf{w}_k d(\mathbf{r}, \mathbf{c}_k) \right) + \mathbf{a}^\top \check{\mathbf{r}}, \end{aligned} \quad (3.2)$$

où $\check{\mathbf{r}} = (u \ v \ a \ 1)^\top$ est la représentation en coordonnées homogènes.

Nous utilisons la fonction de type TPS 3D pour modéliser l'écart-type du bruit qui est de distribution normale. Cette fonction paramétrique est choisie pour plusieurs raisons. Premièrement, elle introduit la propriété de continuité et de régularité du bruit : nous supposons que pour des pixels voisins et pour des valeurs amplitudes proches, le bruit est assez semblable. La fonction de type TPS 3D vérifie cette hypothèse car elle introduit une transformation minimisant un terme de lissage qui est l'énergie de courbure. Deuxièmement, elle nous permet de limiter l'empreinte mémoire. En effet, seuls les $l + 4$ paramètres et les l centres doivent être enregistrés. Le modèle proposé est basé sur les valeurs des écarts-types et est défini par la fonction g :

$$\begin{aligned} g : \Theta &\rightarrow \mathbb{R} \\ \begin{pmatrix} u \\ v \\ a \end{pmatrix} &\rightarrow \sigma, \end{aligned} \quad (3.3)$$

où $\Theta \subset \mathbb{R}^3$, $\Theta = [u_{min}; u_{max}] \times [v_{min}; v_{max}] \times [a_{min}; a_{max}]$ et σ est un scalaire qui représente la valeur de l'écart-type. $g(\mathbf{r}) \stackrel{\text{def}}{=} \boldsymbol{\ell}_\mathbf{r}^\top \mathbf{h}$ et g appartient à $L^2(\Theta)$ ². Les l centres de la fonction TPS correspondent à quelques points des données d'entrée régulièrement positionnés de tel sorte à couvrir tout l'espace Θ . l est choisi égal $l = 6^3$ (6 points dans chacune des directions u , v et a).

Estimation des paramètres de la fonction TPS-3D. Les $l + 4$ coefficients de la fonction TPS dans le vecteur \mathbf{h} sont calculés à partir des valeurs cibles de σ_k (l'écart-type). En appliquant l'équation de la fonction TPS (3.2) aux centres c_r

2. L'espace de Hilbert des fonctions intégrables deux fois.

avec les valeurs cibles σ_r nous obtenons :

$$\left(\sum_{k=1}^l \mathbf{w}_k d(\mathbf{c}_r, \mathbf{c}_k) \right) + \mathbf{a}^\top \check{\mathbf{c}}_r = \sigma_r, \quad (3.4)$$

où $\check{\mathbf{c}}_r$ représente les coordonnées homogènes. En combinant toutes les équations obtenues pour tous les l centres avec les conditions aux bords (voir section B.2.3) en une seule équation matricielle nous obtenons :

$$\underbrace{\begin{pmatrix} \mathbf{K}_\lambda & \mathbf{P} \\ \mathbf{P}^\top & 0 \end{pmatrix}}_{\mathbf{D}} \underbrace{\begin{pmatrix} \mathbf{w} \\ \mathbf{a} \end{pmatrix}}_{\mathbf{h}} = \begin{pmatrix} \sigma \\ \mathbf{0} \end{pmatrix} \text{ avec } \mathbf{K}_\lambda = \begin{cases} \lambda & r = k \\ d(\mathbf{c}_r, \mathbf{c}_k) & r \neq k. \end{cases} \quad (3.5)$$

où λ est un scalaire que nous fixons à une valeur très faible *e.g.* $\lambda = 10^{-4}$, ce qui permet d'assurer que \mathbf{K}_λ et ainsi \mathbf{D} soient bien conditionnées et \mathbf{P} est la matrice dont la $k^{\text{ème}}$ ligne est donnée par $(\mathbf{c}_k \ 1)$. Cela correspond à un système linéaire, les coefficients de la fonction TPS dans \mathbf{h} sont ainsi facilement estimés par la technique de la pseudo-inverse (plus de détails dans la section B.1.2.2).

3.3.2 Modèle position-profondeur- TI

Notre objectif est de fournir un modèle qui dépend à la fois de la position du pixel dans l'image (u, v) , de la profondeur d , du TI et de la réflectivité de l'objet. Cependant, dans un premier temps, nous considérons la réflectivité constante. L'intégration de la réflectivité dans notre modèle constitue une perspective de nos travaux. Tout d'abord, un modèle pour un TI donné est proposé. Ensuite nous proposons de l'étendre pour l'ensemble des TI s. Il modélise la variation du bruit en fonction de la position du pixel dans l'image et de la profondeur pour un TI fixé (choisi égal à 14 ms dans notre modèle). La fonction de type TPS 3D est également choisie pour modéliser cette variation pour les mêmes raisons citées dans la section précédente.

Modèle de type TPS-3D. Le modèle proposé donne la valeur de l'écart-type pour chaque point 2.5D noté $\mathbf{q} = (u \ v \ d)^\top$ et est défini par la fonction f :

$$f : \Omega \rightarrow \mathbb{R} \\ \begin{pmatrix} u \\ v \\ d \end{pmatrix} \rightarrow \sigma, \quad (3.6)$$

où $\Omega \subset \mathbb{R}^3$, $\Omega = [u_{min}; u_{max}] \times [v_{min}; v_{max}] \times [d_{min}; d_{max}]$ et σ est un scalaire qui représente la valeur de l'écart-type. $f(\mathbf{q}) \stackrel{\text{def}}{=} \ell_{\mathbf{q}}^T \mathbf{h}$ et f appartient à $L^2(\Omega)$. Cette fonction est modélisée par la fonction de type TPS 3D. Les paramètres de la fonction TPS sont estimés de la même façon que dans le premier modèle : il suffit de remplacer l'amplitude a par la distance d dans la définition de l'ensemble de l centres ($l = 6^3$).

Modélisation de la variation en fonction du TI . Dans la section précédente, un modèle de bruit pour un TI donné (égal à 14 ms dans notre cas) est proposé. En comparant les distributions du bruit des différents TIs , nous observons qu'elles sont linéairement liées. Il n'est donc pas nécessaire d'ajouter une dimension à la fonction TPS pour prendre en compte le TI , il suffit d'estimer la transformation linéaire. Pour cela, nous commençons par analyser la variation de $\sigma_{TI=i}$ ($i \in \{7 \text{ ms}, 3,5 \text{ ms}, 2 \text{ ms}, 1 \text{ ms}\}$) en fonction de $\sigma_{TI=14}$. La figure 3.8 présente les valeurs de $\sigma_{TI=i}$ pour différents points 2.5D tracées en fonction de $\sigma_{TI=14}$. Comme le montre cette figure, $\sigma_{TI=i}$ croît linéairement avec un coefficient directeur $a = \frac{14}{i}$ et une

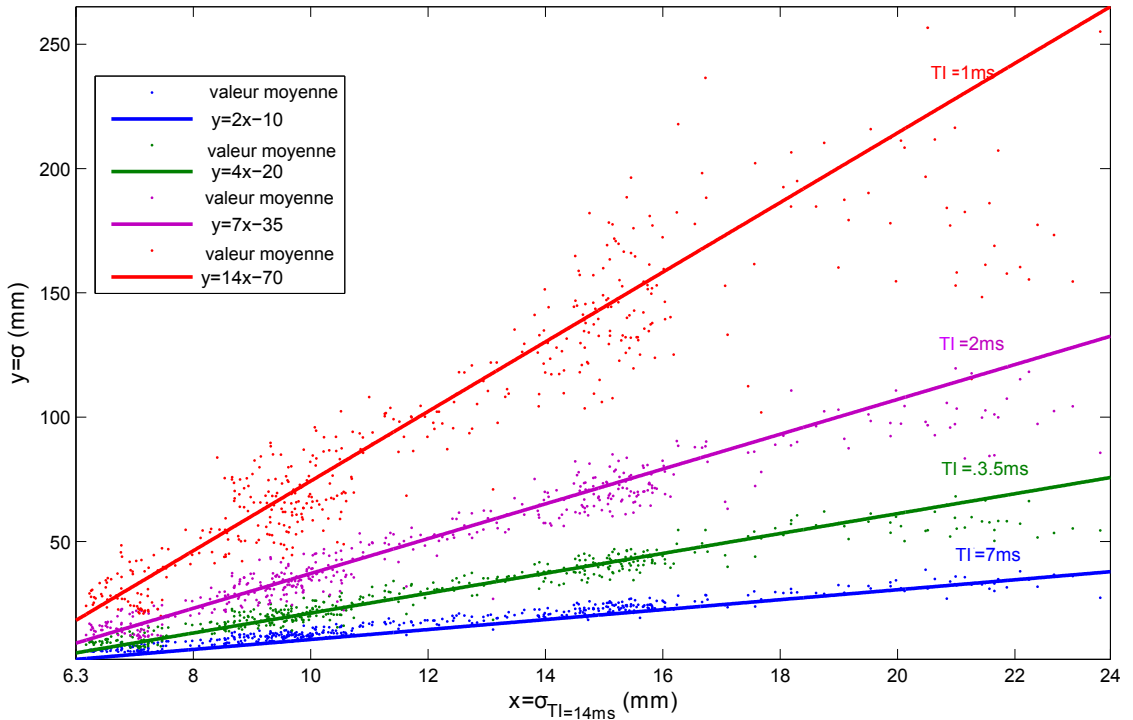


FIGURE 3.8 – Approximation de $\sigma_{TI=7}$, $\sigma_{TI=3,5}$, $\sigma_{TI=2}$ et $\sigma_{TI=1}$, respectivement, en fonction de $\sigma_{TI=14}$ entre 2 m et 6,5 m

ordonnée à l'origine $b = -5a$. Par exemple, la relation entre $\sigma_{TI=2}$ et $\sigma_{TI=14}$ est exprimée par :

$$\begin{aligned}\sigma_{TI=2} &= a \times \sigma_{TI=14} + b \\ &= \frac{14}{2} \times \sigma_{TI=14} - 5 \times \frac{14}{2}.\end{aligned}\tag{3.7}$$

3.4 Évaluation des deux modèles

Une évaluation de la qualité de nos deux modèles est proposée. Le critère d'évaluation est la capacité à fournir l'estimation de l'incertitude pour chacun des points de l'espace 2.5D à partir des images issues du capteur. Les deux modèles dont :

- Notre modèle position-amplitude de type TPS utilisant l'amplitude a et défini par la fonction g (équation 3.3) noté $\text{TPS}_{(u,v,a)}$;
- Notre modèle position-profondeur- TI de type TPS défini par la fonction f (équation 3.6) noté $\text{TPS}_{(u,v,d)}$.

Afin de pouvoir estimer la qualité du modèle, il est nécessaire de disposer d'acquisitions avec la vérité terrain associée (*σ empirique*). Cette vérité terrain est obtenue en capturant 100 images d'une scène statique (un mur blanc) en couvrant différentes distances (d allant de 1 m à 7 m). Un mur blanc est utilisé pour garder la réflectivité constante, car la variation de la réflectivité n'est pas encore intégré dans notre deuxième modèle. L'écart type des 100 observations est calculé pour chaque point 2.5D. Un jeu de données est utilisé pour l'estimation du modèle et un autre pour son évaluation. Deux courbes sont tracées sur la figure 3.9. Chacune présente σ prédit par un modèle de bruit tracé en fonction de σ empirique. Pour un modèle idéal, la réponse doit correspondre à une droite de pente 1. Pour les deux courbes, les points sont bien répartis autour de la droite idéale. En les comparant, nous constatons que notre deuxième modèle $\text{TPS}_{(u,v,d)}$ offre une meilleur précision que le premier $\text{TPS}_{(u,v,a)}$: les points sont moins dispersés le long de la droite et il y a moins de valeurs aberrantes. Notons que ces observations sont valables pour une réflectivité constante. Cependant, pour une scène où la réflectivité varie, le modèle $\text{TPS}_{(u,v,a)}$ peut être une bonne option.

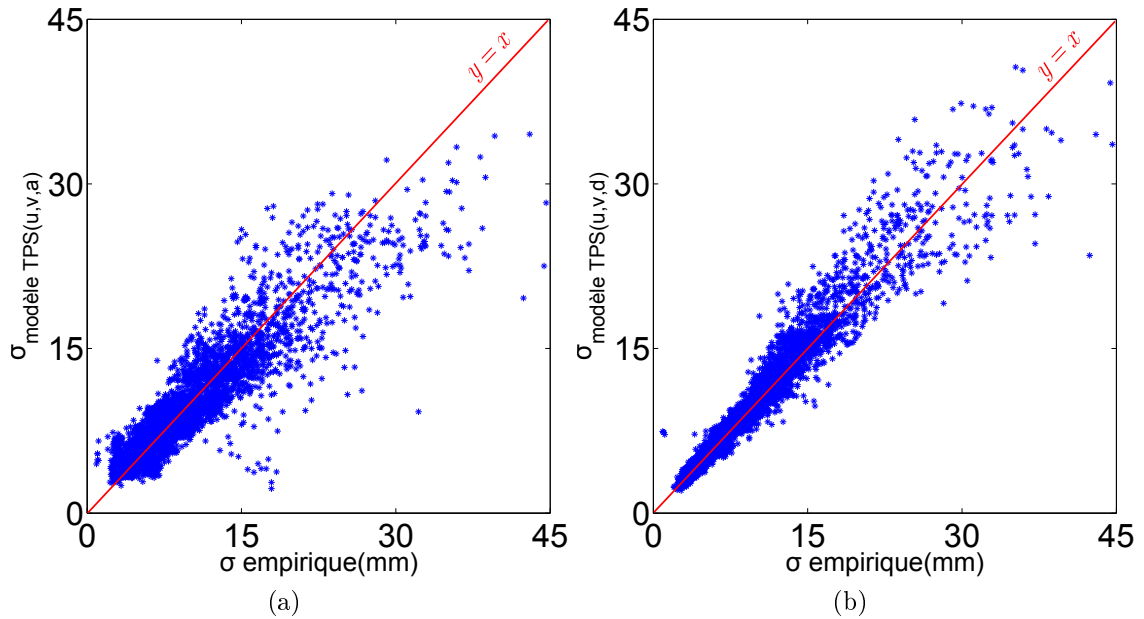


FIGURE 3.9 – Comparaison de (a) notre premier modèle position-amplitude noté $\text{TPS}_{(u,v,a)}$ avec (b) notre deuxième modèle position-profondeur- TI noté $\text{TPS}_{(u,v,d)}$. Les données correspondent à un mur blanc. Pour un modèle idéal, les points doivent être sur la droite $y = x$.

3.5 Exemple d'application : filtrage des images de profondeur

Afin d'illustrer l'intérêt de disposer d'un modèle de bruit précis du capteur Temps-de-Vol, une application de filtrage des images de profondeur basée sur notre modèle est proposée. L'image utilisée dans cette application contient des objets ayant des indices de réflectivité différents. Notre modèle position-amplitude (section 3.3.1) est alors utilisé (car la réflectivité n'est pas encore intégré dans notre deuxième modèle position-profondeur).

3.5.1 Principe

Notre approche de filtrage des images de profondeur des caméras Temps-de-Vol est basée seulement sur l'écart-type du bruit σ obtenu par notre premier modèle position-amplitude modélisé par la fonction TPS de dimension 3 ($\text{TPS}_{(u,v,a)}$). Cela nous permet d'avoir un filtrage plus robuste comparé au filtre bilatéral qui est très sensible aux valeurs de σ_{bf} (σ_{bf} présente la valeur de σ utilisé pour

le filtrage bilatéral). L'image de profondeur est filtrée en utilisant des fenêtres de 3×3 afin de générer des images de profondeur plus lisses. Cela est détaillé dans l'Algorithme 1. Chaque point 2.5D $\mathbf{q} = (u \ v \ d)^\top$ de l'image de profondeur est remplacé par une pondération de la moyenne des valeurs de profondeur du voisinage $\hat{\mathbf{q}} = (u \ v \ \hat{d})^\top$. Ce poids dépend de l'écart-type σ calculé à partir de la fonction g , $\sigma = g(\mathbf{r})$ avec $\mathbf{r} = (u \ v \ a)^\top$ (voir section 3.3.1) et aussi de la distance euclidienne notée Δ_d . Δ_d est la distance entre le point 2.5D considéré et les points du voisinage. Cela donne effectivement un poids plus élevé au point 2.5D les plus précis (ceux qui ont des σ plus faibles).

```

1: for chaque point 2.5D  $\mathbf{q}$  de l'image de profondeur do
2:    $\sigma \leftarrow g(\mathbf{r})$ 
3:   for chaque  $\mathbf{q}_i$  du voisinage  $3 \times 3$  du point  $\mathbf{q}$  do
4:      $\Delta_d = \|d - d_i\|_2$ 
5:      $w_i \leftarrow \exp\left(\frac{-\Delta_d^2}{2\sigma^2}\right)$ 
6:   end for
7:    $\hat{d} \leftarrow \frac{\sum_i d_i w_i}{\sum_i w_i}$ 
8: end for

```

Algorithme 1 : Filtrage des images de profondeur

3.5.2 Résultats

Pour illustrer l'efficacité de la méthode de filtrage proposée, les données d'expérimentation sont réalisées pour un TI différent de 14 ms. La valeur de TI choisie pour l'expérimentation est égale à 0,9 ms, valeur qui donne des images de profondeur plus bruitées. Les résultats sont illustrés par la figure 3.10. Une image de profondeur avec des objets placés à différentes distances est utilisée. Une comparaison des résultats du filtre proposé avec le filtre bilatéral ($\sigma_{bf} = 3$) est présentée dans cette figure. La méthode de filtrage proposée préserve mieux la discontinuité de profondeur comparé au filtre bilatéral. Avec une valeur de $\sigma_{bf} = 3$, les contours des objets disparaissent en utilisant le filtre bilatéral. Cela est plus visible en effectuant une extraction de contour après le filtrage. Pour l'algorithme d'extraction de contours l'opérateur de Sobel est utilisé. La limite principale du filtre bilatéral est sa sensibilité à la variation de σ_{bf} . Cela est illustré par la comparaison des résultats obtenus après extraction des contours en faisant varier la valeur de σ_{bf} . Une valeur élevée de σ_{bf} réduit mieux le bruit, cependant elle ne préserve pas les contours des objets. Inversement, une faible valeur de σ_{bf} préserve mieux les contours, mais réduit moins bien le bruit. En conclusion, nous pouvons dire que notre méthode donne de bons résultats et qu'elle ne nécessite pas d'ajuster la valeur de σ pour

chaque type de scène.

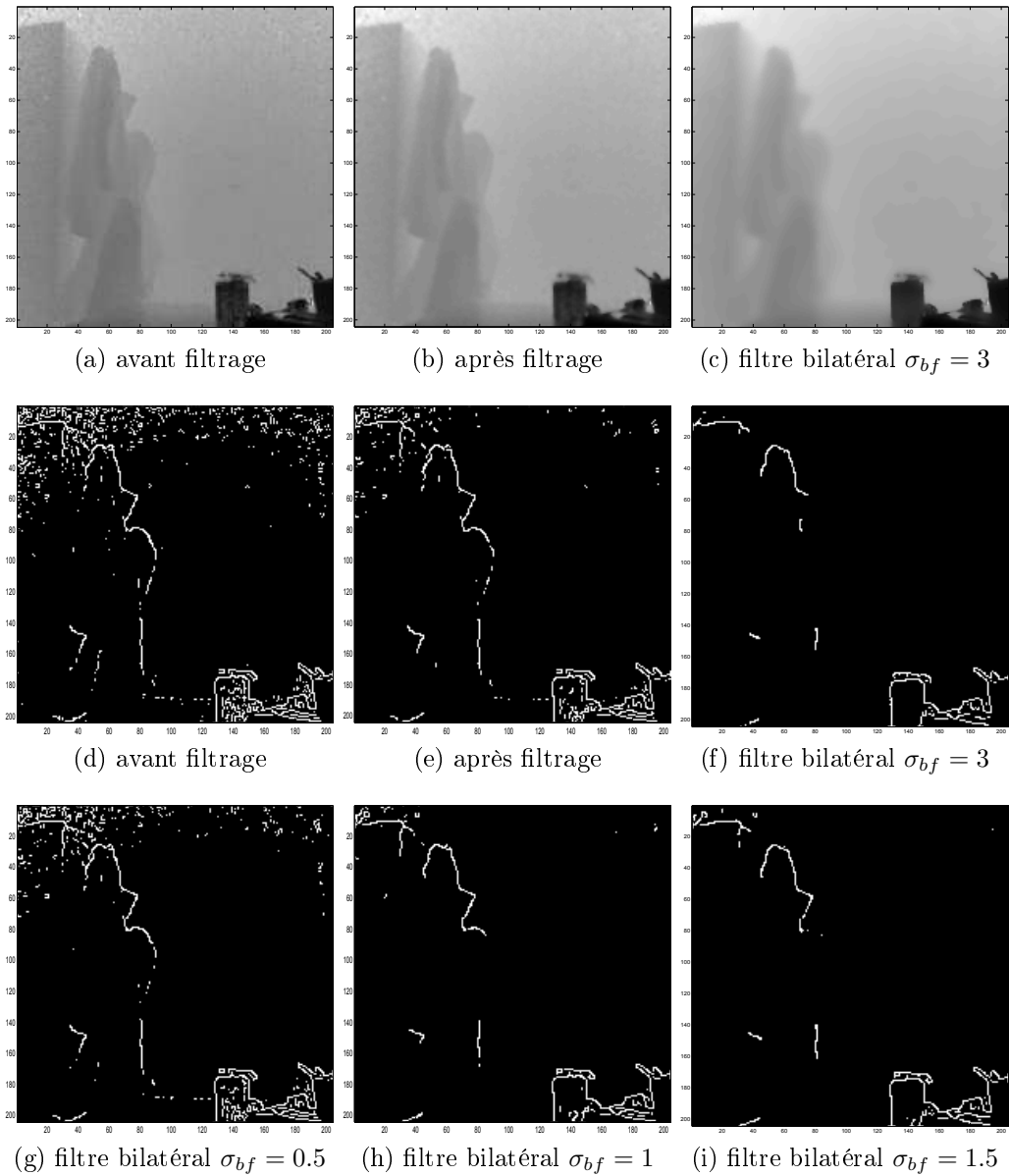


FIGURE 3.10 – Comparaison du filtre proposé avec le filtre bilatéral sur des données réelles. Les résultats sont mieux visibles avec l'extraction des contours effectuée après le filtrage. Le filtrage bilatéral est effectué en faisant varier les valeurs de σ_{bf} : (g) 0,5 mm, (h) 1 mm, (i) 1,5 mm et (f) 3 mm.

3.6 Conclusion

Dans ce chapitre, nous avons montré que le bruit des caméra Temps-de-Vol suit une distribution normale. Nous proposons de modéliser la variation du bruit en considérant d'autres paramètres que la seule amplitude en intégrant d'autres paramètres, tels que la position du pixel dans l'image (u, v) , la profondeur d , le temps d'intégration TI et la réflectivité. Deux modèles ont été proposés. Le premier est une fonction de (u, v, a) (a est l'amplitude). La fonction utilisée pour modéliser ces variations est une spline de type plaque mince de dimension 3 (TPS 3D). Le deuxième est plus complexe, c'est une fonction de (u, v, d, TI) . Il est composé d'une fonction TPS 3D et une transformation linéaire. La réflectivité n'est pas encore introduite dans notre deuxième modèle, cela constitue une perspective de nos travaux. En les comparant, nous constatons que le deuxième modèle offre une meilleure précision que le premier. Une méthode de filtrage des images de profondeur basée sur notre premier modèle est présentée et est testée avec succès sur des données réelles. Notre deuxième modèle de bruit est testé dans le chapitre 6 où nous présentons une application d'estimation de l'incertitude des mesures 3D basées sur des données de la caméra Temps-de-Vol.

CHAPITRE 4

Étalonnage de la distorsion à partir de mesures 2.5D

Dans ce chapitre, nous présentons notre première méthode d'étalonnage de la mesure de profondeur. Elle est basée principalement sur le calcul de la déviation entre les mesures 2.5D réalisées par le capteur Temps-de-Vol et la vérité terrain. Cette méthode présente une solution aux deux premières limites des méthodes de l'état de l'art présentées dans le chapitre 2. Nous pensons que les modèles de la littérature ne garantissent pas la précision requise pour certaines applications. Ceci est dû à l'utilisation de modèles qui ne sont pas assez flexibles ou l'utilisation de deux corrections additives. Nous proposons une méthode basée sur une approche variationnelle. Elle permet d'obtenir une correction plus précise, comparée à d'autres méthodes de la littérature. Elle utilise un modèle plus flexible prenant en compte les variations selon les trois composantes (u, v, d) sous un même formalisme. Dans un premier temps, la méthode proposée est détaillée. Ensuite, son évaluation expérimentale ainsi qu'une comparaison avec les méthodes de l'état de l'art sont présentées. Les travaux décrits dans ce chapitre ont donné lieu à la publication [Belhedi et al. 2012c].

4.1 Introduction

Avant de présenter notre méthode, nous rappelons que nous nous concentrons, dans nos travaux, sur l'étalonnage de la distorsion due au facteur principal : *l'erreur systématique*. Comme il a été montré dans la section 2.2 du chapitre 2, plusieurs modèles paramétriques d'étalonnage de la mesure de profondeur prenant en compte ce facteur ont été proposés dans la littérature. Cependant, il n'existe pas de modèle d'étalonnage physique pour les caméras Temps-de-Vol, ce qui explique la variété des modèles proposés. En l'absence d'une modélisation théorique, les modèles paramétriques de l'état de l'art ont été proposés en se basant sur les caractéristiques de la FCD (Fonction de Correction de la Distorsion) définies empiriquement telles que :

- Variation en fonction de la profondeur d ;
- Variation en fonction de la position du pixel (u, v) dans l'image ;
- Continuité.

La limite des méthodes proposées est le manque de précision dû au choix d'un modèle peu flexible comme le modèle polynômial ou à l'utilisation de deux modèles additifs. Un modèle qui prend en compte les variations selon (u, v) et un autre les variations selon (d) (les variations des trois composantes (u, v, d) n'étant alors pas modélisées sous un même formalisme).

Dans la méthode proposée, nous revisitons la formulation du problème de l'étalonnage. En l'absence de connaissance du modèle exact de la FCD, il est préférable de ne pas utiliser une formulation trop restrictive du problème dans laquelle la FCD est modélisée par une fonction paramétrique prédéfinie. Il est préférable de passer par une formulation flexible. Nous proposons un modèle d'étalonnage basé sur une formulation variationnelle. La FCD est alors estimée à partir de plusieurs jeux de données, par optimisation numérique d'une fonctionnelle. Le coût est la déviation entre les mesures 2.5D (u, v, d) de la vérité terrain et celles issues du capteur. La fonctionnelle est constituée d'un terme d'attache aux données et d'un terme de régularisation. Le premier rapproche la FCD à estimer du jeu de données. Notons qu'aucune fonction particulière n'est imposée à la FCD et que cette dernière permet d'intégrer dans un même formalisme l'ensemble des facteurs de la distorsion (c'est-à-dire les trois composantes (u, v, d)). Le terme de régularisation pénalise la dérivée de la FCD. Il impose de façon simple des hypothèses de régularité sur la FCD. Le modèle estimé vérifie les propriétés de variation continue de la FCD. Nous proposons ainsi une solution permettant d'obtenir une correction plus précise, comparée à d'autres méthodes de la littérature, sans imposer une fonction paramétrique particulière.

Une telle approche de modélisation présente plusieurs avantages. Premièrement, elle n'est pas propre aux caméras Temps-de-Vol : elle peut être appliquée à toutes les caméras 3D. Deuxièmement, elle vérifie les caractéristiques citées ci-dessus (continuité et variation en fonction de (u, v, d)). Troisièmement, elle permet de modéliser sous un même formalisme la distorsion selon d et celle selon la position du pixel dans l'image (u, v) ce qui augmente la précision de la FCD.

Une comparaison avec les deux principales méthodes paramétriques de la littérature, présentée dans la section 4.3.4, valide la précision de notre méthode.

4.2 Présentation de la méthode

Dans cette section, notre méthode basée sur une approche variationnelle est présentée en détail.

4.2.1 Modèle de la FCD

Contrairement aux méthodes de la littérature, nous ne proposons pas une fonction paramétrique pour modéliser la FCD. Notre modèle consiste à estimer une fonction g qui à chaque point 2.5D \mathbf{q} associe la correction correspondante c permettant de compenser la distorsion. Dans le cas continu, cette fonction de correction g peut être modélisée par :

$$g \in L_2^{-1}(\mathbb{G}), g : \mathbb{G} \rightarrow \mathbb{R}, g(\mathbf{q}) = c, \quad (4.1)$$

où \mathbb{G} est un sous ensemble de \mathbb{R}^3 : $\mathbb{G} = [u_{min}; u_{max}] \times [v_{min}; v_{max}] \times [d_{min}; d_{max}]$ ($[u_{min}; u_{max}] \times [v_{min}; v_{max}]$ est défini par la résolution de la caméra et $[d_{min}; d_{max}]$ est défini par la portée de la caméra (voir section 1.5.2), \mathbf{q} le vecteur $(u, v, d)^T$ et c un scalaire qui représente la correction. Notre objectif est donc d'estimer la FCD g .

4.2.2 Formulation variationnelle

Le modèle proposé permet de représenter n'importe quelle fonction, y compris des fonctions avec des discontinuités. Il faut donc que la propriété de continuité de la FCD soit introduite dans la formulation de cette fonction. Comme présenté précédemment, nous proposons de modifier la formulation du problème d'étalonnage de la mesure de profondeur en utilisant une formulation flexible intégrant une contrainte de régularité sur la FCD à estimer. Pour cette raison, nous utilisons une

approche variationnelle par minimisation d'une fonctionnelle de coût qu'on note C avec :

$$C[g] = D[g] + S[g]. \quad (4.2)$$

Elle est composée d'un terme d'attache aux données D alignant g à des mesures de vérité terrain et d'un terme de régularisation S pénalisant la dérivée de g . Notons qu'un poids de régularisation contrôlant l'influence de chacun de ces termes est intégré dans l'équation précédente et sera présenté en détail dans ce qui suit. Cela permet de garantir la continuité de la FCD sans imposer une fonction paramétrique précise. Nous détaillons dans ce qui suit chacun des termes.

Terme d'attache aux données. Les données d'entrée de notre méthode est un ensemble Ω de $m \times n$ points-de-profondeur \mathbf{q} (m points dans n images). Pour chaque point $\mathbf{q} \in \Omega$, la vérité terrain d_{GT} est calculée. Elle représente la profondeur de référence dépourvue de toute distorsion. La méthode de calcul de cette dernière est détaillée dans la section 4.3.1. Nous notons $\alpha(\mathbf{q})$ la déviation entre la profondeur d_{TOF} mesurée par la caméra et la vérité terrain d_{GT} , elle est donnée par :

$$\alpha(\mathbf{q}) = d_{GT}(\mathbf{q}) - d_{TOF}(\mathbf{q}) \quad (4.3)$$

Le terme d'attache aux données est la distance entre la correction $g(\mathbf{q})$ et la déviation $\alpha(\mathbf{q})$. Ce terme est simplement la somme de l'erreur quadratique :

$$D[g] = \sum_{\mathbf{q} \in \Omega} \|g(\mathbf{q}) - \alpha(\mathbf{q})\|_2^2, \quad (4.4)$$

Notons que ce terme d'attache aux données n'est pas robuste aux données aberrantes. Il intègre des données des images de profondeur calculées par le capteur Temps-de-Vol qui présentent des bruits assez importants. Comme déjà présenté précédemment, ce bruit suit une distribution normale, c'est pour cette raison que la médiane sur une centaine d'images est utilisée pour le filtrer.

Terme de régularisation. Ce terme vise à introduire la propriété de continuité de la FCD sans imposer une fonction paramétrique précise. En plus, il présente une solution robuste aux données manquantes. Nous supposons que la FCD est lisse : pour des pixels voisins et pour des distances proches, les corrections de la mesure de profondeur à appliquer sont assez semblables. Dans notre cas, une régularisation 3D de nos données en u , v et d est appliquée. Une grande variété de termes de régularisation ou termes de lissage est proposée dans la littérature. Ils

sont construits à partir des dérivées de la fonction, comme par exemple le terme de lissage quadratique d'ordre 1 :

$$S[g] = \int_{\mathbb{G}} \left(\frac{\partial g}{\partial u} \right)^2 + \left(\frac{\partial g}{\partial v} \right)^2 + \left(\frac{\partial g}{\partial d} \right)^2 d_{\mathbf{q}}, \quad (4.5)$$

ou le terme de lissage quadratique d'ordre 2, appelée énergie de courbure ("bending energy"). Dans notre cas, nous choisissons de minimiser l'énergie de courbure :

$$S[g] = \int_{\mathbb{G}} \left\| \frac{\partial^2 g}{\partial \mathbf{q}^2} \right\|_2^2. \quad (4.6)$$

4.2.3 Discrétisation

Le modèle proposé ci-dessus permet de représenter la FCD dans l'espace \mathbb{G} . En pratique, il nécessite que la FCD prenne un nombre de valeurs fini. C'est pourquoi, dans l'approche proposée l'espace \mathbb{G} doit être discrétisé en points (un ensemble de triplets u, v, d) pour obtenir l'espace discret que nous notons \mathbb{H} . La discrétisation spatiale la plus naturelle est formée par les pixels de l'image $\{1 \dots u_{max}\} \times \{1 \dots v_{max}\}$. La discrétisation en profondeur est effectuée avec un pas de distance régulier Δd . Nous définissons alors un ensemble de distances $\mathcal{D} = \{d_1 \dots d_q\}$ (voir figure 4.1) pour lesquels la correction associée à chaque pixel de l'image (u, v) doit être estimée. Ce choix de discrétisation nous permet de contrôler le nombre de valeurs (de correction) à conserver en mémoire (qui dépend de la valeur de Δd) et par suite de réduire l'empreinte mémoire. Cela permet aussi de gagner en temps de calcul dans l'étape d'étalonnage ainsi que dans l'étape de correction (voir section 2.2.1). La fonction de correction de la distorsion (FCD) discrétisée est notée :

$$f : \mathbb{H} \rightarrow \mathbb{R}, \quad (4.7)$$

avec $\mathbb{H} = \{1 \dots u_{max}\} \times \{1 \dots v_{max}\} \times \{d_1 \dots d_q\}$. En pratique, nous cherchons à estimer la fonction f dans l'espace discret \mathbb{H} . Nous présentons dans ce qui suit comment exprimer le terme d'attache aux données et le terme de régularisation sous la forme discrète (en fonction de la FCD discrète f).

Terme d'attache aux données. Nous pouvons exprimer la FCD continue g (présentée dans l'équation (4.4)) en fonction de la FCD discrète f par une interpolation linéaire. Pour expliquer le principe, nous considérons l'exemple du point $\mathbf{q} = (u \ v \ d)^\top \in \Omega$ illustré par la figure 4.1. Il peut être exprimé en fonction de $(u \ v \ d')^\top$ et $(u \ v \ d'')^\top$ (voir figure) comme suit :

$$d = a(d)d' + (1 - a(d))d'',$$

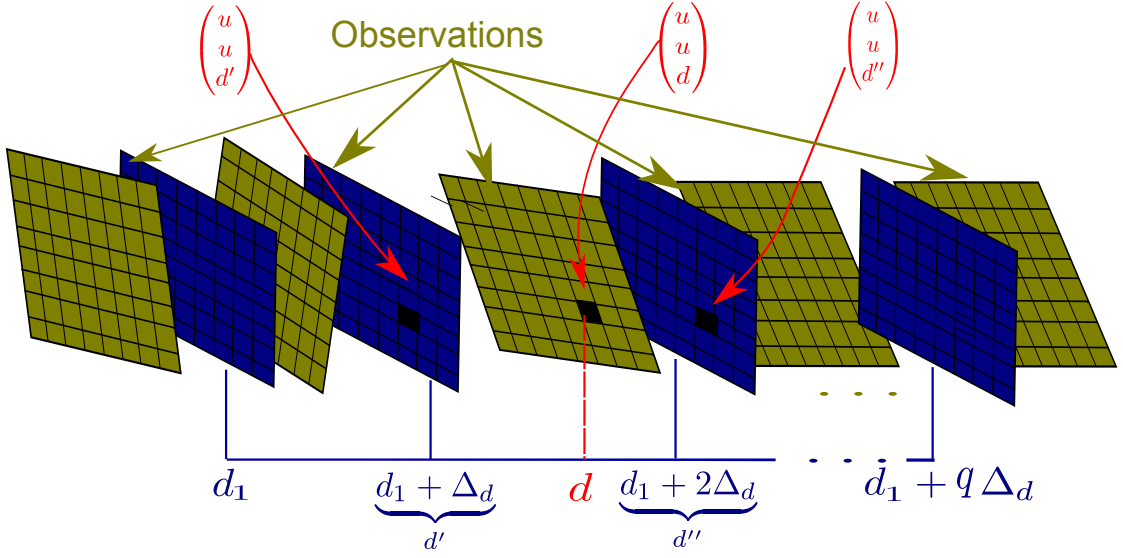


FIGURE 4.1 – Espace \mathbb{H} obtenu par discrétisation de l'espace continu \mathbb{G} avec un pas de distance régulier Δd . \mathbb{H} (ensemble des images de profondeur en bleu) obtenu à partir des observations (les images de profondeur en jaune).

avec $d' \in \mathcal{D}, d'' \in \mathcal{D}$ représentant respectivement les distances $d_1 + \Delta_d$ et $d_1 + 2\Delta_d$ et $a(d)$ un scalaire (le coefficient de pondération de l'interpolation linéaire). g peut alors être exprimée en fonction de f comme suit :

$$g \begin{pmatrix} u \\ u \\ d \end{pmatrix} = a(d) f \begin{pmatrix} u \\ v \\ d' \end{pmatrix} + (1 - a(d)) f \begin{pmatrix} u \\ v \\ d'' \end{pmatrix},$$

Le terme d'attache aux données D peut donc être exprimé sous la forme discrète comme :

$$D[f] = \sum_{\mathbf{q} \in \Omega} \|a(d)f(\mathbf{q}') + (1 - a(d))f(\mathbf{q}'') - \alpha(\mathbf{q})\|_2^2, \quad (4.8)$$

Terme de régularisation. Nous notons $\tilde{f}_{\mathbf{q}}$ la dérivée partielle seconde de f par rapport à \mathbf{q} dans l'espace discret \mathbb{H} pour éviter de reprendre la formule présentée dans l'équation (4.6). Le terme de régularisation sous la forme discrète est donné par :

$$S[f] = \sum_{\mathbf{q} \in \mathbb{H}} \|\tilde{f}_{\mathbf{q}}\|_2^2, \quad (4.9)$$

Notons qu'une fois la FCD discrète f estimée, la FCD continue g peut être obtenue à partir de f par simple interpolation comme ci-dessus.

En conclusion nous pouvons dire que notre méthode permet d'approximer la FCD et que la qualité de cette approximation dépend du pas de discrétisation Δd . Le choix de ce pas est critique. En effet, il doit assurer un bon compromis entre la précision souhaitée de la FCD et l'empreinte mémoire.

4.2.4 Résolution

Notre système composé du terme d'attache aux données (4.8) et du terme de régularisation (4.9) peut être réécrit sous une forme matricielle. Nous détaillons dans ce qui suit le passage en représentation matricielle pour chacun des termes.

Terme d'attache aux données. Soit \mathbf{U} le vecteur concaténant toutes les corrections à estimer par f dans l'espace discret \mathbb{H} . En représentation matricielle

$f \begin{pmatrix} u \\ v \\ d \end{pmatrix}$ est notée $\mathbf{U}(u, v, d)$. D est donnée par :

$$D(\mathbf{U}) = \|\mathbf{L}\mathbf{U} - \mathbf{b}\|_2^2, \quad (4.10)$$

avec :

- $\mathbf{U} \in \mathbb{R}^{mq}$, mq nombre d'inconnus ;
- $\mathbf{L} \in \mathbb{R}^{mn \times mq}$: la matrice des indices $a(d)$ et $1 - a(d)$;
- $\mathbf{b} \in \mathbb{R}^{mn}$: le vecteur contenant les $m \times n$ observations $\alpha(\mathbf{q})$.

Terme de régularisation. Le terme de régularisation est composé des dérivées partielles secondes de f par rapport aux trois composantes u, v, d (équation (4.9)). L'approximation de ces dérivées repose sur la méthode des différences finies. Nous notons ses dérivées respectivement $\mathbf{U}_u, \mathbf{U}_v, \mathbf{U}_d$. Nous rappelons que \mathbf{U} est le vecteur concaténant toutes les corrections à estimer par f dans \mathbb{H} . Par exemple, dans notre cas nous avons choisi la dérivée à droite :

$$\mathbf{U}_u(u, v, d) = \mathbf{U}(u + 1, v, d) - \mathbf{U}(u, v, d).$$

Elles sont représentées par les matrices $\mathbf{M}_u, \mathbf{M}_v, \mathbf{M}_d$ tel que :

$$\mathbf{U}_u = \mathbf{M}_u \mathbf{U} \quad \mathbf{U}_v = \mathbf{M}_v \mathbf{U} \quad \mathbf{U}_d = \mathbf{M}_d \mathbf{U}.$$

Les trois matrices $\mathbf{M}_u, \mathbf{M}_v, \mathbf{M}_d$ représentent les opérateurs de différences finies pour l'approximation des dérivées premières. Ces matrices ont une structure très creuse comme illustré par la figure 4.2.

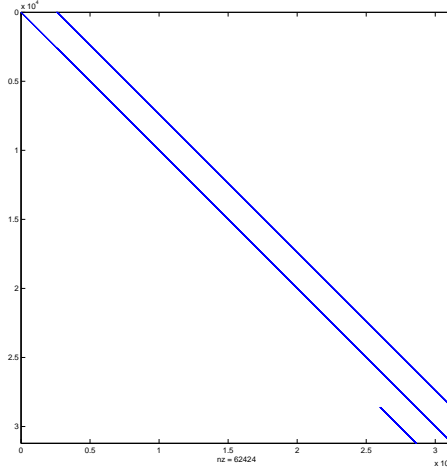


FIGURE 4.2 – Structure de la matrice M_u . Cette matrice a une structure très creuse. Seules les lignes de la matrice ayant au moins un élément non-nul sont représentées. Les matrices M_v et M_d ont la même structure.

D'une manière similaire, M_{uu} , M_{uv} , M_{ud} , M_{vd} , M_{vv} , M_{dd} sont calculées, correspondants aux opérateurs de différences finies pour l'approximation des dérivées secondes. Nous obtenons alors :

$$S(\mathbf{U}) = \mathbf{U}^T \mathbf{M} \mathbf{U}, \quad (4.11)$$

avec :

$$\begin{aligned} \mathbf{M} = & \lambda_u^2 \mathbf{M}_{uu}^T \mathbf{M}_{uu} + 2\lambda_u \lambda_v \mathbf{M}_{uv}^T \mathbf{M}_{uv} + 2\lambda_u \lambda_d \mathbf{M}_{ud}^T \mathbf{M}_{ud} \\ & + 2\lambda_v \lambda_d \mathbf{M}_{vd}^T \mathbf{M}_{vd} + \lambda_v^2 \mathbf{M}_{vv}^T \mathbf{M}_{vv} + \lambda_d^2 \mathbf{M}_{dd}^T \mathbf{M}_{dd}, \end{aligned} \quad (4.12)$$

où $\mathbf{M} \in \mathbb{R}^{mn \times mn}$ ($m \times n$ est le nombre d'observations) et $\lambda_u, \lambda_v, \lambda_d$ contrôlent respectivement l'influence de la régularisation en u, v, d , ce qui nous permet d'utiliser un poids de régularisation adéquat pour chaque direction. Plus de détails sur le calcul de \mathbf{M} sont donnés dans [Prasad *et al.* 2006].

Fonction de coût. A partir des équations ((4.10)) et ((4.11)), la fonction de coût (4.2) peut s'écrire sous la forme suivante :

$$C(\mathbf{U}) = \mathbf{U}^T \mathbf{M} \mathbf{U} + \|\mathbf{L} \mathbf{U} - \mathbf{b}\|_2^2 \quad (4.13)$$

Résolution pour un jeu de λ donné. Minimiser C revient à résoudre le système matriciel suivant :

$$(\mathbf{M} + \mathbf{L}^T \mathbf{L}) \mathbf{U} = \mathbf{L}^T \mathbf{b} \quad (4.14)$$

Notons que pour un jeu de $\lambda_x, \lambda_y, \lambda_d$ donné, ce système d'équations peut être résolu par inversion de matrice :

$$\mathbf{U} = (\mathbf{M} + \mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T \mathbf{b}, \quad (4.15)$$

La solution est obtenue par inversion d'une matrice qui peut être de grande dimension. Cependant, cette dernière est extrêmement creuse ce qui rend le système soluble. Les matrices \mathbf{M} et \mathbf{L} ont une structure très creuse comme illustré par la figure 4.3.

Nous notons qu'il existe d'autres méthodes pour résoudre ce système d'équations. Citons par exemple la méthode itérative de "*Gauss-Seidel*" qui limite l'empreinte mémoire comparée à la méthode de la pseudo-inverse.

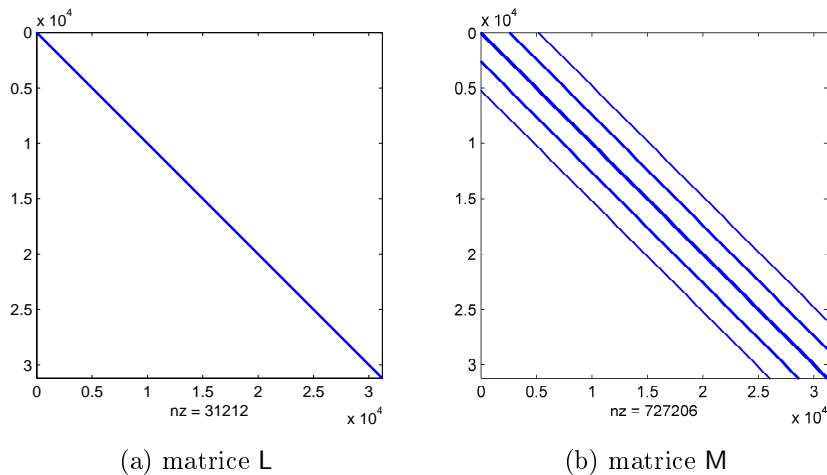


FIGURE 4.3 – Structure des matrices \mathbf{L} et \mathbf{M} . Seules les lignes des matrices ayant au moins un élément non-nul sont représentées. Ces matrices ont une structure très creuse.

Une des limites de l'approche proposée est qu'elle peut être coûteuse en espace mémoire dans quelques cas (*e.g.* caméra de haute résolution, de grande portée). Cependant, ceci ne constitue pas un grand problème dans notre cas étant donné la faible résolution des images et le domaine de distance limité par l'espace discret \mathbb{H} . Néanmoins, il est possible, lors de l'étape d'étalonnage de réduire l'empreinte mémoire en effectuant un échantillonnage des cartes de profondeur dans le plan

image (c.a.d sous échantillonner u et v). Ensuite, lors de l'étape de correction, une interpolation en ligne est utilisée pour couvrir tous les pixels. Dans notre expérimentation, pour chaque bloc de 4×4 pixels de l'image, la médiane est considérée.

4.3 Résultats expérimentaux

Dans cette section, le protocole expérimental mis en œuvre est tout d'abord présenté, ensuite les résultats obtenus et enfin une comparaison avec deux méthodes de l'état de l'art.

4.3.1 Protocole expérimental

Comme il a été exposé dans la section 4.2.1, l'approche proposée nécessite de disposer d'un jeu de cartes de profondeur pour lequel une vérité terrain est disponible. Ces données sont obtenues en observant un mur blanc à différentes distances. Nous présentons tout d'abord l'approche de calcul de vérité terrain proposée et ensuite les données utilisées par notre méthode.

4.3.1.1 Calcul de la vérité terrain

La méthode proposée se base sur le calcul de la déviation entre la vérité terrain et les mesures de profondeur associées (issues du capteur). La vérité terrain représente les mesures de profondeur de référence dépourvues de toute distorsion. Le calcul d'une vérité terrain précise est donc une étape importante dans notre méthode. Comme déjà présenté dans la section 2.2.4 du chapitre 2, il existe deux méthodes de calcul de la vérité terrain. La première est l'utilisation d'outils de mesure indépendants qui sont coûteux et ne sont pas disponibles dans notre laboratoire. La deuxième est l'utilisation d'un objet d'étalonnage observé par un système mono-caméra ou multi-caméra. Comme nous favorisons la précision au détriment de la mise en œuvre, nous avons choisi une solution basée sur un système multi-caméra. Bien que la mise en œuvre ne soit pas simple, elle repose sur un équipement peu coûteux, et sur des algorithmes d'étalonnage classiques et disponibles.

Nous avons choisi de coupler la caméra Temps-de-Vol à une caméra RGB standard de résolution 1600×1200 pixels. Le couple stéréo des deux caméras (RGB-TOF) rigidement liées a été étalonné (plus de détail sur l'étalonnage d'un couple stéréo sont donnés dans la section B.1.3.3 de l'Annexe B). Il existe plusieurs objets de calibrage. Nous avons choisi d'utiliser une mire plane avec de petits cercles.

La détection des points de la mire est effectuée avec une méthode de détection sous-pixelique ce qui permet de garantir une meilleure précision.

Une fois les caméras étalonnées (leurs paramètres intrinsèques et extrinsèques sont connus), nous calculons la vérité terrain par l'approche d'estimation de pose en utilisant la paire de caméra stéréo (RGB-TOF). La figure 4.4 explique le principe de calcul de la vérité terrain (les cartes de profondeur de référence). Un mur d'étalonnage instrumenté sur lequel sont disposées des cibles dont les positions sont connues (exemple le point M) est utilisé. Ce mur représente notre objet d'étalonnage. Le calcul de la vérité est effectué en trois étapes :

- La pose du mur dans le repère de la caméra RGB est calculée à partir des cibles collées sur le mur. Cette pose est notée $[\mathbf{R}_1 \ \mathbf{T}_1]$ (\mathbf{R}_1 matrice de rotation et \mathbf{T}_1 vecteur de translation). Elle est calculée par l'approche d'estimation de pose d'une caméra à partir de correspondances de points 3D/2D qui est détaillée dans la section B.1.3.2 de l'Annexe B.
- La pose $[\mathbf{R}_1 \ \mathbf{T}_1]$ est par la suite utilisée pour déduire la pose du mur dans le repère de la caméra Temps-de-Vol. Cette dernière est notée $[\mathbf{R}_2 \ \mathbf{T}_2]$. Elle est calculée en utilisant la transformation rigide (les paramètres stéréo) reliant les deux caméras notée $\mathbf{R}_{RGB \rightarrow TOF} | \mathbf{T}_{RGB \rightarrow TOF}$:

$$\begin{aligned} \mathbf{R}_2 &= \mathbf{R}_{RGB \rightarrow TOF} \mathbf{R}_1 \\ \mathbf{T}_2 &= \mathbf{R}_{RGB \rightarrow TOF} (\mathbf{T}_1 + \mathbf{T}_{RGB \rightarrow TOF}) \end{aligned} \quad (4.16)$$

- Pour chaque pixel m de l'image de la caméra Temps-de-Vol, la distance de référence d_{GT} associée est calculée. Cette distance correspond à la distance entre le centre de la caméra C et le point 3D M. Les coordonnées de ce point sont obtenues en calculant l'intersection du plan du mur Π et la droite D qui passe à la fois par C et le point m (projeté de M dans le plan image). L'équation de la droite D est donnée par l'équation (B.15) de la rétroprojection perspective introduite dans la section B.1.1.1 de l'annexe B. Le plan Π est défini par la donnée de trois positions non colinéaires des cibles.

La vérité terrain est calculée pour différentes positions de la caméra de telle sorte qu'elle couvre la plage de distance de la caméra Temps-de-Vol. A chaque position, l'image de profondeur associée doit aussi être sauvegardée. C'est une procédure difficile à mettre en œuvre. En effet, à chaque position nous effectuons les quatre étapes suivantes :

- Les cibles sont collés de telle sorte qu'elles couvrent tout le mur pour augmenter la précision de l'estimation de pose ;
- La vérité terrain est calculée ;
- Les cibles sont enlevés du mur pour éviter les distorsions liées aux autres facteurs de distorsion ;

- L'image de profondeur mesurée par la caméra Temps-de-Vol est sauvegardée.

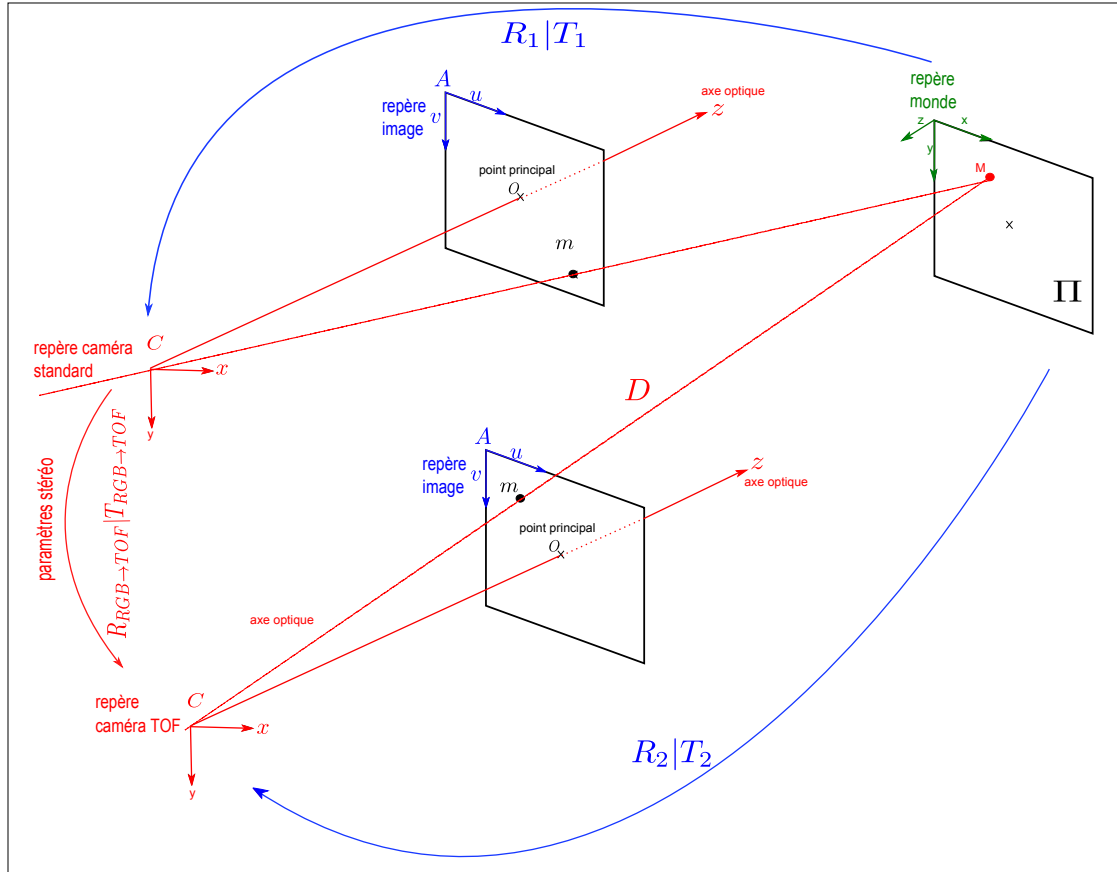


FIGURE 4.4 – Principe du calcul de la vérité terrain par l’approche d’estimation de pose utilisant la paire de caméra stéréo (RGB-TOF) et un mur instrumenté.

Nous n’avons pas de moyen robuste pour évaluer la précision de la vérité terrain calculée par l’approche d’estimation de pose. Par contre, nous avons une idée sur les caractéristiques de la distorsion de la mesure de profondeur. Comme déjà présenté dans la section 1.5.3.2 du chapitre 1, cette dernière augmente du centre de l’image vers les bords et oscille en fonction de la profondeur. Nous avons choisi un échantillon de point 2.5D et avons calculé la déviation entre la vérité terrain et la mesure de profondeur calculée par la caméra Temps-de-Vol. La figure 4.5 présente le résultat obtenu. La courbe en rouge correspond au pixel central et les courbes bleue et verte correspondent respectivement au pixel du coin supérieur gauche et au pixel du coin inférieur droit de la carte de profondeur. D’après cette figure, les caractéristiques de la distorsion sont vérifiées. La distorsion varie en fonction de la distance et de la position du pixel dans l’image. Pour les pixels présentés sur

cette figure, l'erreur varie de -2cm à 14cm . Cependant, nous nous sommes pas en mesure de garantir une précision millimétrique de la vérité terrain. En effet, plusieurs sources d'erreurs peuvent s'accumuler avec notre approche de calcul de la vérité terrain et seront détaillées dans la section 4.4.

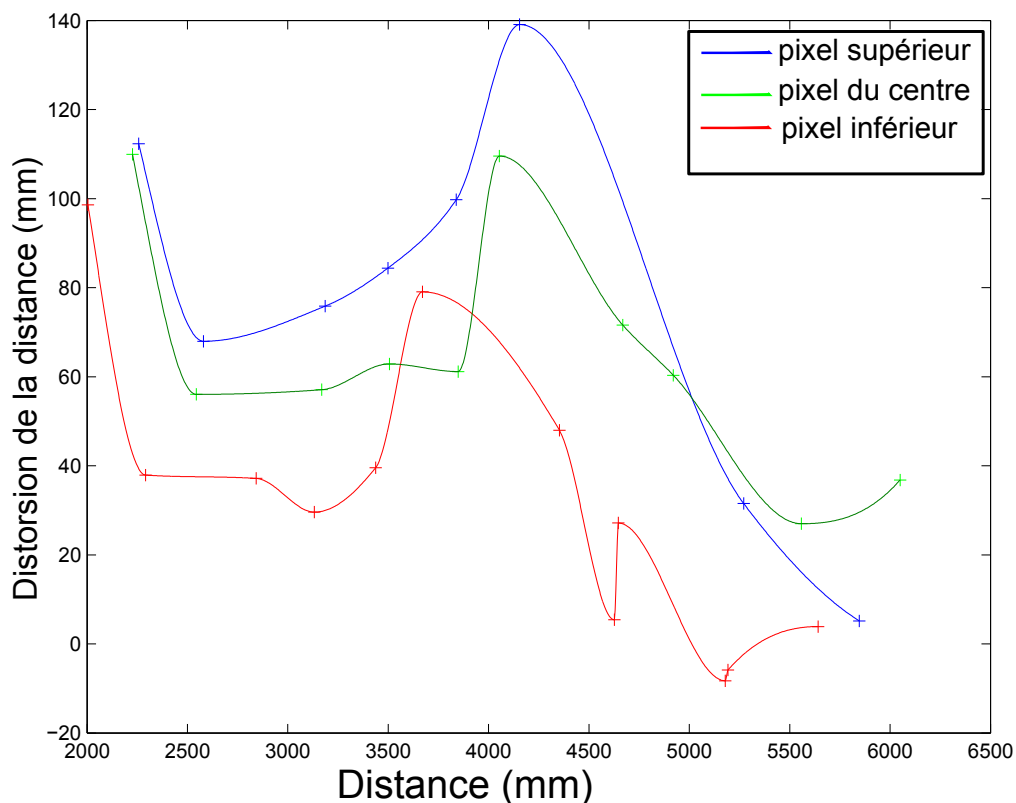


FIGURE 4.5 – Distorsion de la mesure de profondeur calculée pour trois différents points 2.5D de l'image et à différentes distances. La courbe en rouge correspond au pixel central et les courbes bleue et verte correspondent respectivement au pixel du coin supérieur gauche et au pixel du coin inférieur droit de la carte de profondeur.

4.3.1.2 Données utilisées

Comme présenté dans la section 1.4 du chapitre 1, la caméra Temps-de-Vol utilisée dans l'expérimentation est la PMD Vision CamCube3.0 de résolution 204×204 avec une f_{mod} réglée à 20 MHz et un TI fixé à 14 ms.

La méthode d'étalonnage proposée a été réalisée en utilisant un jeu de 14 cartes de profondeur couvrant une plage de distance approximative allant de 2,4 m à 6,9 m pour lequel la vérité terrain associée est calculée. L'ensemble Ω est défini par $m = 204 \times 204$ et $n = 14$. La FCD f est estimée dans l'espace discret

$\mathbb{H} = \{1 \dots 204\} \times \{1 \dots 204\} \times \{2,5 \text{ m}, 2,8 \text{ m}, \dots 5,8 \text{ m}\}$. Notons que nous nous limitons à cet intervalle de profondeur pour éviter de représenter les mesures de profondeur aux bords qui sont invalides.

La méthode proposée a été évaluée, par la suite, sur un jeu de données non utilisé au cours de l'étalonnage. Ces données représentent des cartes de profondeur qui ont été choisies à différentes distances et ne se trouvant pas dans le jeu d'entraînement. Notons que la médiane est considérée pour toutes les cartes de profondeur utilisées pour filtrer le bruit.

4.3.2 Détermination des valeurs optimales de λ

Dans la section 4.2.4, nous avons présenté la méthode de résolution de notre système en supposant que les valeurs optimales des poids de régularisation (λ_u , λ_v , λ_d) sont déjà connus. Cependant, ces valeurs restent à déterminer. Pour ce faire, différents étalonnages ont été effectués en variant (λ_u , λ_v , λ_d) pour déterminer leurs valeurs optimales (avec $\lambda_u = \lambda_v$ parce que la variation en u et v est similaire). Les différents étalonnages sont testés sur un ensemble d'images de profondeur non utilisées au cours de l'étalonnage. Après correction de la distorsion de la mesure de profondeur, pour chaque pixel de ces images, l'erreur de la mesure de profondeur résiduelle est calculée en comparant les mesures de profondeur après correction (de la distorsion de la mesure de profondeur) à la vérité terrain. La valeur médiane de l'erreur résiduelle de tous les pixels des images de test est considérée pour comparer les différents étalonnages. Les valeurs optimales de λ sont choisies de telle sorte qu'elles minimisent l'erreur médiane sur tous les pixels observés. Les résultats illustrés par la figure 4.6 représentent les erreurs médianes calculées après correction de la distorsion de la mesure de profondeur. Ils sont représentés par une surface sur la figure 4.6a : l'axe des X présente les valeurs de $\lambda_u = \lambda_v$ et l'axe des Y présente les valeurs de λ_d . La figure 4.6b correspond à un zoom autour de la valeur minimale des erreurs médianes (le carré sur la figure 4.6a). Les valeurs optimales des poids de régularisation $\lambda_u = \lambda_v$ et λ_d sont respectivement 2 et 1 comme le montre la figure.

4.3.3 Résultats expérimentaux de notre étalonnage

La figure 4.7 montre une comparaison entre les erreurs avant et après correction de la mesure de profondeur des images de test. L'erreur, pour chaque pixel, est obtenue en comparant la carte de profondeur corrigée à la vérité terrain. Chaque ligne correspond à une image de test, à gauche, l'erreur avant correction et à droite celle après correction. Par exemple, pour la première image de test, avant

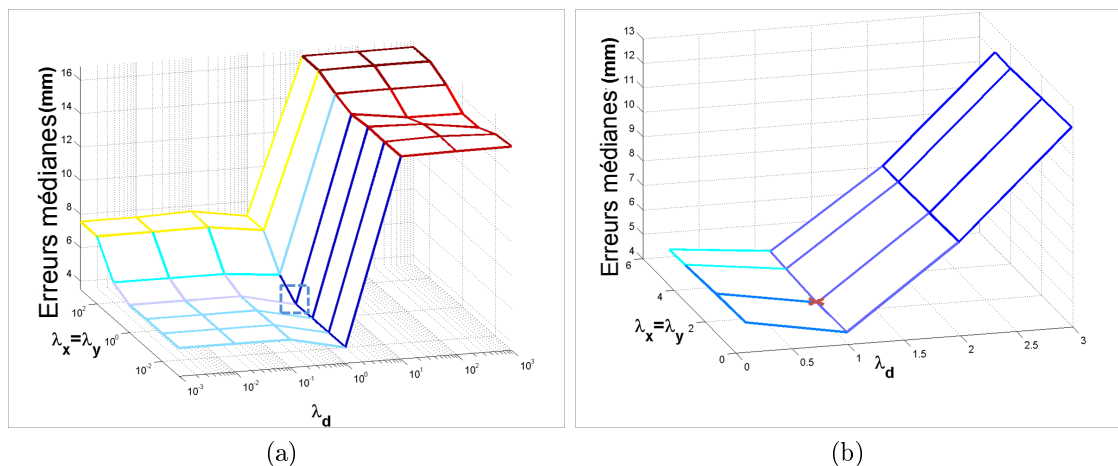


FIGURE 4.6 – Erreurs médianes avec différentes valeurs de $\lambda_u = \lambda_v$ et λ_d .

correction l'erreur varie entre 20 mm et 75 mm (figure 4.7a). Nous observons un manque de précision plus élevé au niveau des bords de l'image qu'au centre de l'image. Comparé aux résultats après correction de la mesure de profondeur (figure 4.7b), l'erreur restante est centrée en zéro et est homogène sur toute l'image. Ceci est aussi visible sur la figure 4.8 qui représente une comparaison de la distribution des erreurs calculées avant et après correction pour les mêmes images de test. Dans le cas de la première image, par exemple, avant correction (figure 4.8a) l'histogramme présente un pic autour de 40mm (53% des pixels de l'image), alors qu'après correction (figure 4.8b), le pic est autour de zéro (65% des pixels de l'image).

4.3.4 Comparaison avec des méthodes de l'état de l'art

Une comparaison de nos résultats avec des méthodes de l'état de l'art sur des images de profondeur issues de notre camera est présentée. Les méthodes d'étalonnage de la mesure de profondeur présentées dans [Lindner & Kolb 2006] et dans [Kahlmann *et al.* 2006] ont été implémentées et testées sur le même jeu de données. Le tableau 4.1 présente les résultats. Une amélioration de la précision des cartes de profondeur est obtenue avec l'approche proposée. En effet, l'erreur résiduelle après correction est estimée en valeur médiane sur différentes images à 4,42 mm alors qu'elle est de 6,39 mm et de 8,49 mm avec les autres méthodes.

Comme présenté précédemment, notre méthode de calcul de la vérité terrain ne peut pas garantir une précision millimétrique de cette dernière. De plus, nous n'avons pas de moyen robuste pour évaluer la précision de la vérité terrain que

nous avons obtenue. Nous proposons donc, dans le chapitre 6, une façon pour évaluer la qualité de la méthode de l'étalonnage proposée sans dépendre de la vérité terrain. Nous observons un objet de dimensions connues et comparons ses dimensions mesurées avec la caméra Temps-de-Vol avant et après correction de la mesure de profondeur.

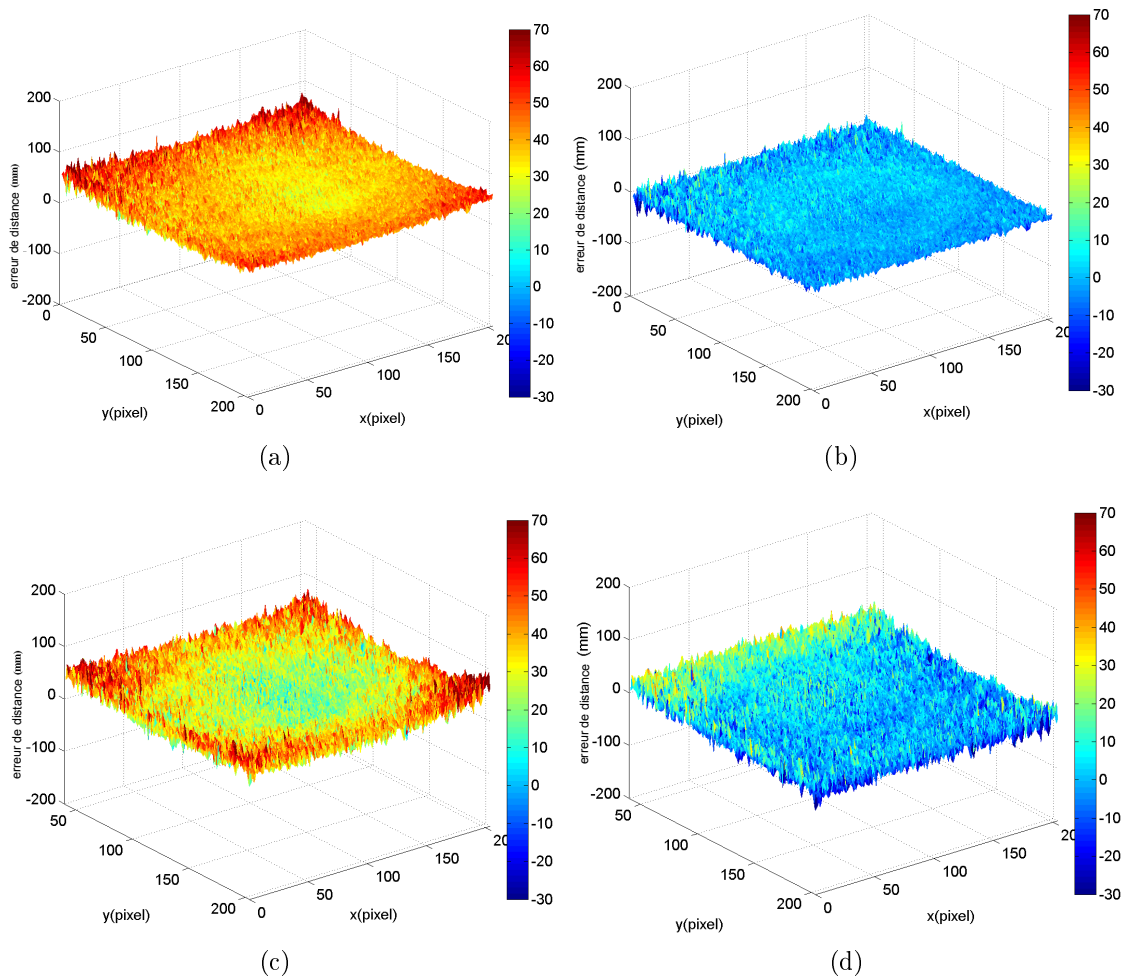


FIGURE 4.7 – Comparaison de l'erreur avant (a),(c) et après (b),(d) correction de la mesure de profondeur.

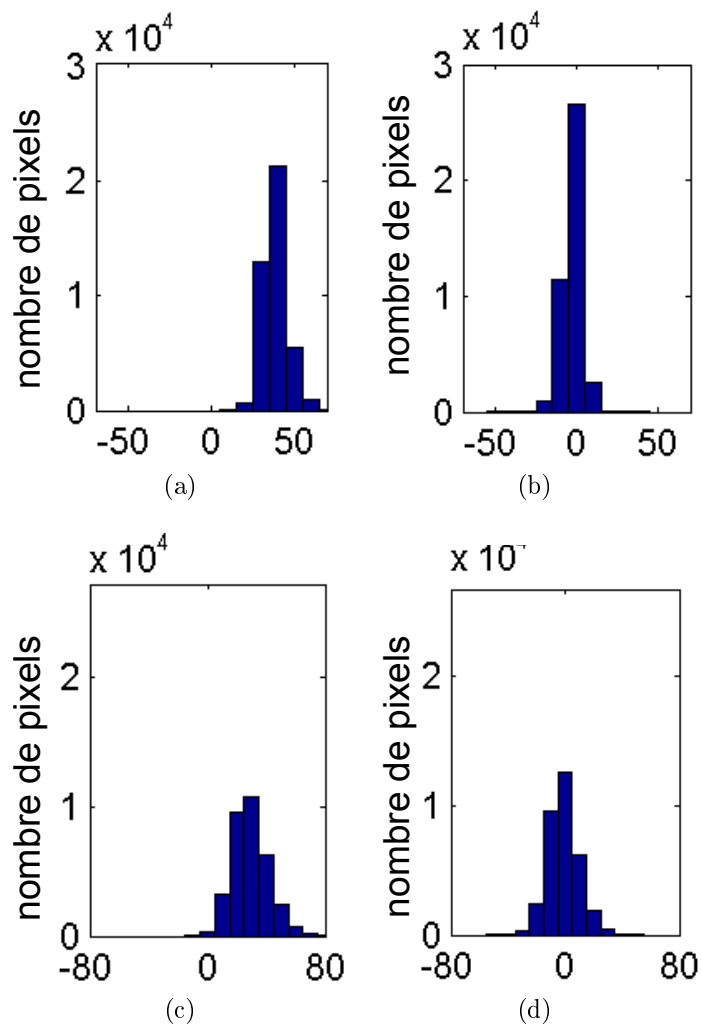


FIGURE 4.8 – Comparaison de la distribution de l’erreur avant (a),(c) et après (b),(d) correction de la mesure de profondeur.

Tableau 4.1 – Comparaison avec deux approches de la littérature

| | | Erreurs résiduelle (mm) | |
|-----------------------------|-------------------------------|-------------------------|---------|
| | | Médiane | Moyenne |
| Avant correction | | 35,03 | 34,30 |
| Après correction avec | Notre méthode | 4,42 | 6,16 |
| | [Lindner & Kolb 2006] | 6,39 | 8,70 |
| | [Kahlmann <i>et al.</i> 2006] | 8,49 | 10,83 |

4.4 Discussion et conclusion

Ce chapitre a décrit la première méthode d'étalonnage de la mesure de profondeur proposée. Elle se base sur le calcul de la déviation entre les mesures 2.5D issues du capteur Temps-de-Vol et celles obtenues à partir du calcul de la vérité terrain. Une formulation variationnelle introduisant une contrainte de régularité est adoptée dans notre méthode. Une telle formulation est utilisée, à notre connaissance, pour la première fois dans ce contexte présentant ainsi une solution permettant d'obtenir une correction plus précise, comparée à d'autres méthodes de la littérature, sans imposer une fonction paramétrique particulière. Les résultats expérimentaux montrent la validité de l'approche proposée. En effet, une meilleure précision est atteinte, comparée aux méthodes de l'état de l'art. Une amélioration de l'approche serait de prendre en compte les autres facteurs d'erreur. Notre approche peut être étendue facilement notamment en adaptant l'étalonnage de la mesure de profondeur en fonction du TI . Des FCDs avec différents TI peuvent être modélisées, ensuite la FCD pour un TI donné est obtenue par simple interpolation $2D$. Le facteur de réflectivité des objets de la scène peut aussi être pris en compte en introduisant l'information d'intensité. Une façon très simple de faire est de modéliser des FCDs pour différents niveaux de gris (blanc, gris, noir), ensuite la FCD pour un niveau de gris donné est obtenue par simple interpolation $2D$. Il serait aussi intéressant de tester l'approche proposée sur d'autres capteurs comme le Kinect.

La limite principale de cette méthode est la mise en œuvre difficile pour le calcul de la vérité terrain. Comme déjà présenté précédemment, les cibles utilisées pour calculer la vérité terrain doivent être collées sur le mur puis enlevées. Cela est répété plusieurs fois au cours du processus de calcul de la vérité terrain. De plus, l'approche de calcul de la vérité terrain ne fournit pas des mesures de haute précision. En effet, l'erreur est accumulée au cours des trois processus suivants :

- étalonnage de la caméra Temps-de-Vol (TOF).
- étalonnage de la caméra RGB.
- étalonnage du système stéréo TOF-RGB.

Notons que l'erreur associée au processus d'étalonnage de la caméra Temps-de-Vol est plus importante que celle des caméras standard RGB. En effet, l'estimation de pose est peu précise car la faible résolution de la caméra Temps-de-Vol engendre une extraction imprécise des points de la cible.

Il serait intéressant de disposer d'une méthode d'étalonnage de la mesure de profondeur plus facile à mettre en œuvre. Nous proposons une deuxième méthode qui ne nécessite que peu de correspondances de points entre la vérité terrain et les images du capteur Temps-de-Vol. Elle est basée principalement sur la contrainte

de planarité et sera présentée en détail dans le chapitre 5.

Étalonnage de la distorsion à partir de structures planes

Dans le chapitre précédent, nous avons présenté notre première méthode d'étalonnage de la mesure de profondeur visant à obtenir une correction plus précise que les autres méthodes de la littérature. Cependant, sa limite principale est la mise en œuvre délicate pour obtenir la vérité terrain pour chaque pixel et pour chaque pas de distance. Ce problème est commun à toutes les méthodes de la littérature. Dans ce chapitre, nous présentons notre deuxième méthode d'étalonnage visant la simplicité et la rapidité de la mise en œuvre. La méthode proposée est basée principalement sur des contraintes géométriques de coplanarité : les données d'entrée sont des images de structures planes. Elle nécessite très peu de correspondances de points entre la vérité terrain et les images du capteur Temps-de-Vol (4 points minimum). Dans notre méthode, nous utilisons différents hyper-paramètres tels que le nombre de point contrôlant le modèle d'étalonnage et le paramètre de régularisation. Nous proposons de déterminer leurs valeurs optimales par la technique de "golden section search". Cette méthode d'étalonnage a été testée sur des données de simulation ainsi que sur des données réelles. En plus de sa simplicité de mise en œuvre, la méthode proposée donne de bons résultats. Une comparaison avec notre première méthode d'étalonnage ainsi que deux méthodes de la littérature est également présentée. Les travaux décrits dans ce chapitre ont donné lieu à la publication [Belhedi et al. 2012b].

5.1 Motivation

Notre méthode d'étalonnage de la mesure de profondeur présentée dans le chapitre précédent permet d'avoir une correction précise. Cependant sa mise en œuvre est délicate. La difficulté est avant tout liée à l'obtention d'une vérité terrain précise pour chaque pixel des images d'étalonnage et pour une large gamme de distances. Ce problème est commun à l'ensemble des méthodes de la littérature actuelles puisqu'elles reposent toutes sur le principe de minimisation de l'erreur entre la FCD à estimer et la déviation entre les mesures 2.5D issues du capteur Temps-de-Vol et celles obtenues à partir du calcul de la vérité terrain (*e.g.* de 0,75 m jusqu'à 7,5 m avec un pas de 10 cm pour chaque pixel [Lindner & Kolb 2006]). Comme déjà présenté dans la section 2.2.4, l'acquisition de cette vérité terrain est délicate pour différentes raisons. En effet, un système additionnel est nécessaire, *i.e.* un système de rail de mesure "*track line system*" [Lindner & Kolb 2006, Kahlmann *et al.* 2006] ou un système utilisant un objet d'étalonnage dont la géométrie doit être parfaitement connue (basé sur une extraction de cibles détaillée dans la section 2.2.4)[Schiller *et al.* 2008, Lindner & Kolb 2007, Belhedi *et al.* 2012c] (voir figure 5.1).


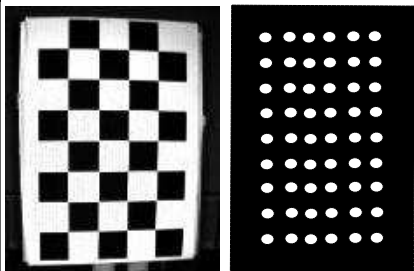

| Méthodes existantes basées sur le calcul de la vérité terrain par: | | Notre méthode utilise des vues d'un plan |
|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
|  |  |  |
| système de rail de mesure | système d'extraction de cibles | sans rail de mesure ni système d'extraction de cibles |

FIGURE 5.1 – Les approches classiques nécessitent un grand ensemble de mesures de vérité terrain obtenu par un système de rail de mesure "*track line system*" ou un système d'extraction de cibles. Le premier système est coûteux. Le second système nécessite une mise en œuvre délicate, longue et ne permet pas d'obtenir des mesures d'une grande précision à de longues distances. L'approche proposée utilise des vues d'un plan et nécessite très peu de correspondances de points entre la vérité terrain et les images du capteur (4 points suffisent).

Le premier système est coûteux. Le second système nécessite une mise en œuvre délicate et longue comme déjà expliqué dans la section 4.3.1.1. De plus, la précision recherchée est difficilement atteignable. Comme expliqué dans la section 4.4, le manque de précision est dû à l'accumulation des erreurs lors de l'étalonnage des caméras utilisées, principalement, celui de la caméra Temps-de-Vol.

L'objectif de ce chapitre est de proposer une deuxième méthode d'étalonnage plus simple à utiliser comparée à notre première méthode et aux méthodes de la littérature. La méthode proposée ne nécessite pas de disposer de la vérité terrain pour chaque pixel et pour chaque pas de distance. Elle utilise principalement une vérité terrain structurelle : elle est basée sur des contraintes géométriques de coplanarité. Elle nécessite un nombre réduit de correspondances de points entre la vérité terrain et les images du capteur. Son principe est expliqué en détail dans la section suivante.

Une comparaison avec notre première méthode et avec les deux méthodes principales de la littérature montre qu'en plus de sa simplicité de mise en œuvre, la méthode proposée garantit une bonne qualité de correction.

5.2 Présentation de la méthode

5.2.1 Introduction

Notre objectif est d'avoir une méthode d'étalonnage de la mesure de profondeur rapide à mettre en œuvre. Nous proposons d'estimer la FCD à partir d'une vérité terrain structurelle, c'est-à-dire en exploitant les propriétés géométriques intrinsèques à la scène observée. Il s'agit d'estimer la FCD qui permet d'aligner au mieux la reconstruction de la scène issue du capteur Temps-de-Vol et le modèle de la scène dont la géométrie est connue. Une scène ayant des propriétés géométriques remarquables est utilisée. Il s'agit d'une scène plane : un mur. Ce choix nous évite d'avoir une étape préalable d'estimation du modèle de la scène. Il s'agit de filmer un mur avec le capteur Temps-de-Vol afin d'avoir différentes vues. Les points 3D reconstruits à partir de ces vues ne sont pas coplanaires : ceci est causé par la distorsion de la mesure de profondeur. Notre objectif est d'estimer la FCD de telle façon que les points 3D de chaque vue du plan, après correction par cette fonction, tendent vers des points coplanaires. Cependant, l'équation du plan pour chaque vue n'est pas donnée. Elle doit être estimée simultanément avec la FCD. Nous optons pour un processus itératif qui alterne l'estimation de la FCD et l'équation du plan pour chaque vue. Il sera détaillé dans la suite.

La FCD obtenue est définie, comme nous le montrerons, à une transformation

affine près. Afin de lever cette ambiguïté, nous proposons une seconde étape d'étalonnage qui, contrairement aux méthodes de la littérature, nécessite une vérité terrain réduite. En effet, 4 correspondances de points entre la vérité terrain et les images du capteur au minimum sont nécessaires.

Avant de présenter plus en détail chacune des deux étapes de notre méthode, nous commençons par définir notre espace de travail. Notons que la première méthode présentée dans le chapitre 4 ainsi que toutes les méthodes de la littérature opèrent dans l'espace 2.5D (correction des points 2.5D $\mathbf{q} = (u \ v \ d)$). Dans la méthode proposée, nous opérons dans l'espace 3D. Ce choix d'estimer la FCD dans l'espace 3D permet de simplifier les calculs puisque l'objectif est de rendre coplanaires des points 3D notés $\mathbf{Q} = (X \ Y \ Z)$. Notons qu'une fois la FCD notée F estimée, celle définie dans l'espace 2.5D peut être obtenue à partir de F par une simple transformation présentée dans la section B.3.2. Notons également que les nuages de points 3D sont calculés à partir des images de profondeur issues de la caméra Temps-de-Vol comme décrit dans l'équation (B.37).

Nous présentons dans ce qui suit les deux étapes de notre méthode.

Correction de la Non Planarité (CNP). En observant des nuages de points 3D correspondants à des vues d'un mur plan, nous constatons que les points 3D de chaque vue ne sont pas coplanaires. Ceci est causé par la distorsion de la mesure de profondeur des caméras Temps-de-Vol. La première étape de notre méthode vise à estimer une fonction F de correction de la non planarité des nuages de points 3D correspondants à différentes vues d'un mur plan. Les données d'entrée sont des nuages de points d'un mur plan qui s'intersectent pour couvrir tout l'espace d'étalonnage (voir figure 5.2). Cela est facile à mettre en œuvre. En effet, il s'agit de filmer un mur plan sans contrainte sur l'orientation de la caméra et sans avoir à connaître la vérité terrain associée. L'essentiel est de couvrir tout l'espace d'étalonnage, c'est-à-dire toute la plage de distance supportée par la caméra avec différentes orientations de telle sorte à avoir des redondances de données dans l'espace 3D d'étalonnage.

Nous rappelons que dans cette étape nous estimons simultanément deux inconnus : la FCD et les équations des plans qui approximent au mieux les points 3D de chaque vue du mur. Cela est réalisé avec un processus itératif qui alterne l'estimation des deux inconnus (présenté en détail dans la section 5.2.4).

La CNP contraint la fonction de correction de la non planarité F à une transformation affine près que nous notons A .

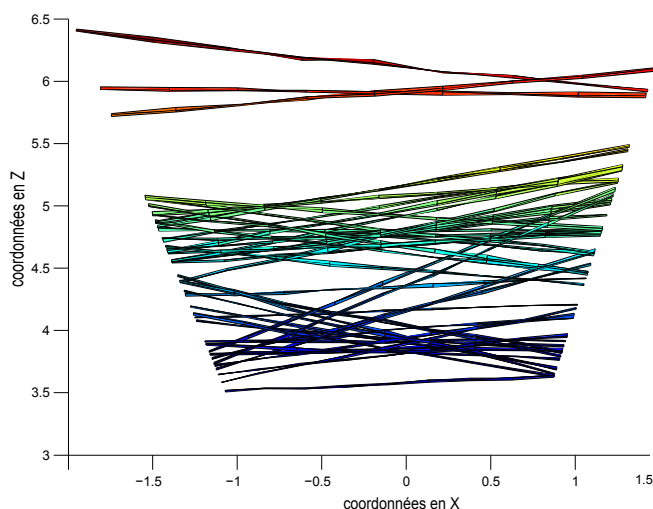


FIGURE 5.2 – Données d’entrée de notre méthode d’étalonnage projetées dans le plan XZ : un grand ensemble de vues d’un mur plan issues du capteur Temps-de-Vol qui s’intersectent pour couvrir tout l’espace d’étalonnage.

Correction Affine (CA). La correction obtenue après l’étape de CNP permet d’estimer une FCD définie à une transformation affine près. Cette transformation est choisie pour deux raisons : c’est la transformation la plus générale qui conserve le critère géométrique de coplanarité et qui conserve aussi l’énergie de lissage. Pour estimer cette transformation affine A , il suffit de disposer d’un faible ensemble de correspondances de points entre la vérité terrain et les images du capteur. En effet, A ne dépendant que de 12 paramètres, il suffit de disposer de 4 correspondances de points. Cependant, en pratique, nous utilisons plus de points pour avoir une transformation plus précise et robuste au bruit. La CA correspond à un problème d’optimisation de moindre carrés linéaires qui peut être résolu facilement.

L’avantage de la méthode proposée est qu’elle est plus facile à mettre en œuvre comparée aux autres méthodes puisqu’elle ne nécessite pas un grand nombre de mesures de vérité terrain.

Dans ce qui suit, nous nous concentrons sur la première étape qui est l’étape de CNP, puisque l’étape de CA est triviale une fois la vérité terrain obtenue.

5.2.2 Modèle de la CNP

La CNP estime la fonction de correction des points 3D $\mathbf{Q} = (X \ Y \ Z)$. Nous montrerons qu’il suffit de calculer la correction de la troisième composante Z que nous notons C_Z . La correction des deux autres composantes X et Y peut être

déduite à partir de C_Z . Nous définissons alors la fonction F de la CNP comme suit :

$$\begin{aligned} F : \quad \psi &\rightarrow \mathbb{R}, \psi \subset \mathbb{R}^3 \\ \mathbf{Q} &\rightarrow C_Z \end{aligned} \quad (5.1)$$

où ψ est un sous ensemble de \mathbb{R}^3 : $\psi = [X_{min}; X_{max}] \times [Y_{min}; Y_{max}] \times [Z_{min}; Z_{max}]$ et C_Z est un scalaire qui présente la correction de la troisième composante Z . Notons que F appartient à $L^2(\psi)$ ¹.

Notons $\mathbf{S} = (X_{corr} \quad Y_{corr} \quad Z_{corr})$ le point après correction de la mesure de profondeur, la composante X_{corr} peut être exprimée en fonction de la correction $C_Z = F(\mathbf{Q})$ et du point avant correction noté \mathbf{Q} en remplaçant dans l'équation (B.38) Z par $Z + C_Z$ comme suit :

$$\begin{aligned} X_{corr} &= (Z + C_Z) \frac{(u - c_u)d_u}{f_c} \\ &= X + \frac{(u - c_u)d_u}{f_c} C_Z \\ &= X + \frac{1}{Z} X C_Z \\ &= X + \frac{1}{Z} X F(\mathbf{Q}). \end{aligned} \quad (5.2)$$

De la même façon, nous exprimons la composante Y_{corr} en fonction de $F(\mathbf{Q})$. D'où l'expression du point corrigé \mathbf{S} :

$$\mathbf{S} = (X + \frac{1}{Z} X F(\mathbf{Q}) \quad Y + \frac{1}{Z} Y F(\mathbf{Q}) \quad Z + F(\mathbf{Q})) \quad (5.3)$$

qui peut être réécrit comme suit :

$$\mathbf{S} = \mathbf{Q} + \frac{1}{Z} F(\mathbf{Q}) \mathbf{Q} \quad (5.4)$$

5.2.3 Formulation variationnelle de la CNP

Le modèle de la CNP proposé intègre des données des images de profondeur issues du capteur Temps-de-Vol qui présentent des bruits assez forts. Il serait possible de considérer la médiane sur une centaine d'images pour filtrer le bruit. Cependant, dans la méthode proposée nous optons pour une solution rapide et simple à mettre en œuvre. Nous évitons cette solution qui augmente le temps

1. L'espace de Hilbert des fonctions intégrables deux fois.

d'acquisition. Il faut donc que la prise en compte du bruit soit introduite dans la formulation de cette fonction. Nous proposons alors une formulation intégrant une contrainte de régularité sur la FCD à estimer. Pour cette raison, nous utilisons une approche variationnelle par minimisation d'une fonctionnelle de coût qu'on note C :

$$\min_{F \in L^2, \mathcal{P}} C[F, \mathcal{P}] \stackrel{\text{def}}{=} \min_{F \in L^2, \mathcal{P}} C_d[F, \mathcal{P}] + \lambda C_s[F]. \quad (5.5)$$

Elle est composée d'un terme d'attache aux données C_d , d'un terme de régularisation C_s et d'un poids de régularisation λ contrôlant l'influence de chacun des termes. Le premier (C_d) estime simultanément F et l'ensemble des plans inconnus noté \mathcal{P} . Le deuxième (C_s) pénalise la dérivée de F . Nous détaillons dans ce qui suit chacun des deux termes.

Terme d'attache aux données. Les données d'entrée de la CNP correspondent à n vues d'un plan avec m points chacune :

$$\mathbf{Q}_{ij} \stackrel{\text{def}}{=} (X_{ij} \ Y_{ij} \ Z_{ij})^\top \quad (5.6)$$

où i est le numéro de la vue du plan et j est la position du pixel dans l'image. L'objectif est de rendre les points $\{\mathbf{Q}_j\}_{j=1}^m$ coplanaires. Si nous notons le point 3D après correction \mathbf{S}_{ij} , la contrainte géométrique de coplanarité est exprimée comme suit :

$$\mathbf{S}_{ij}, j=1, \dots, m \in \Pi_i, i=1, \dots, n, \quad (5.7)$$

où Π_i est le plan 3D inconnu, correspondant à la vue i . Nous représentons un plan Π_i par un vecteur $\mathbf{P}_i \in \mathbb{R}^3$ dont la norme est l'inverse de la distance au centre de la caméra et qui est colinéaire à la normale au plan. Les plans inconnus $\mathcal{P} = \{\mathbf{P}_1, \dots, \mathbf{P}_n\} \in \mathbb{R}^{3n}$ sont des variables latentes.

Le terme d'attache aux données minimise la distance entre les points corrigés S et le plan Π :

$$C_d[F, \mathcal{P}] \stackrel{\text{def}}{=} \sum_{i=1}^n \sum_{j=1}^m d^2(\mathbf{S}_{ij}, \Pi_i), \quad (5.8)$$

où $d^2(\mathbf{S}, \Pi)$ est la distance au carré du point \mathbf{S} au plan Π . Un choix pour d peut être la simple distance algébrique exprimée par :

$$d^2(\mathbf{S}, \Pi) \stackrel{\text{def}}{=} (\mathbf{S}^\top \mathbf{P} + 1)^2. \quad (5.9)$$

ou la distance euclidienne point-plan correspond à la distance orthogonale au plan exprimée par :

$$d^2(\mathbf{S}, \Pi) \stackrel{\text{def}}{=} \frac{(\mathbf{S}^\top \mathbf{P} + 1)^2}{\|\mathbf{P}\|_2^2}. \quad (5.10)$$

Dans notre cas, nous optons pour la deuxième distance, car la distance algébrique n'a pas de sens strictement physique contrairement à la distance Euclidienne.

Terme de régularisation. Ce terme vise à introduire la propriété de continuité de la FCD et présente une solution robuste au bruit. Nous supposons que la FCD est lisse : pour des pixels voisins et pour des distances proches, les corrections des points 3D à appliquer sont assez semblables. Dans notre cas, une régularisation 3D de nos données en $(X, Y$ et $Z)$ est appliquée. Une grande variété de termes de régularisation ou termes de lissage est proposée dans la littérature. Nous choisissons le même terme de lissage choisi dans notre première méthode, quadratique d'ordre 2, appelée énergie de courbure ("*bending energy*") définie comme suit :

$$C_s[F] \stackrel{\text{def}}{=} \int_{\psi} \left\| \frac{\partial^2 F}{\partial \mathbf{Q}^2} \right\|_2^2 d\mathbf{Q}. \quad (5.11)$$

Nous montrerons dans la suite que la fonction choisie pour modéliser la FCD possède la propriété d'introduire une transformation minimisant l'énergie de courbure. Cette énergie sera donc exprimée sous une autre forme propre à ce modèle qui sera détaillé dans la section 5.3.2.

5.2.4 Résolution itérative de la CNP

L'objectif est d'estimer la fonction F qui minimise C (équation (5.5)). Cela correspond à un problème d'optimisation non-convexe et non-linéaire où les deux inconnues \mathcal{P} et F doivent être estimées. En effet, les plans Π_i représentés par l'ensemble des vecteurs \mathcal{P} ne sont pas des données d'entrée de notre méthode.

Nous proposons un processus itératif qui alterne l'estimation de \mathcal{P} (ensemble des plans qui approximent au mieux les points 3D de chaque vue) et l'estimation de F par minimisation de la distance des points 3D aux plans \mathcal{P} . Les différentes

étapes du processus itératif sont résumées dans l'Algorithme 2.

| | |
|--------------------------------------------------------------------------|----------------------------------------------------|
| $k \leftarrow 0$ | % première itération |
| $F^k \leftarrow 0$ | % initialisation de F (0 est la fonction "zero") |
| repeat | |
| $\mathcal{P}^k \leftarrow \arg \min_{\mathcal{P}} C_d[F^k, \mathcal{P}]$ | % étape EP |
| $F^{k+1} \leftarrow \arg \min_{F \in L^2} C_d[F, \mathcal{P}^k]$ | % étape EF |
| $k \leftarrow k + 1$ | |
| until $F^{k+1} \approx F^k$ ou $k = k_{max}$ | % $k_{max} = 20$ (notre mise en œuvre) |

Algorithme 2 : Processus de Résolution itérative de la CNP.

- Au début, la fonction F de la CNP est initialisée à la fonction zéro (F^0 , correction nulle)
- Ensuite, F est mise à jour à chaque itération (k est le numéro de l'itération) se rapprochant ainsi à chaque itération de la solution.

Le processus peut être vu comme deux étapes de minimisation alternatives.

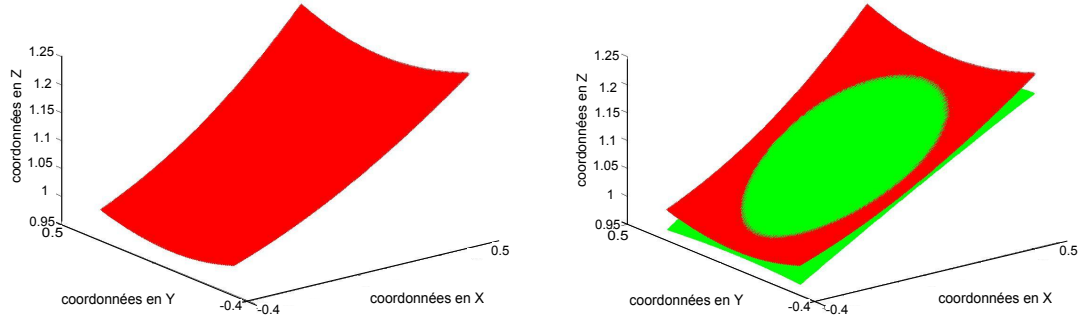
- **Estimation de \mathcal{P} (EP).** Cette étape consiste à fixer la fonction de CNP estimée à la $k^{\text{ème}}$ itération F^k afin d'estimer l'ensemble des plans \mathcal{P}^k . Ce sont les plans qui minimisent la fonction de coût $C_d[F^k, \mathcal{P}]$. Cela est réalisé comme suit :
 - Tout d'abord, les points corrigés $\{\mathbf{S}_{ij}\}_{j=1}^m$ par la fonction de correction F^k sont calculés. Un exemple d'une des vues du mur utilisée dans l'étalonnage est illustré par la figure 5.3. Les points 3D avant correction (corrigés avec la fonction de CNP zéro F^0) sont illustrés sur la figure 5.3a.
 - Ensuite, le plan Π_i qui approxime au mieux ces points par minimisation de la distance point-plan est estimé. Le plan approximé à partir des points 3D est illustré par la figure 5.3b.

Notons que cela doit être effectué pour l'ensemble des plans \mathcal{P}^k . Cette étape EP correspond à un problème d'optimisation de moindres carrés linéaires qui peut être résolu facilement.

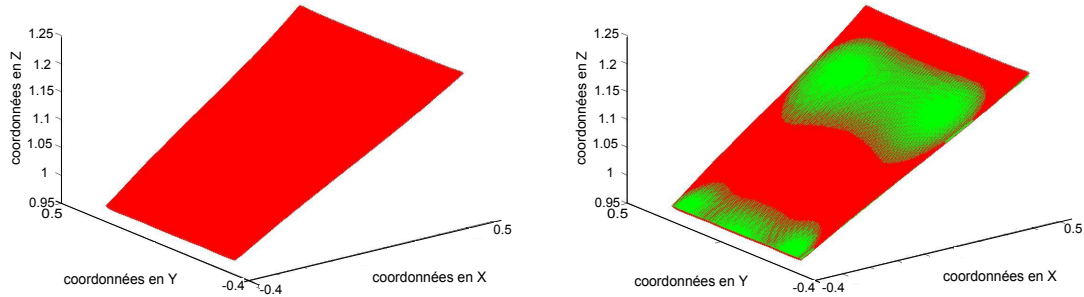
- **Estimation de F (EF).** Cette étape consiste à fixer l'ensemble des plans estimé \mathcal{P}^k afin d'estimer la fonction de CNP F^{k+1} par minimisation de la fonction de coût $C[F, \mathcal{P}^k]$. La figure 5.3c illustre les points 3D après correction par la fonction de correction F^1 estimée à la première itération. La figure 5.3d montre que la distance entre les points 3D après correction et le plan approximé a diminué.

L'étape EF dépend du modèle de correction choisi. Dans nos travaux, nous avons choisi d'utiliser la fonction Spline de type plaque mince de dimension 3 "*3D Thin-Plate Spline*" (TPS 3D) pour modéliser la fonction de correction. Dans la section suivante, nous introduisons tout d'abord ce modèle et ensuite

nous présentons une solution de l'étape EF basée sur ce modèle.



(a) Points 3D après correction par F^0 ($k = 0$). (b) *Estimation de Π_i* ($k = 0$). Surface constituée des points 3D (en rouge) et le plan Π_i (en vert) qui approxime au mieux ces points.



(c) Points 3D après correction par F^1 ($k = 1$).

(d) *Estimation de Π_i* ($k = 1$).

FIGURE 5.3 – Exemple d'une des vues du mur utilisée lors de l'étalonnage. Les résultats des deux premières itérations sont illustrés. La surface constituée des points 3D est présentée en rouge. Le plan Π_i qui approxime au mieux ces points est présenté en vert.

5.3 Modélisation et estimation de la FCD

Comme présenté précédemment, l'étape EF dépend du modèle de F (la fonction de correction de la mesure de profondeur) choisi. Contrairement à la première méthode d'étalonnage où un modèle non paramétrique a été employé, nous optons dans celle-ci pour un modèle paramétrique. Le choix d'un modèle paramétrique est

imposé par la grande quantité de données d'entrée. En effet, dans notre première méthode, pour chaque point utilisé dans l'étalonnage la vérité terrain associée est connue. L'utilisation d'un modèle non paramétrique était possible et a été privilégié. Cependant, pour notre deuxième méthode il n'y a pas de vérité terrain associée à chaque point. Nous avons donc besoin de plus de points qui se recoupent afin que le problème soit bien contraint. Une solution non paramétrique n'est donc pas possible : un grand nombre de données ne peut pas être facilement géré avec une modélisation non paramétrique.

Le modèle paramétrique doit vérifier les propriétés suivantes :

- Variation en fonction des trois composantes X , Y et Z ;
- Robustesse au bruit. Les données d'entrée de notre modèle sont les mesures issues de la caméra Temps-de-Vol qui présentent des discontinuités de profondeur et du bruit assez fort. Nous rappelons que nous évitons tout traitement de filtrage des données pour ne pas alourdir la mise en œuvre.

Les variations en fonction des trois composantes ne peuvent pas être bien modélisées par une simple fonction linéaire. Un modèle plus complexe est nécessaire dans notre cas. La fonction paramétrique Spline de type plaque mince de dimension 3 "*3D Thin-Plate Spline*" (TPS 3D) est alors choisie pour différentes raisons. Premièrement, elle vérifie toutes ces propriétés citées ci-dessus. En effet, c'est une fonction flexible connue pour être une bonne approximation pour des déformations complexes. De plus, elle est naturellement régularisée, minimisant l'énergie de courbure, ce qui présente une solution robuste pour les données aberrantes. Deuxièmement, nous montrerons qu'elle nous permet de contrôler l'empreinte mémoire.

Dans cette section, le modèle de fonction de type TPS 3D est tout d'abord présenté et ensuite la solution de l'estimation de F (étape EF du processus de Correction de la Non Planarité) est détaillée.

5.3.1 Modélisation de la FCD avec une fonction TPS 3D

Nous utilisons la fonction TPS 3D pour modéliser la fonction de CNP F . La fonction TPS 3D est présentée en détail dans la section B.2.3 de l'Annexe B. Elle est contrôlée par l centres 3D notés \mathbf{c}_k positionnés dans tout l'espace ψ et est paramétrée par un vecteur de $l + 4$ coefficients $\mathbf{h}^T = (\mathbf{w}^T \quad \mathbf{a}^T)$. Soit un point 3D noté \mathbf{Q} et $\ell_{\mathbf{Q}}^T$ le vecteur défini par :

$$\ell_{\mathbf{Q}}^T = ((d(\mathbf{Q}, \mathbf{c}_1)) \cdots (d(\mathbf{Q}, \mathbf{c}_l)) \quad \mathbf{q}^T \quad 1),$$

avec d est la distance euclidienne, la fonction TPS 3D au point \mathbf{Q} est donnée par :

$$F(\mathbf{Q}) \stackrel{\text{def}}{=} \ell_{\mathbf{Q}}^T \mathbf{h}. \quad (5.12)$$

La fonction de type TPS 3D nous permet de contrôler l’empreinte mémoire, en effet, seulement les $l + 4$ paramètres et les l centres doivent être sauvegardés. l est l’hyper-paramètre inconnu qui est inclus dans ce modèle et qui doit être estimé (voir section 5.4.1 pour l’estimation de ce paramètre). Le point corrigé \mathbf{S} (5.3) peut être exprimé, en introduisant notre modèle, par :

$$\mathbf{S} = \left(X + \frac{1}{Z}X\ell_{\mathbf{Q}}^T\mathbf{h} \quad Y + \frac{1}{Z}Y\ell_{\mathbf{Q}}^T\mathbf{h} \quad Z + \ell_{\mathbf{Q}}^T\mathbf{h} \right). \quad (5.13)$$

5.3.2 Estimation de la FCD

Nous réécrivons tout d’abord notre système composé du terme d’attache aux données (5.8) et du terme de régularisation (5.11) en intégrant notre modèle de type TPS 3D. Ensuite, pour le résoudre nous l’exprimons sous une forme matricielle. Nous détaillons dans ce qui suit ce passage pour chacun des termes.

Terme d’attache aux données. En remplaçant le terme \mathbf{S} qui représente le point corrigé par son expression qui introduit notre modèle de type TPS 3D (5.13), le terme d’attache aux données (5.8) peut être réécrit comme suit :

$$C_d(\mathbf{h}, \mathcal{P}) = \sum_{i=1}^n \sum_{j=1}^m \left(\frac{\frac{1}{Z_{ij}}\mathbf{P}_i^T\mathbf{Q}_{ij}\ell_{\mathbf{Q}_{ij}}^T}{\|\mathbf{P}_i\|_2}\mathbf{h} + \frac{\mathbf{P}_i^T\mathbf{Q}_{ij} + 1}{\|\mathbf{P}_i\|_2} \right)^2 \quad (5.14)$$

ou sous la forme matricielle suivante :

$$C_d(\mathbf{h}, \mathcal{P}) = \|\mathbf{D}\mathbf{h} - \mathbf{b}\|_2^2, \quad (5.15)$$

avec :

- $\mathbf{D} \in \mathbb{R}^{nm \times (l+4)}$, nm nombre des points 3D utilisés dans l’étalonnage et $l + 4$ nombre des inconnus (paramètres de la fonction TPS 3D).
- $\mathbf{b} \in \mathbb{R}^{nm \times 1}$: le vecteur contenant les scalaires $-\frac{\mathbf{P}_i^T\mathbf{Q}_{ij} + 1}{\|\mathbf{P}_i\|_2}$.
- $\mathbf{h} \in \mathbb{R}^{(l+4) \times 1}$: le vecteur contenant les inconnus, c’est-à-dire les $l + 4$ paramètres de la fonction TPS 3D.

Notre terme de donnée C_d décrit par l’équation (5.15) intègre le modèle d’une fonction TPS 3D. Comme expliqué dans la section B.2.3 de l’Annexe B, pour assurer l’existence d’une solution à un tel système, il y a quatre conditions aux bords des fonctions TPS 3D qui doivent être vérifiées. Ces conditions sont exprimées par :

$$\check{\mathbf{P}}^T\mathbf{w} = \mathbf{0},$$

où la $k^{\text{ème}}$ ligne de $\check{\mathbf{P}}^T$ est donnée par $(\mathbf{c}_k^T \ 1)$ et \mathbf{w} est le vecteur contenant les l premiers paramètres de la fonction TPS 3D ($\mathbf{h}^T = (\mathbf{w}^T \ \mathbf{a}^T)$).

Pour intégrer ces conditions dans notre système, nous remplaçons dans l'équation (5.15) \mathbf{h} par $\mathbf{V}_2\mathbf{h}_1$, avec :

- $\mathbf{h}_1 \in \mathbb{R}^l$ le nouveau vecteur des inconnus à estimer. Notons que le vecteur contenant des paramètres de la fonction TPS sera ensuite déduit de \mathbf{h}_1 .
- $\mathbf{V}_2 \in \mathbb{R}^{(l+4) \times l}$ est la matrice qui satisfait les conditions aux limites. C'est-à-dire, $(\check{\mathbf{P}}^\top \ \mathbf{0})\mathbf{V}_2 = \mathbf{0}$. Cette matrice \mathbf{V}_2 peut être déterminée simplement par la méthode de décomposition en valeurs singulières ("SVD") (plus de détails dans la section B.1.2.1 de l'Annexe B). Elle correspond aux l vecteurs de la base de l'espace nul de $(\check{\mathbf{P}}^\top \ \mathbf{0})$.

Le terme de donnée C_d devient alors :

$$C_d(\mathbf{h}, \mathcal{P}) = \|\mathbf{D}\mathbf{V}_2\mathbf{h}_1 - \mathbf{b}\|_2^2 \quad (5.16)$$

Terme de régularisation. Le terme de régularisation est défini (pour la fonction TPS) comme suit :

$$C_s(\mathbf{h}) \stackrel{\text{def}}{=} \mathbf{w}^\top \mathbf{K} \mathbf{w}, \quad (5.17)$$

avec \mathbf{K} la matrice de coefficients calculant la distance entre les centres de la fonction TPS 3D : $\mathbf{K}_{i,j} \stackrel{\text{def}}{=} d(c_i, c_j)$. Pour réécrire le terme de régularisation C_s sous la même forme que le terme d'attache aux données C_d , c'est-à-dire une norme matricielle dépendante en \mathbf{h}_1 , la matrice \mathbf{K} peut être exprimée comme $\mathbf{K} = \mathbf{B}^\top \mathbf{B}$ avec $\mathbf{B} \in \mathbb{R}^{l \times l}$. Nous montrons dans la suite, que dans la mise en œuvre, nous calculons directement \mathbf{K} , nous n'avons pas besoin d'estimer la matrice \mathbf{B} . La matrice \mathbf{V}_2 peut être décomposée en deux matrices : $\mathbf{V}_3 \in \mathbb{R}^{l \times l}$ correspond à ses l premières lignes et $\mathbf{V}_4 \in \mathbb{R}^{4 \times l}$ ses quatre dernières lignes :

$$\mathbf{h} = \begin{pmatrix} \mathbf{w} \\ \mathbf{a} \end{pmatrix} = \mathbf{V}_2\mathbf{h}_1 = \begin{pmatrix} \mathbf{V}_3 \\ \mathbf{V}_4 \end{pmatrix} \mathbf{h}_1. \quad (5.18)$$

Le vecteur \mathbf{w} peut alors être réécrit comme suit :

$$\mathbf{w} = \mathbf{V}_3\mathbf{h}_1, \quad (5.19)$$

Ainsi le terme de régularisation C_s (5.17) peut être réécrit comme norme matricielle sous la forme suivante :

$$C_s(\mathbf{h}) = \|\mathbf{B}\mathbf{V}_3\mathbf{h}_1\|_2^2 \quad (5.20)$$

Fonction de coût. En combinant les équations (5.16) et (5.20), la fonctionnelle de coût (5.5) peut s'écrire sous la forme suivante :

$$C(\mathbf{h}, \mathcal{P}) = \left\| \begin{pmatrix} \mathbf{D}\mathbf{V}_2 \\ \sqrt{\lambda}\mathbf{B}\mathbf{V}_3 \end{pmatrix} \mathbf{h}_1 - \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix} \right\|_2^2. \quad (5.21)$$

Résolution de la EF. le système décrit par l'équation (5.21) peut être résolu par la technique de la pseudo-inverse (plus de détails dans la section B.1.2.2 de l'Annexe B). La solution est donnée par :

$$\mathbf{h}_1 = (\mathbf{V}_2^T \mathbf{D}^T \mathbf{D} \mathbf{V}_2 + \lambda \mathbf{V}_3^T \mathbf{B}^T \mathbf{B} \mathbf{V}_3)^{-1} \mathbf{V}_2^T \mathbf{D}^T \mathbf{b}. \quad (5.22)$$

Notons que la taille de la matrice \mathbf{D} est très grande. Nous ne pouvons pas sauvegarder en mémoire \mathbf{D}^T et \mathbf{D} . En effet, un très grand nombre de points 3D est utilisé ($n \times m = 46 \times (204 \times 204)$) plus de détails dans la section 5.4.2). Par conséquent dans notre mise en œuvre nous calculons directement les coefficients de \mathbf{T} un par un :

$$\mathbf{T} = \mathbf{D}^T \mathbf{D}, \quad \mathbf{T} \in \mathbb{R}^{(l+4) \times (l+4)} \quad (5.23)$$

sans formation de la matrice \mathbf{D} et de \mathbf{D}^T . Comme présenté précédemment, la matrice \mathbf{B} est utilisée seulement pour exprimer le terme de régularisation C_s sous forme d'une norme matricielle, dans notre mise en œuvre, cette matrice n'est pas calculé, la matrice $\mathbf{K} \in \mathbb{R}^{l \times l}$ est directement utilisée. Ainsi, $\mathbf{D}^T \mathbf{D}$ est remplacée par \mathbf{T} et $\mathbf{B}^T \mathbf{B}$ est remplacée par \mathbf{K} dans l'équation (5.22), la solution de notre système est donnée par :

$$\mathbf{h}_1 = (\mathbf{V}_2^T \mathbf{T} \mathbf{V}_2 + \lambda \mathbf{V}_3^T \mathbf{K} \mathbf{V}_3)^{-1} \mathbf{V}_2^T \mathbf{D}^T \mathbf{b}. \quad (5.24)$$

Rappelons que les paramètres de la fonction TPS 3D \mathbf{h} sont obtenus par :

$$\mathbf{h} = \mathbf{V}_2 \mathbf{h}_1. \quad (5.25)$$

5.4 Résultats expérimentaux

Une première évaluation de notre méthode sur des données de synthèse est proposée. La distorsion de la mesure de profondeur est simulée. Cette expérience avec les données de synthèse nous permet de montrer les différentes étapes de l'étalonnage et d'évaluer sa précision. Les résultats de la CNP au cours de plusieurs itérations sont présentés. Cette expérience nous permet également de valider l'étape de Correction Affine (CA) puisque nous disposons de la vérité terrain associée à chaque point. Nous rappelons que le modèle théorique de la FCD n'est pas connu. Néanmoins, ses caractéristiques ont été définies empiriquement comme suit :

- Variation en fonction de la profondeur d ;
- Variation en fonction de la position du pixel ($u \ v$) dans l'image ;
- Continuité.

Nous avons alors généré les données de simulation à partir d'une fonction de distorsion estimée en nous basant sur ces caractéristiques. Les résultats de cette première expérience sont discutés dans la section 5.4.1.

Afin de compléter notre évaluation, une seconde expérience sur des données réelles est présentée. Deux mesures sont utilisées pour évaluer la précision de la correction :

- La mesure de planarité des surfaces après correction. Cette mesure nous permet d'évaluer la précision de la correction de la non planarité ;
- La mesure de l'erreur par rapport à une vérité terrain. Les données de validation de notre première méthode dont la vérité terrain associée est acquise selon la procédure présentée dans la section 4.3.1.1 sont utilisées. Ces données sont utilisées pour se comparer à notre première méthode ainsi qu'aux deux méthodes principales de la littérature.

Les résultats de cette deuxième expérience sont discutés dans la section 5.4.2.

5.4.1 Expériences sur des données simulées

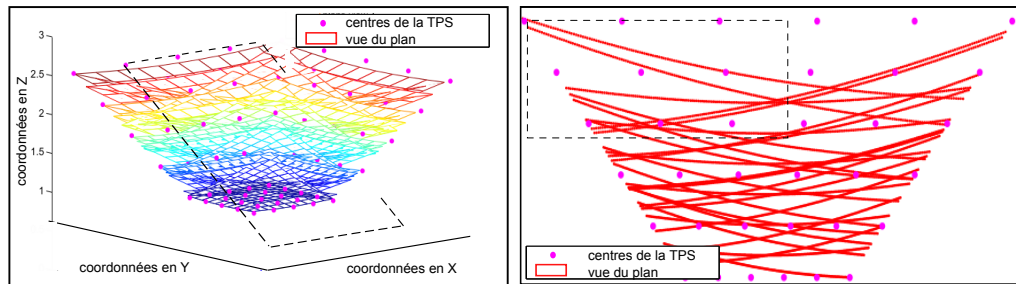
Protocole expérimental. L'ensemble des données utilisées dans cette expérience correspond à différentes vues d'un plan qui s'intersectent pour couvrir tout l'espace 3D d'étalonnage. Il est divisé en trois parties : données d'entraînement, de test et de validation. L'ensemble d'entraînement est utilisé pour effectuer différents étalonnages (estimation de la Correction de la Non Planarité (CNP) et de la Correction Affine (CA)) en faisant varier l'ensemble des hyper-paramètres (l, λ) . Ces étalonnages sont ensuite évalués sur l'ensemble des données de test afin de déterminer les valeurs optimales de (l, λ) (qui minimise le RMS² de la distance d'un point à un plan). Le résultat final utilisé pour évaluer la méthode proposée est obtenu avec l'ensemble des données de validation utilisant l'ensemble optimal de (l, λ) .

Comme présenté précédemment, les données simulées sont obtenues en distordant un plan avec une fonction de distorsion qui vérifie trois caractéristiques :

- Elle varie en fonction de la profondeur et augmente du centre de l'image vers les bords ;
- Elle varie en fonction de la distance. Les données simulées couvrent un intervalle de distance de 1m à 3m. Un ensemble d'entraînement de 36 vues d'un plan est utilisé (figure 5.4a) ;
- Elle vérifie la propriété de continuité.

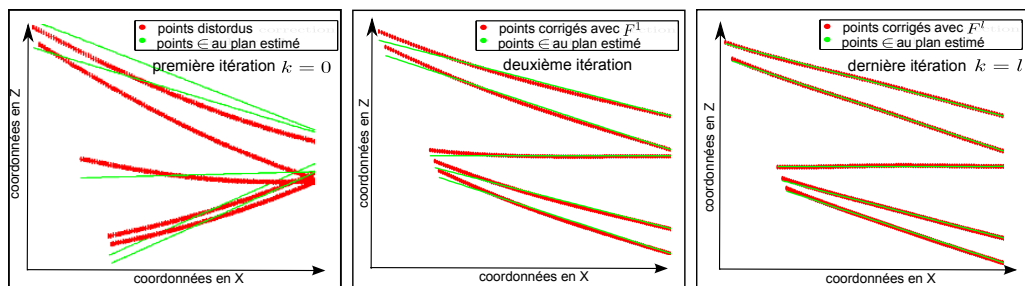
Notons que les ensembles des données de test et de validation contiennent chacun 10 vues d'un plan.

2. La valeur RMS : (Root Mean Square) pour une série de valeurs discrètes est la racine carré de la moyenne du carré de ces valeurs.



(a) Données d'entraînement tracées ensemble avec les centres de la fonction TPS ($l = 5^3$).

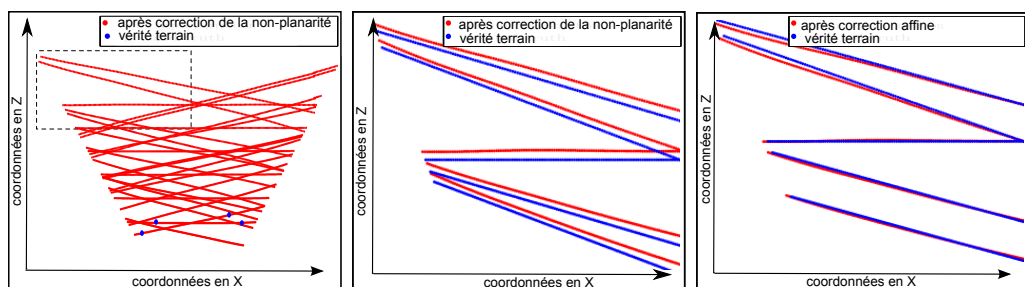
(b) Section de (a).



(c) CNP : $k = 0$

(d) CNP : $k = 1$

(e) CNP : $k = l$



(f) Après CNP

(g) Avant CA

(h) Après CA

FIGURE 5.4 – Résultats avec les données simulées durant le processus d'étalonnage de la mesure de profondeur. (a) Une partie de l'espace d'étalonnage allant de 1 m à 2,5 m. Une section de (b) est considérée pour montrer les résultats obtenus à (c) la première itération, (d) la seconde itération et (e) la dernière itération de l'étape de CNP. (f) Données d'entraînement obtenues après l'étape de CNP et les points utilisés pour estimer A . Comparaison des résultats (g) avant et (h) après l'étape de CA.

Étalonnage. Les étapes d'étalonnage sont présentées sur la figure 5.4. À l'initialisation ($k = 0$), la distance d'un point à un plan est grande (figure 5.4c), elle

diminue à l'itération suivante (figure 5.4d) et devient très faible à la dernière itération (figure 5.4e). Après l'étape de correction de la non-planarité (CNP), les points sont coplanaires mais non alignés avec la vérité terrain (figure 5.4g). Ensuite, l'étape de correction affine (CA) est effectuée. Une telle transformation est choisie car les transformations affines forment le noyau de l'énergie de courbure. Elle conserve alors la courbure obtenue avec la CNP. La transformation affine 3D de la CA peut être calculée à partir de 4 correspondances de points des données après la CNP et la vérité terrain. Nous avons utilisé 8 correspondances de points pour que la CA soit plus précise. Nous avons choisis des points à des distances proches comme le montre la figure 5.4f. Cela nous permet d'évaluer la précision de la CA avec un tel choix de disposition des points. En effet, pour un étalonnage avec des données réelles un tel choix nous permet de garantir une meilleure précision de la vérité terrain. Cela facilite aussi la mise en œuvre. Nous n'avons pas besoin de calculer la vérité terrain pour les différents pas de distances. La figure 5.4h montre qu'après la CA les points sont très proches des points de la vérité terrain.

Ensemble optimal des hyper-paramètres (l, λ) . Nous voulons déterminer les valeurs optimales de (l, λ) . Différents étalonnages sont effectués en faisant varier (l, λ) et ensuite testés. Les résultats présentés par le tableau 5.1 représentent l'erreur RMS calculée après la CNP. Ils montrent que, à partir de $l = 5^3$ (5 centres suivant $X-$, $Y-$ et $Z-$), le système est stable pour un large intervalle de λ . Notons que l doit être choisi aussi petit que possible pour limiter le temps de calcul, sans dégrader la qualité des résultats. Pour cette raison, $l = 5^3$ est choisi, puisque c'est la plus petite valeur qui donne un faible RMS. Maintenant, l est fixé, nous voulons déterminer la valeur optimale de λ . Nous proposons d'utiliser la technique de "*golden section search*". La valeur optimale de λ calculée avec cette technique est égale à 0,12 (RMS = 1,365 mm avec les données de test).

Résultats. La méthode proposée est évaluée sur l'ensemble des données de validation. La figure 5.5 montre un exemple de vues fronto-parallèles d'un plan que nous avons choisi pour illustrer les résultats. Cette vue est située à une distance de 1,20 m de la caméra. La distorsion augmente du centre de l'image vers les bords (la distorsion maximale est 60 mm). Après la CNP, un plan parfait est obtenu (figure 5.5c) : tous les points sont coplanaires et situés à 1,22 m. Cependant, ils ne sont pas alignés avec les points de la vérité terrain (1,20 m). La figure 5.5d montre qu'après la CA le plan corrigé final est très proche des données de vérité terrain. Pour évaluer la méthode proposée sur tous les pixels de l'ensemble des données de validation, nous comparons la distance d'un point à un plan avant et après cor-

| | | l | | | | | |
|-----------|-----------|-------|-------|--------------|-------|-------|-------|
| | | 3^3 | 4^3 | 5^3 | 6^3 | 7^3 | 8^3 |
| λ | 10^{-4} | 2,429 | 2,187 | 1,369 | 1,383 | 1,428 | 1,497 |
| | 10^{-3} | 2,429 | 2,187 | 1,369 | 1,383 | 1,427 | 1,352 |
| | 10^{-2} | 2,429 | 2,187 | 1,369 | 1,384 | 1,330 | 1,313 |
| | 10^{-1} | 2,429 | 2,187 | 1,369 | 1,397 | 1,273 | 1,655 |
| | 10^0 | 2,430 | 2,187 | 1,366 | 1,415 | 6,199 | N |
| | 10^1 | 2,395 | 2,675 | 8,563 | N | 9,550 | N |
| | 10^2 | 4,843 | N | 10,096 | N | N | N |
| | 10^3 | N | N | N | N | N | N |
| | 10^4 | N | N | N | N | N | N |

Tableau 5.1 – Erreur RMS calculée après CNP en faisant varier les valeurs des hyper-paramètres : l allant de 3^5 à 8^3 et λ de 10^{-4} à 10^4 en échelle logarithmique. Les cas où le système ne converge pas sont représentés par N. A partir de $l = 5^3$ (5 centres suivant $X-$, $Y-$ et $Z-$), le système est stable pour un large intervalle de λ .

rection. Pour la validation de la CNP, la distance d'un point à un plan $d^2(S, \Pi_S)$ (Π_S est le plan qui approxime au mieux les points S) est considérée, tandis que pour la validation de la CA, la distance d'un point au plan de la vérité terrain $d^2(S, \Pi_{GT})$ (Π_{GT} est le plan de la vérité terrain) est utilisée. Les résultats sont présentés dans le tableau 5.2. Nous observons une importante amélioration. Après la CNP, l'erreur RMS est égal à 1,47 mm, tandis qu'elle est égale à 15,64 mm avant correction. Après la CA, l'erreur RMS est égale à 2,78 mm, tandis qu'elle est égale à 24,41 mm avant correction. Nous observons une amélioration de la précision des images de profondeur avec l'approche proposée.

| | avant CNP | après CNP |
|--------------------|-----------|-----------|
| $d^2(S, \Pi_S)$ | 15,64 mm | 1,47 mm |
| | avant CA | après CA |
| $d^2(S, \Pi_{GT})$ | 24,41 mm | 2,78 mm |

Tableau 5.2 – Erreur RMS (mm).

5.4.2 Expériences sur des données réelles

Protocole expérimental. Comme dans la première expérience, l'ensemble des données utilisées correspond à différentes vues d'un plan qui s'intersectent pour

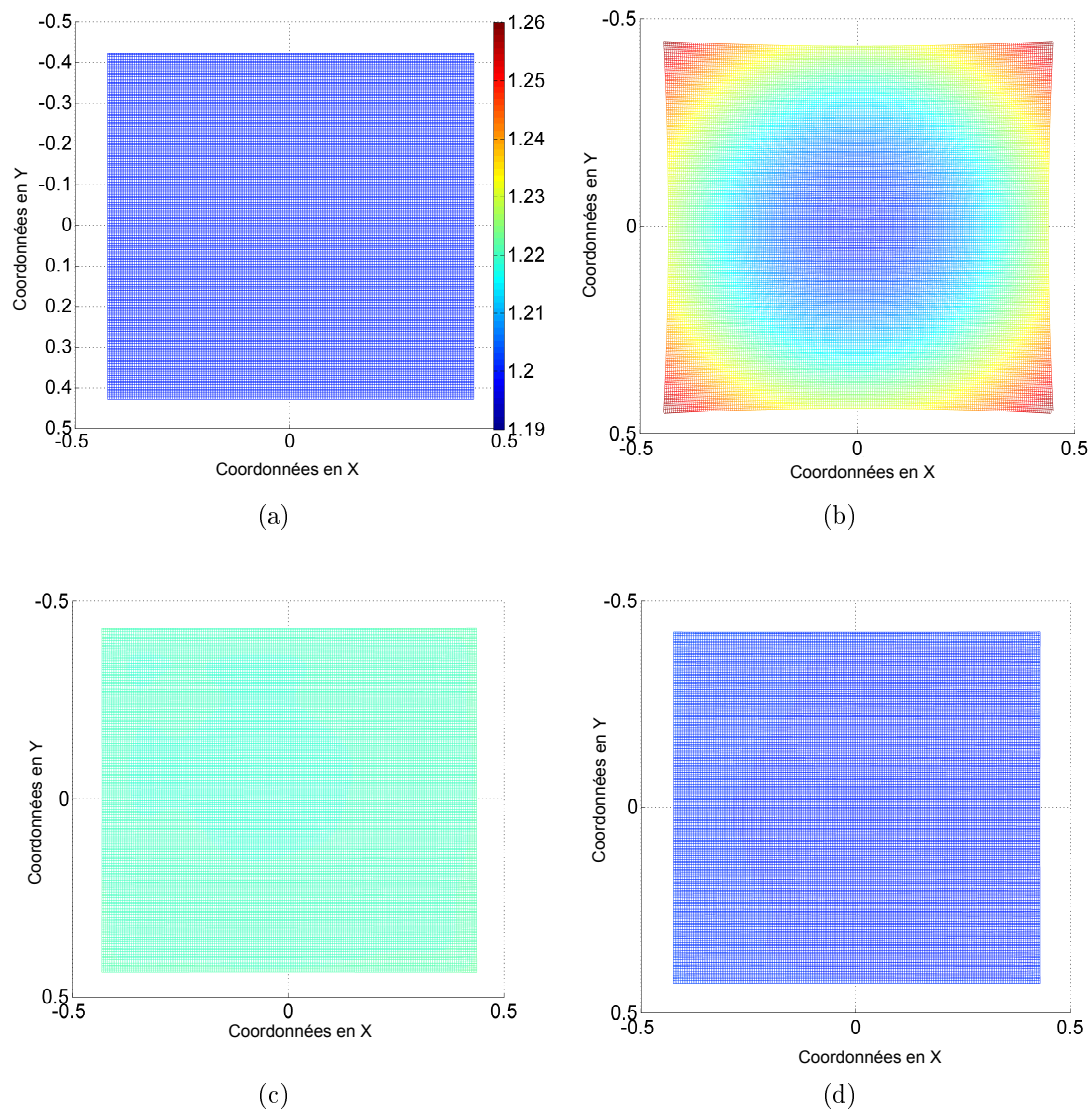


FIGURE 5.5 – Une vue fronto-parallèle d’un plan (a) non distordu (vérité terrain) (b) avant correction, (c) après CNP et (d) après CA.

couvrir tout l’espace 3D d’étalonnage et est divisé en trois parties : des données d’entraînement, de test et de validation.

Ces données sont obtenues en filmant un mur plan sans aucune contrainte d’orientation. Notons que, dans nos travaux, l’extraction des plans à partir des images est effectuée manuellement. Cela pourrait être amélioré par une détection et extraction automatique des plans. La caméra Temps-de-Vol utilisée est la PMD

CamCube2 avec une résolution de 204×204 pixels et l'objet observé est un mur blanc. Les données couvrent une plage de distance approximative de 3 m à 6,5 m. Notons que nous nous limitons à cet intervalle de profondeur pour éviter de représenter les mesures de profondeur aux bords qui sont invalides. Un ensemble d'entraînement de 46 vues d'un plan est utilisé. Nous avons utilisé 8 correspondances de points entre la vérité terrain et les images du capteur pour que l'estimation de la CA soit plus précise. Ces correspondances de points sont extraites des données de la première méthode d'étalonnage (section 4.3.1.1). Ils sont choisis à de courtes distances pour garantir une meilleure précision et simplifier la mise en œuvre comme expliqué précédemment (figure 5.4f). L'ensemble des données de test et de validation contient chacun 10 vues d'un plan.

Nous utilisons également les données de validation de notre première méthode pour lesquelles la vérité terrain associée a été calculée (voir section 4.3.1). Ces données sont utilisées pour comparer la méthode proposée avec notre première méthode et aux deux méthodes principales de la littérature. Elles correspondent à différentes vues fronto-parallèle d'un mur blanc.

Résultats. La méthode proposée est évaluée sur l'ensemble des données réelles de validation. Les valeurs optimales de l, λ sont respectivement 5^3 et 0,1. Ces valeurs sont déterminées par la technique de recherche de minimum par "*golden section search*" présentée précédemment.

Une comparaison de la distance d'un point à un plan avant et après l'étape de CNP est effectuée pour évaluer la précision de cette étape pour tous les pixels. La figure 5.6 montre la distribution de la distance point-plan pour tous les points des images de validation avant et après correction. Nous observons une distribution normale sur les deux histogrammes avec une moyenne égale à zéro. Avant correction, l'écart type (ou RMS) est égal à 11,5 mm, tandis qu'il est égale à 8,45 (ou RMS) après correction. L'histogramme après correction présente un pic plus grand en zéro. Nous estimons que le RMS résiduel après correction correspond au bruit du capteur. En effet, une évaluation du bruit du capteur est effectuée et a été présentée dans le chapitre 3. L'écart-type d'un mur blanc est calculé à partir de 100 mesures en chaque pixel. Il varie approximativement de 0 mm à 30 mm avec une valeur moyenne égale à 12,5 mm (voir figure 3.4). La figure 5.7 montre un exemple d'une vue d'un mur blanc. La distance point-plan $d^2(S, \Pi_S)$ (Π_S est le plan qui approxime au mieux les points) de chaque point 3D est représentée par la couleur. Avant l'étape de CNP, les points 3D ne sont pas coplanaires : la distance point-plan augmente du centre de l'image vers les bords. Après l'étape de correction de la non-planarité (CNP), les points 3D sont coplanaires et la distance point-plan est homogène sur toute l'image. En conclusion, nous pouvons donc dire,

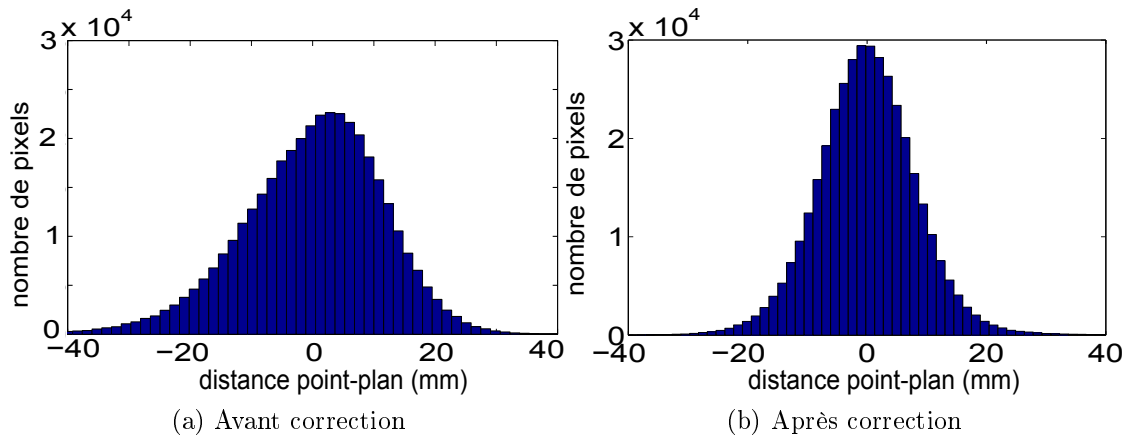


FIGURE 5.6 – Comparaison de la distribution de la distance point-plan pour tous les points des images de validation avant et après correction.

qu'en dépit du bruit du capteur, de bons résultats sont obtenus.

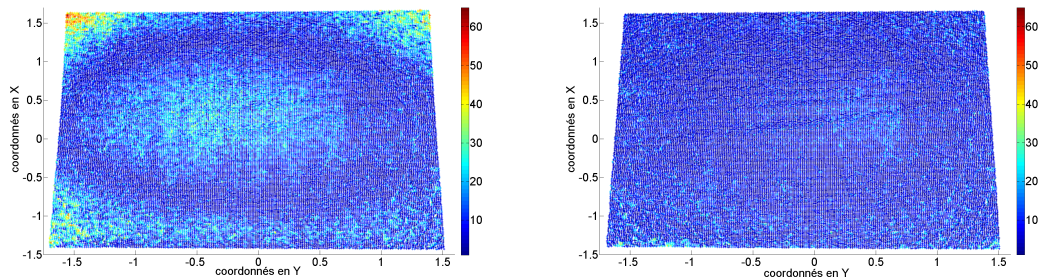


FIGURE 5.7 – Une vue d'un mur blanc (à gauche) avant la correction de la distorsion de la mesure de profondeur et (à droite) après l'étape de CNP. La couleur représente la distance d'un point à un plan mesurée en mm.

Comparaison. Les résultats présentés dans la section précédente nous permettent de valider seulement l'étape de CNP de notre méthode. Pour évaluer notre méthode (composée des deux étapes CNP et CA), nous utilisons l'ensemble des données de validation de notre première méthode car nous disposons de la vérité terrain pour ces données. Nous n'avons pas utilisé cet ensemble de données dans la section précédente parce qu'il ne contient pas beaucoup de mesures, comparée à l'ensemble de données de validation utilisé dans la section précédente et ne contient pas d'intersection de points. Il correspond à différentes vues frontoparallèle d'un mur blanc. La valeur médiane de l'erreur résiduelle de tous les pixels

de ces images est calculée en comparant les mesures de profondeur après correction (de la distorsion de la mesure de profondeur) à la vérité terrain. Le résultat est présenté dans le tableau 5.3. Nous reprenons dans ce tableau les résultats obtenus après correction avec notre première méthode ainsi les deux méthodes de la littérature [Lindner & Kolb 2006] et [Kahlmann *et al.* 2006] présentés précédemment dans le tableau 4.1. L'erreur médiane résiduelle après la correction de la profondeur avec notre deuxième méthode est égale à 6,99 mm. En la comparant avec la méthode [Kahlmann *et al.* 2006] (erreur résiduelle égale à 8,49 mm), nous observons une amélioration de la précision. Après correction avec la méthode [Lindner & Kolb 2006], l'erreur résiduelle est égale à 6,39 mm. Nous pouvons dire que les résultats obtenus avec cette méthode sont assez proches de ceux obtenus avec la méthode que nous proposons. En comparant nos deux méthodes, nous constatons qu'une meilleure précision est obtenue avec notre première méthode : l'erreur médiane résiduelle est égale à 4,42 mm. Avec notre deuxième méthode, nous perdons en précision. Cependant, l'étalonnage est beaucoup plus simple et rapide à mettre en oeuvre comparée à notre première méthode et aux méthodes de la littérature. En effet, au lieu d'utiliser des correspondances de points entre la vérité terrain et les images du capteur pour chaque pixel et pour chaque pas de distance, nous utilisons seulement 8 correspondances de points. Notons aussi qu'on est dans l'incertitude de la vérité terrain.

Tableau 5.3 – Comparaison avec la première méthode et deux approches de la littérature.

| | Erreurs de profondeur (mm) | |
|-------------------------------|----------------------------|----------------|
| | Médiane | Moyenne |
| Avant étalonnage | 35,03 mm | 34,30 mm |
| Méthode proposée | 6,99 mm | 8,18 mm |
| Notre première méthode | 4,42 mm | 6,16 mm |
| [Lindner & Kolb 2006] | 6,39 mm | 8,70 mm |
| [Kahlmann <i>et al.</i> 2006] | 8,49 mm | 10,83 mm |

En comparant la correction de la non-planarité des points 3D obtenus après correction avec nos deux méthodes, nous observons une meilleure correction avec notre deuxième méthode. Les résultats sont présentés dans le tableau 5.4. La distance d'un point à un plan Π_S (Π_S est le plan qui approxime au mieux les points) est considérée pour comparer la correction de la non-planarité. Le RMS restant après étalonnage avec notre deuxième méthode est égal à 7,25 mm, tandis qu'il est égal à 7,82 mm avec la première (RMS avant étalonnage est égal à 9,25 mm).

Nous proposons dans le chapitre 6, une façon pour évaluer la qualité de la méthode d'étalonnage proposée sans dépendre de la vérité terrain. Nous observons un objet de dimensions connues et comparons ses dimensions mesurées avec la caméra Temps-de-Vol avant et après correction de la mesure de profondeur.

Tableau 5.4 – Comparaison de la correction de la planarité avec la première méthode proposée.

| | $d^2(S, \Pi_S)$ |
|------------------------|-----------------|
| Avant étalonnage | 9,24 |
| Méthode proposée | 7.25 |
| Notre première méthode | 7,82 |

5.5 Conclusion

Nous avons présenté une méthode d'étalonnage de la mesure de profondeur basée sur la correction de la non planarité qui nécessite un grand ensemble des vues d'un plan facile à acquérir et une correction affine qui nécessite un faible ensemble de correspondances de points entre la vérité terrain et les images du capteur. Une fonction TPS 3D est utilisée pour modéliser la fonction de correction de la mesure de profondeur. Ce type de modèle est utilisé à notre connaissance, pour la première fois dans ce contexte. La méthode proposée a l'avantage d'être facile à utiliser comparée aux autres méthodes de la littérature. En effet, elle ne requiert pas un grand nombre de correspondances de points entre la vérité terrain et les images du capteur (4 points au minimum). Les résultats expérimentaux sur des données de simulation ainsi que sur des données réelles démontrent la validité et la précision de l'approche proposée. Une comparaison avec notre première méthode et avec les deux méthodes principales de la littérature montre qu'en plus de sa simplicité de mise en œuvre, la méthode proposée garantit une bonne qualité de correction. La méthode proposée pourrait être combinée avec une détection et extraction automatique des plans à partir des images. Une autre amélioration serait de prendre en compte les autres facteurs d'erreur. Notre approche peut être étendue facilement comme notre première méthode en adaptant l'étalonnage de la mesure de profondeur en fonction du TI et de la variation de la réflectivité comme expliqué dans le chapitre précédent.

CHAPITRE 6

Application : mesure 3D et incertitude associée

Dans les chapitres précédents, nous nous sommes intéressés à deux limites principales des caméras Temps-de-Vol : l'incertitude des mesures de profondeur due au bruit stochastique du capteur et la distorsion de ces mesures due à l'erreur systématique. Nous avons présenté deux modèles de bruit (chapitre 3) et deux méthodes d'étalonnage de la mesure de profondeur (chapitres 4 et 5).

L'objectif de ce chapitre est de présenter une application illustrant l'intérêt de nos travaux. Comme présenté dans le chapitre 4, notre méthode de calcul de la vérité terrain (voir section 4.3.1.1) ne peut pas garantir une précision millimétrique de cette dernière. Nous proposons donc, dans ce chapitre, une manière de comparer nos deux méthodes d'étalonnage aux principales méthodes de l'état de l'art sans dépendre de la vérité terrain. Nous utilisons un objet étalon (de dimensions connues) et comparons ses dimensions mesurées avec la caméra Temps-de-Vol avant et après correction de la mesure de profondeur. Nous estimons également l'incertitude de ces mesures calculée par propagation de l'erreur à travers les différentes formules de mesure en se basant sur notre modèle de bruit position-profondeur-TI (présenté dans la section 3.3.2). Ce modèle est utilisé car, comme cela a été montré dans le chapitre 3, il est plus précis pour une réflectivité constante.

6.1 Principe

Le processus de calcul d'une mesure 3D (*e.g.* mesure de distance entre deux points 3D) basé sur des mesures de profondeur d'une caméra Temps-de-Vol se fait en quatre étapes :

- Détection et extraction de deux extrémités 2D dans une image ;
- Extraction de la profondeur associée à chaque point 2D dans la carte de profondeur. La combinaison des coordonnées pixéliques des points 2D et de leurs profondeurs nous donne les points 2.5D ;
- Passage en coordonnées 3D ;
- Calcul de la distance entre les deux points 3D.

Les données utilisées seront corrigées avec nos deux méthodes d'étalonnage. La correction peut intervenir soit après le calcul des points 2.5D si nous utilisons notre première méthode d'étalonnage (chapitre 4) soit après le calcul des points 3D si nous utilisons notre deuxième méthode (chapitre 5).

La première étape, qu'elle soit manuelle ou automatique, ne peut pas être effectuée sans perte de précision. En plus de l'erreur de positionnement manuel du point 2D, la profondeur associée (point 2.5D) extraite d'une carte de profondeur d'une caméra Temps-de-Vol est sujette au bruit du capteur. La correction de la distorsion de la mesure de profondeur est sujette aux erreurs d'étalonnage. Par conséquent, toute mesure 3D (*e.g.* mesure de distance entre deux point 3D) basée sur des mesures de profondeur d'une caméra Temps-de-Vol est entachée d'erreur. Notre objectif est d'estimer l'incertitude de la mesure 3D finale (mesure de distance). Comme exemple de mesures 3D, nous avons choisi des mesures géométriques telles que la longueur et la largeur d'un objet. Pour obtenir l'incertitude de ces mesures, il faut donc intégrer et propager l'ensemble des incertitudes à travers les différentes transformations du processus de mesure permettant de passer des points 2D à leur distance 3D. Nous utilisons la méthode analytique de propagation d'erreur au premier ordre appelée dans la section B.2.1. Nous nous basons sur notre deuxième modèle de bruit (position-profondeur-*TI*) proposé dans la section 3.3.2). Ce modèle est utilisé car la réflectivité des points utilisés dans cette application est constante et il a été montré dans le chapitre 3 que ce modèle est plus précis dans ce cas. A partir des points cliqués ou détectés dans l'image, nous voulons estimer les mesures géométriques et leurs incertitudes par propagation d'erreur du point 2D (pixel dans l'image) noté $\mathbf{p} = (u \ v)$ aux mesures géométriques. Nous procédons en trois étapes : la première est la propagation d'erreur du point 2D (le pixel \mathbf{p}) au point 2.5D (point $\mathbf{q} = (u \ v \ d)$), la deuxième est la propagation d'erreur du point 2.5D \mathbf{q} au point 3D $\mathbf{Q} = (X \ Y \ Z)$ et la troisième est la propagation d'erreur des points 3D à la mesure géométrique finale (*i.e.* longueur et largeur d'un

objet dans notre cas).

6.1.1 Incertitude du point 2.5D

L'incertitude d'un pixel (point 2D \mathbf{p}) correspond à l'erreur de détection ou du clic dans l'image. Cette erreur est définie par la matrice de covariance en \mathbf{p} notée $\Sigma^{\mathbf{p}}$ (matrice de dimension 2×2). $\Sigma^{\mathbf{p}}$ dépend de la méthode d'extraction du point 2D. La méthode de calcul de cette matrice sort du propos de ce chapitre, nous supposons qu'elle est donnée. La transformation T_1 entre un pixel \mathbf{p} et un point 2.5D \mathbf{q} est définie par :

$$T_1 : \Gamma \rightarrow \Omega$$

$$\begin{pmatrix} u \\ v \end{pmatrix} \rightarrow \begin{pmatrix} u \\ v \\ d \end{pmatrix}, \quad (6.1)$$

où $d = d(u, v)$, $\Gamma = [u_{min}; u_{max}] \times [v_{min}; v_{max}]$ est défini par la résolution de l'image et $\Omega = [u_{min}; u_{max}] \times [v_{min}; v_{max}] \times [d_{min}; d_{max}]$, où $[d_{min}; d_{max}]$ est défini par la portée de la caméra (voir section 1.5.2). Une approximation au premier ordre de la matrice de covariance $\Sigma_{inter}^{\mathbf{q}}$ de \mathbf{q} est donnée par (voir l'équation (B.33)) :

$$\Sigma_{inter}^{\mathbf{q}} = J_{T_1} \Sigma^{\mathbf{p}} J_{T_1}^T, \quad (6.2)$$

où J est la matrice Jacobienne de dimension 3×2 de la fonction T_1 définie par :

$$J_{T_1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ d(u+1, v) - d(u, v) & d(u, v+1) - d(u, v) \end{pmatrix}. \quad (6.3)$$

Nous utilisons une approximation par différences finies avancée (ou à droite) des dérivées. Cependant, une dérivée à gauche ou centrée pourrait également être utilisée.

En plus de l'erreur du clic du point 2D, la mesure de profondeur associée est sujette à l'incertitude due au bruit du capteur. Comme il a été montré dans le chapitre 3, cette incertitude est définie par l'écart-type σ modélisé par la fonction TPS 3D f (voir équation (3.6)). L'intégration de la variance de la profondeur σ^2 dans la matrice de covariance $\Sigma_{inter}^{\mathbf{q}}$ donne :

$$\Sigma^{\mathbf{q}} = \Sigma_{inter}^{\mathbf{q}} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}. \quad (6.4)$$

6.1.2 Propagation de l'incertitude du 2.5D au 3D

Pour obtenir la position du point dans le repère cartésien, la transformation présentée dans la section B.3.1 est appliquée à la mesure 2.5D. Nous notons cette transformation T_2 :

$$T_2 : \quad \Omega \quad \rightarrow \quad \psi$$

$$\begin{pmatrix} u \\ v \\ d \end{pmatrix} \rightarrow \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad \text{avec} \quad \begin{cases} X = Z \frac{(u-c_u)d_u}{f_c} \\ Y = Z \frac{(v-c_v)d_v}{f_c} \\ Z = d \frac{f_c}{\sqrt{f_c^2 + ((u-c_u)d_u)^2 + ((v-c_v)d_v)^2}} \end{cases} \quad (6.5)$$

où $(c_u \ c_v)$ est le centre optique, f_c la distance focale, $(d_u \ d_v)$ la taille d'un pixel dans la direction u (resp. v), ψ est un sous ensemble de \mathbb{R}^3 : $\psi = [X_{min}; X_{max}] \times [Y_{min}; Y_{max}] \times [Z_{min}; Z_{max}]$. Pour obtenir l'incertitude exprimée dans le même repère cartésien, il est nécessaire de propager l'incertitude au travers de la transformation T_2 . Pour cela, nous proposons d'effectuer une approximation au premier ordre de la matrice de covariance Σ^Q de T_2 :

$$\Sigma^Q = J_{T_2} \Sigma^a J_{T_2}^T, \quad (6.6)$$

où J_{T_2} est la matrice Jacobienne de dimension 3×3 de la fonction T_2 définie par :

$$J_{T_2} = \begin{pmatrix} \frac{\partial T_{21}}{\partial u} & \frac{\partial T_{21}}{\partial v} & \frac{\partial T_{21}}{\partial d} \\ \frac{\partial T_{22}}{\partial u} & \frac{\partial T_{22}}{\partial v} & \frac{\partial T_{22}}{\partial d} \\ \frac{\partial T_{23}}{\partial u} & \frac{\partial T_{23}}{\partial v} & \frac{\partial T_{23}}{\partial d} \end{pmatrix}. \quad (6.7)$$

Nous avons présenté la propagation d'incertitude en considérant les mesures de profondeur dépourvues de distorsion. Nous présentons dans la section suivante comment la correction de cette distorsion peut être prise en compte dans l'estimation de l'incertitude.

6.1.3 Prise en compte de l'erreur d'étalonnage

Comme présenté précédemment, pour effectuer la mesure de distance de manière précise, on applique la fonction de correction de la distorsion (FCD). Il faut donc propager l'incertitude au travers de cette transformation. Nous souhaitons également comparer nos deux méthodes d'étalonnages. Nous allons donc corriger la distorsion de la mesure de profondeur avec les deux méthodes. Cependant, notre première méthode opère dans l'espace 2.5D alors que la deuxième opère dans l'espace 3D. Nous distinguons alors deux cas :

- Premier cas : utilisation de notre première méthode d'étalonnage. Dans ce cas, nous commençons par intégrer l'erreur d'étalonnage dans la matrice de covariance $\Sigma^{\mathbf{q}}$ du point 2.5D \mathbf{q} . Ensuite, nous propageons l'incertitude au point 3D ;
- Deuxième cas : utilisation de notre deuxième méthode d'étalonnage. Dans ce cas, nous commençons par propager l'incertitude au point 3D. Ensuite, nous intégrons l'erreur d'étalonnage dans la matrice de covariance $\Sigma^{\mathbf{Q}}$ du point 3D \mathbf{Q} .

Nous présentons dans ce qui suit les deux cas.

Premier cas : utilisation de notre première méthode. Dans ce cas, nous corrigeons les points 2.5D en utilisant notre première méthode d'étalonnage. Nous devons alors prendre en compte l'erreur de cet étalonnage dans le calcul de l'incertitude. Nous définissons la transformation entre un point 2.5D avant correction et après correction comme suit :

$$T_g : \Gamma \rightarrow \Omega$$

$$\begin{pmatrix} u \\ v \\ d \end{pmatrix} \rightarrow \begin{pmatrix} u \\ v \\ d + c \end{pmatrix}, \quad (6.8)$$

où $c = g \begin{pmatrix} u \\ v \\ d \end{pmatrix}$ est la correction estimée par la FCD g (voir équation (4.1)). La matrice de covariance $\Sigma^{\mathbf{q}}$ (équation (6.4)) est modifiée pour prendre en compte l'erreur d'étalonnage :

$$\Sigma^{\mathbf{q}} = J_{T_g} \left(\Sigma_{inter}^{\mathbf{q}} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix} \right) J_{T_g}^{\mathbf{T}}. \quad (6.9)$$

où J_{T_g} est la matrice Jacobienne de dimension 3×3 de la fonction T_g définie par :

$$J_{T_g} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} & 1 + \frac{\partial g}{\partial d} \end{pmatrix}. \quad (6.10)$$

Dans ce cas, l'incertitude d'un point 3D obtenue par propagation d'incertitude du point 2D \mathbf{p} est définie par la matrice de covariance $\Sigma^{\mathbf{Q}}$:

$$\Sigma^{\mathbf{Q}} = J_{T_2} J_{T_g} \left(J_{T_1} \Sigma^{\mathbf{p}} J_{T_1}^{\mathbf{T}} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix} \right) J_{T_g}^{\mathbf{T}} J_{T_2}^{\mathbf{T}}, \quad (6.11)$$

Deuxième cas : utilisation de notre deuxième méthode. Dans ce deuxième cas, nous corrigeons les points 3D en utilisant notre deuxième méthode d'étalonnage. Nous devons alors prendre en compte l'erreur de cet étalonnage dans le calcul de l'incertitude. Nous rappelons alors la transformation entre un point 3D avant correction et après correction (voir équation (5.13)) :

$$T_F : \Gamma \rightarrow \Omega$$

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \rightarrow A \begin{pmatrix} X + \frac{1}{Z}XC_Z \\ Y + \frac{1}{Z}YC_Z \\ Z + C_Z \\ 1 \end{pmatrix}, \quad (6.12)$$

où $C_Z = F \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ est la correction estimée par la première étape de notre méthode (la correction de la non-planarité, équation (5.1)) et A est la transformation affine estimée dans la deuxième étape. La matrice de covariance $\Sigma^{\mathbf{Q}}$ (6.6) ($\mathbf{Q} = (X \ Y \ Z)$) est modifiée pour prendre en compte l'erreur d'étalonnage :

$$\Sigma^{\mathbf{Q}} = J_{T_F} J_{T_2} \Sigma^{\mathbf{a}} J_{T_2}^T J_{T_F}^T, \quad (6.13)$$

où J_{T_F} est la matrice Jacobienne de dimension 3×3 de la fonction T_F définie par :

$$J_{T_F} = \begin{pmatrix} \frac{\partial T_{F1}}{\partial X} & \frac{\partial T_{F1}}{\partial Y} & \frac{\partial T_{F1}}{\partial Z} \\ \frac{\partial T_{F2}}{\partial X} & \frac{\partial T_{F2}}{\partial Y} & \frac{\partial T_{F2}}{\partial Z} \\ \frac{\partial T_{F3}}{\partial X} & \frac{\partial T_{F3}}{\partial Y} & \frac{\partial T_{F3}}{\partial Z} \end{pmatrix}. \quad (6.14)$$

Dans ce cas, l'incertitude d'un point 3D obtenue par propagation d'incertitude du point 2D \mathbf{p} est définie par la matrice de covariance $\Sigma^{\mathbf{Q}}$:

$$\Sigma^{\mathbf{Q}} = J_{T_F} J_{T_2} \left(J_{T_1} \Sigma^{\mathbf{p}} J_{T_1}^T + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix} \right) J_{T_2}^T J_{T_F}^T, \quad (6.15)$$

6.1.4 Propagation de l'incertitude des points 3D aux mesures de distance

En effectuant des mesures utilisant des points 3D \mathbf{Q}_i , l'incertitude provient de l'erreur de localisation de ces points 3D. Cette incertitude de localisation est modélisée par les matrices de covariance des points $\Sigma^{\mathbf{Q}_i}$. Nous avons choisi comme

mesure 3D la mesure de distance entre deux points 3D (distance donnant la longueur et la largeur d'un objet). Étant donné deux points $\mathbf{Q}_1, \mathbf{Q}_2$ et leurs matrices de covariance associées $\Sigma^{\mathbf{Q}_1}$ et $\Sigma^{\mathbf{Q}_2}$, la distance entre \mathbf{Q}_1 et \mathbf{Q}_2 est définie par la fonction D :

$$D : \psi^2 \rightarrow \mathbb{R}$$

$$\begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{pmatrix} \rightarrow \|\mathbf{Q}_1 - \mathbf{Q}_2\|_2 . \quad (6.16)$$

En supposant que les points \mathbf{Q}_1 et \mathbf{Q}_2 sont indépendants, une approximation au premier ordre de la variance σ_D^2 est donnée par :

$$\sigma_D^2 = J_D \begin{pmatrix} \Sigma^{\mathbf{Q}_1} & 0 \\ 0 & \Sigma^{\mathbf{Q}_2} \end{pmatrix} J_D^T, \quad (6.17)$$

où J_D est la matrice Jacobienne de dimension 1×6 de la fonction D .

6.2 Résultats

Nous présentons un simple exemple d'utilisation de notre modèle de bruit sur des données réelles. La caméra Temps-de-Vol utilisée dans cet expérimentation ainsi que les paramètres de capture sont présentés dans la section 1.4. La caméra est préalablement étalonnée (ses paramètres intrinsèques sont connus).

L'exemple consiste à mesurer les dimensions (longueur et largeur) d'un damier d'étalonnage. Ces mesures sont obtenues à partir des 4 pixels cliqués dans l'image $\mathbf{p}_1, \mathbf{p}_2$ et $\mathbf{p}_3, \mathbf{p}_4$ (voir figure 6.1b). Pour chaque pixel $(\mathbf{p}_i)_{i=1}^4$, la matrice de variance $\Sigma^{\mathbf{p}_i}$ est calculée : des clics multiples sont effectués et l'écart-type en chaque direction (u, v) est calculé. Ensuite, les points 3D correspondants $(\mathbf{Q}_i)_{i=1}^4$ et les matrices de covariance $(\Sigma^{\mathbf{q}_i})_{i=1}^4$ et $(\Sigma^{\mathbf{Q}_i})_{i=1}^4$ sont calculées par propagation d'erreur comme expliqué précédemment dans la section 6.1. Le premier pixel \mathbf{p}_1 est considéré comme un exemple pour présenter les résultats de chaque étape de propagation d'erreur. Les matrices de $\Sigma^{\mathbf{p}_1}$, $\Sigma^{\mathbf{q}_1}$ et $\Sigma^{\mathbf{Q}_1}$ en prenant en compte l'erreur d'étalonnage de notre deuxième méthode sont :

$$\Sigma^{\mathbf{p}_1} = \begin{pmatrix} 0,60^2 & 0 \\ 0 & 0,60^2 \end{pmatrix} \quad \Sigma^{\mathbf{q}_1} = \begin{pmatrix} 0,36 & 0 & 1,09 \\ 0 & 0,36 & 0,36 \\ 1,09 & 0,36 & 61,13 \end{pmatrix}$$

$$\Sigma^{\mathbf{Q}_1} = \begin{pmatrix} 39,39 & 4,11 & 7,47 \\ 4,11 & 40,95 & 0,78 \\ 7,47 & 0,78 & 60,95 \end{pmatrix} .$$

A partir de ces matrices, l'ellipse d'incertitude de \mathbf{p}_1 (figure 6.1c) et les ellipsoïdes d'incertitude de \mathbf{q}_1 (figure 6.1d) et de \mathbf{Q}_1 (figure 6.1e) sont calculées. L'ellipse ou l'ellipsoïde d'incertitude permet de représenter graphiquement les matrices de covariances des vecteurs aléatoires gaussiens. Ensuite, la longueur du damier $h = D(\mathbf{Q}_1, \mathbf{Q}_2)$ et sa largeur $w = D(\mathbf{Q}_3, \mathbf{Q}_4)$ sont calculées à partir de l'équation (6.16).

Nous avons corrigé les mesures issues du capteur Temps-de-Vol avec nos deux méthodes d'étalonnage. Une comparaison des résultats obtenus est présentée dans le tableau 6.1. Les dimensions du damier calculées avant l'étalonnage sont $w=1246$

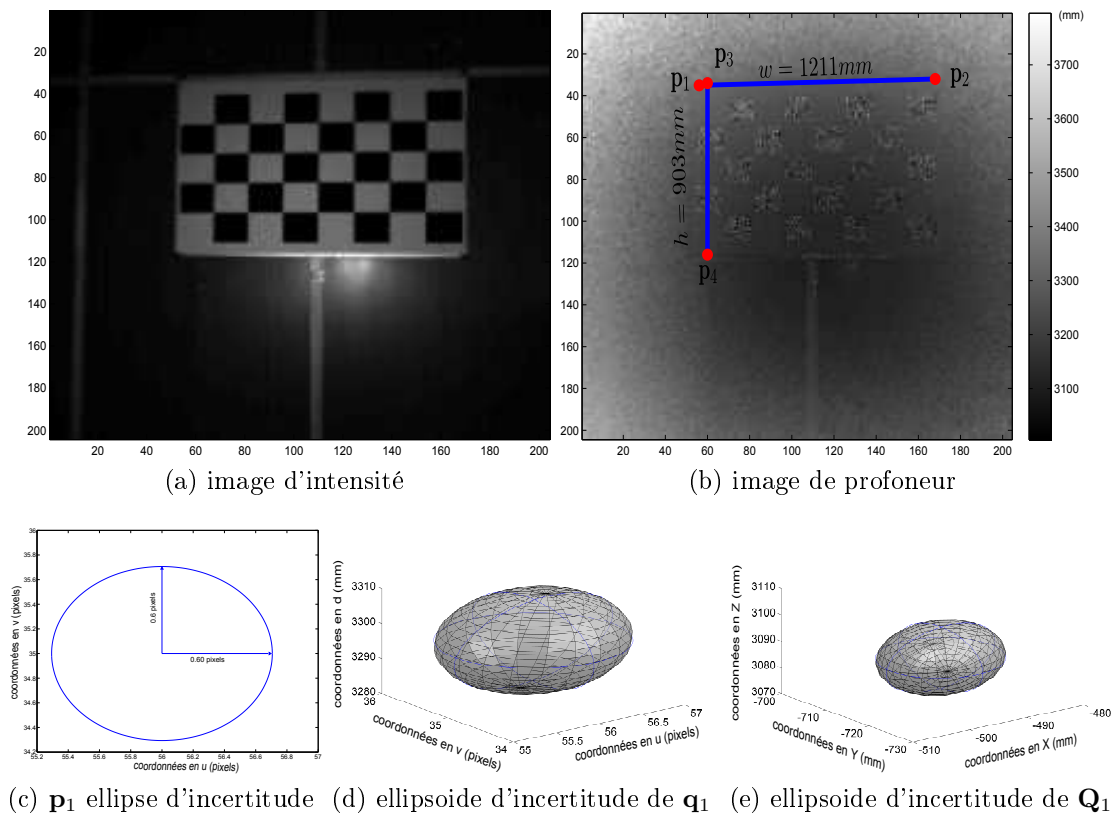


FIGURE 6.1 – Calcul des dimensions du damier d'étalonnage (longueur et largeur) à partir de points 2.5D : (a) image d'intensité (non utilisée dans les calculs) et (b) image de profondeur après correction de la distorsion optique. Les valeurs de la longueur et de la largeur calculées sont respectivement égales à $1211 \pm 11,18$ mm et à $903 \pm 8,97$ mm. Ces mesures sont obtenues en utilisant notre deuxième méthode d'étalonnage. (c) l'ellipse d'incertitude de \mathbf{p}_1 et l'ellipsoïde d'incertitude de (d) \mathbf{q}_1 et (e) de \mathbf{Q}_1 sont présentées. Leurs ellipses d'incertitude associées sont tracées en bleu autour des ellipsoïdes.

mm et $h=927$ mm. Leurs valeurs respectives en corrigeant la distorsion de la mesure de profondeur sont égales à $w=1228$ mm et $h=913$ mm avec la première méthode et à $w=1211$ mm et $h=903$ mm avec la deuxième méthode. Les mesures de la vérité terrain respectives sont $w=1200$ mm et $h=900$ mm. Nous observons une amélioration de la précision avec les deux méthodes d'étalonnage, les dimensions calculées après correction se rapprochent de la vérité terrain. En comparant les résultats des deux méthodes, nous constatons, contrairement à ce que nous avons présenté dans le tableau 5.3, une meilleure correction avec la deuxième méthode. Cela montre qu'en plus de sa simplicité, notre deuxième méthode est plus robuste car elle ne dépend pas d'une vérité terrain qui peut être entachée d'erreurs.

| | | largeur w (mm) | σ_w | longueur h (mm) | σ_h |
|------------------|-----------|------------------|------------|-------------------|------------|
| vérité terrain | | 1200 | | 900 | |
| avant correction | | 1246 | 11,60 | 927 | 9,31 |
| après correction | méthode 1 | 1228 | 11,37 | 913 | 9,19 |
| | méthode 2 | 1211 | 11,22 | 903 | 9,03 |

Tableau 6.1 – Comparaison des dimensions du damier (largeur et longueur) après correction de la mesure de profondeur avec nos deux méthodes d'étalonnage et de l'incertitude associée.

Après le calcul de la largeur et de la longueur du damier, leurs valeurs de variance sont calculées avec l'équation (6.17) et les incertitudes (qui sont égales aux écarts-type) est déduite. Leurs valeurs respectives obtenues, dans le cas où nous intégrons l'erreur d'étalonnage de notre deuxième méthode, sont $\sigma_w=11,22$ mm et $\sigma_h=9,03$ mm. Notons que les valeurs de la vérité terrain sont dans l'intervalle de confiance $[w - \sigma_w; w + \sigma_w]$ et $[h - \sigma_h; h + \sigma_h]$ avec un degré de confiance ¹ égal à 68%.

Dans les tests présentés précédemment, une seule image de profondeur a été utilisée. Nous proposons maintenant de refaire les mêmes tests en considérant plusieurs images de profondeur afin d'évaluer nos deux méthodes d'étalonnage sur un plus grand ensemble de données. Nous nous comparons également aux principales méthodes de l'état de l'art [Lindner & Kolb 2006] et [Kahlmann *et al.* 2006]. Pour ce faire, des images de profondeur correspondes à différentes vues du même damier d'étalonnage sont acquises avec le capteur Temps-de-Vol. Les vues sont choisies de telle sorte à couvrir tout l'espace d'étalonnage (différentes positions

1. Le degrés de confiance indique la probabilité de l'intervalle de confiance de contenir la valeur de la vérité terrain. L'augmentation du degré de confiance entraîne un étalement de l'intervalle de confiance et donc une diminution de la précision.

dans l'image et différentes distances par rapport au capteur sont considérées). 15 images sont utilisées. Pour chacune des images, les dimensions du damier (largeur et longueur) sont calculées avant et après correction de la mesure de profondeur comme présenté précédemment, ensuite la moyenne de l'erreur résiduelle par rapport à la vérité terrain est calculée. Les résultats sont illustrés dans le tableau 6.2. Nous observons une meilleure précision avec notre deuxième méthode pour les deux mesures (largeur et longueur). Pour la largeur, par exemple, l'erreur résiduelle moyenne après correction est égale à 16,28 mm avec notre deuxième méthode, tandis qu'elle est égale à 28,10 mm avec notre première méthode, 28,75 mm avec celle de [Lindner & Kolb 2006] et 29,17 mm avec celle de [Kahlmann *et al.* 2006]. Ces résultats prouvent que notre deuxième méthode est plus robuste que les autres méthodes basées entièrement sur des mesures de vérité terrain dont la précision ne peut pas être garantie principalement pour de longues distances. Pour notre première méthode, les résultats obtenus sont proches de ceux de la méthode [Lindner & Kolb 2006] et meilleurs que ceux de [Kahlmann *et al.* 2006].

| | | erreur résiduelle moyenne de la | |
|-----------------------------|-------------------------------|---------------------------------|-------------------|
| | | largeur w (mm) | longueur h (mm) |
| avant correction | | 45,50 | 27,83 |
| après correction avec | méthode 1 | 28,10 | 15,34 |
| | méthode 2 | 16,28 | 10,62 |
| | [Lindner & Kolb 2006] | 28,75 | 15,37 |
| | [Kahlmann <i>et al.</i> 2006] | 29,94 | 17,34 |

Tableau 6.2 – Erreur résiduelle moyenne des dimensions du damier (largeur et longueur) calculée avant et après correction de la mesure de profondeur. La correction est estimée avec les deux méthodes d'étalonnage proposées ainsi que les principales méthodes de l'état de l'art [Lindner & Kolb 2006] et [Kahlmann *et al.* 2006]. Un ensemble de 15 images du damier d'étalonnage positionné à différentes distances et à différentes positions dans l'image est considéré.

6.3 Conclusion

Nous avons présenté dans ce chapitre une simple application de nos travaux. Nous proposons une manière d'évaluer et de comparer nos deux méthodes d'étalonnage aux principales méthodes de l'état de l'art sans dépendre de la vérité terrain. Nous utilisons un objet étalon (dont les dimensions sont connues) et comparons ses

dimensions avant et après correction de la mesure de profondeur. Les résultats obtenus montrent que nos deux méthodes améliorent la précision de nos mesures. Les dimensions calculées après correction se rapprochent des dimensions réelles. Une meilleure correction est obtenue avec notre deuxième méthode, comparée à notre première méthode et à celles de l'état de l'art. Cela prouve que notre deuxième méthode est plus robuste car elle ne dépend pas d'une vérité terrain qui peut être entachée d'erreurs. Nous estimons également l'incertitude de calcul de ces dimensions en se basant sur notre modèle de bruit position-profondeur- TI (section 3.3.2). Cela est réalisé par propagation d'incertitude à travers les différentes formules de mesure.

Conclusion et perspectives

Nos travaux présentés dans ce mémoire traitent deux types d'erreur affectant les mesures de profondeur des caméras Temps-de-Vol : le bruit stochastique et la distorsion systématique. Nous avons caractérisé ces deux types d'erreur et avons présenté des contributions pour leur traitement. La première porte sur la modélisation du bruit des caméras Temps-de-Vol. La seconde consiste à modéliser et étalonner la distorsion systématique. Pour le bruit comme pour la distorsion systématique, deux manières de les modéliser et de les étalonner sont proposées. La première vise une meilleure précision, tandis que la seconde privilégie la simplicité de mise en œuvre.

Travaux effectués. Nous donnons ci-dessous nos conclusions par rapport aux contributions apportées dans cette thèse.

- **Modélisation du bruit stochastique.** Après avoir montré que le bruit suit une loi normale, nous avons proposé deux manières de le modéliser. Alors que les modèles de la littérature reposent uniquement sur l'amplitude, le premier modèle que nous avons proposé tient compte en plus de la position du pixel dans l'image. Quant au second, qui vise une meilleure précision, il remplace l'amplitude par la profondeur et le temps d'intégration. Une comparaison sur des données réelles montre qu'une meilleure précision est obtenue avec nos deux modèles, comparée à un modèle basé uniquement sur l'amplitude. Nous avons également montré que le second modèle offre une meilleure précision que le premier. Cependant, il est limité actuellement à une seule réflectivité. Afin de montrer l'intérêt de nos travaux, deux exemples d'application basés sur le modèle de bruit sont présentés : filtrage des cartes

de profondeur et calcul de l'incertitude de mesures géométriques formées à partir de ces cartes ;

- **Modélisation et étalonnage des mesures de profondeur.** Afin d'obtenir une modélisation de meilleure précision que celle de la littérature, nous avons proposé une première méthode basée sur une formulation variationnelle introduisant une contrainte de régularité. Elle utilise un modèle plus flexible prenant en compte l'ensemble des distorsions sous un même formalisme. Pour la mise en œuvre, cette méthode conserve l'approche classique de calcul de vérité terrain pour chaque point utilisé dans l'étalonnage. Une comparaison avec des méthodes d'état de l'art montre qu'elle permet d'obtenir une correction plus précise. Cependant, elle est sensible à la précision de la vérité terrain dont l'acquisition nécessite une mise en œuvre délicate. Pour être moins sensible à la vérité terrain et avoir une mise en œuvre plus simple, nous proposons une deuxième méthode qui utilise une vérité terrain structurelle liée à la géométrie plane de la scène observée. Elle est basée principalement sur la correction de la non planarité qui nécessite un ensemble de vues d'une surface plane faciles à acquérir. En la comparant à deux méthodes d'état de l'art sur des données réelles, nous observons qu'en plus de sa simplicité de mise en œuvre, elle garantit une bonne précision. Pour ne pas dépendre de la vérité terrain dans notre évaluation et comparaison de nos deux méthodes, nous proposons de mesurer les dimensions d'un objet étalon à partir de sa carte de profondeur et de les corriger avec nos deux méthodes. En comparant les résultats obtenus, nous avons observé une meilleure précision avec notre seconde méthode. Elle est, en effet, plus robuste car elle ne dépend pas d'une vérité terrain qui pourrait être entachée d'erreurs.

Nous avons illustré l'intérêt d'avoir ces deux modèles au travers d'une application de mesure 3D à partir de cartes de profondeur issues du capteur Temps-de-Vol. Pour cela, nous proposons d'estimer les mesures 3D ainsi que leurs incertitudes associées par propagation d'incertitude en nous basant sur le modèle du bruit et en corrigeant les distorsions.

Perspectives. Quelques améliorations peuvent être apportées à nos travaux afin d'obtenir une modélisation plus complète des deux types d'erreur considérés :

- **Prise en compte de la réflectivité dans la modélisation du bruit.** Dans notre premier modèle de bruit, la réflectivité était incorporée implicitement dans l'information d'amplitude. Alors que dans le second, elle était considérée constante. Il est donc important de l'intégrer dans le second modèle et de pouvoir comparer nos deux modèles sur des données où la réflec-

tivité varie. La manière de modéliser la variation du bruit en fonction de la réflectivité doit être étudiée ;

- **Extension des modèles d'étalonnage de la mesure de profondeur.** Les deux méthodes d'étalonnage proposées prennent en compte le principal facteur de la distorsion de la mesure de profondeur qui est *l'erreur systématique*. Ces méthodes peuvent être améliorées en prenant en compte d'autres facteurs d'erreur comme le temps d'intégration et la réflectivité.
- **Test des méthodes d'étalonnage sur d'autres données.** Il serait intéressant de pouvoir tester et évaluer nos méthodes d'étalonnage de la mesure de profondeur en se basant sur une vérité terrain de haute précision. Une telle précision pourrait être obtenue avec les systèmes de mesure tels qu'un rail de mesure ou une *'total station'* ;
- **Test des méthodes d'étalonnage sur d'autres capteurs.** Les modèles utilisés dans nos deux méthodes d'étalonnage n'intègrent aucune contrainte liée aux caméras Temps-de-Vol. Ils ne sont donc pas limités à ce type de caméras et il serait possible de les tester sur d'autres capteurs comme le Kinect ;
- **Auto-étalonnage** Les résultats expérimentaux ont montré que notre seconde méthode d'étalonnage est à la fois plus simple et plus robuste. Cependant, sa limite est la nécessité de disposer d'une grande surface plane qui n'est pas toujours disponible. Cette contrainte peut être supprimée en l'intégrant au cours du processus d'étalonnage pour faire de "l'auto-étalonnage". Ainsi, Kinect fusion fournit une estimation de la géométrie de la scène qui peut remplacer la contrainte de planarité. Une manière de réaliser l'étalonnage consisterait à optimiser, dans un processus itératif, la reconstruction de la scène ainsi que les paramètres de distorsion de telle sorte à minimiser l'erreur entre chaque vue issue du modèle reconstruit par la caméra Temps-de-Vol et celui reconstruit avec Kinect fusion.

ANNEXE A

Analyse du bruit de la caméra Kinect

Nous nous intéressons, dans cette annexe, à l'analyse du bruit des caméras Kinect. Cette étude est réalisée afin de comparer l'intensité et les facteurs de variation du bruit de cette caméra avec ceux des caméras Temps-de-Vol que nous avons présenté dans le chapitre 3.

Notre étude du bruit est basée sur le calcul de la variation des mesures de profondeur de différentes images de profondeur prises dans les mêmes conditions. Des images d'un mur blanc (même objet utilisé dans l'étude du bruit des caméra Temps-de-Vol) sont acquises pour des distances allant de 0,5 m à 3,5 m. A chaque distance, 100 images de profondeur sont acquises. L'écart-type σ est calculé pour chaque point 2.5D des images de profondeur. Nous nous intéressons à la variation du bruit en fonction de la position du pixel dans l'image. La figure A.1 présente

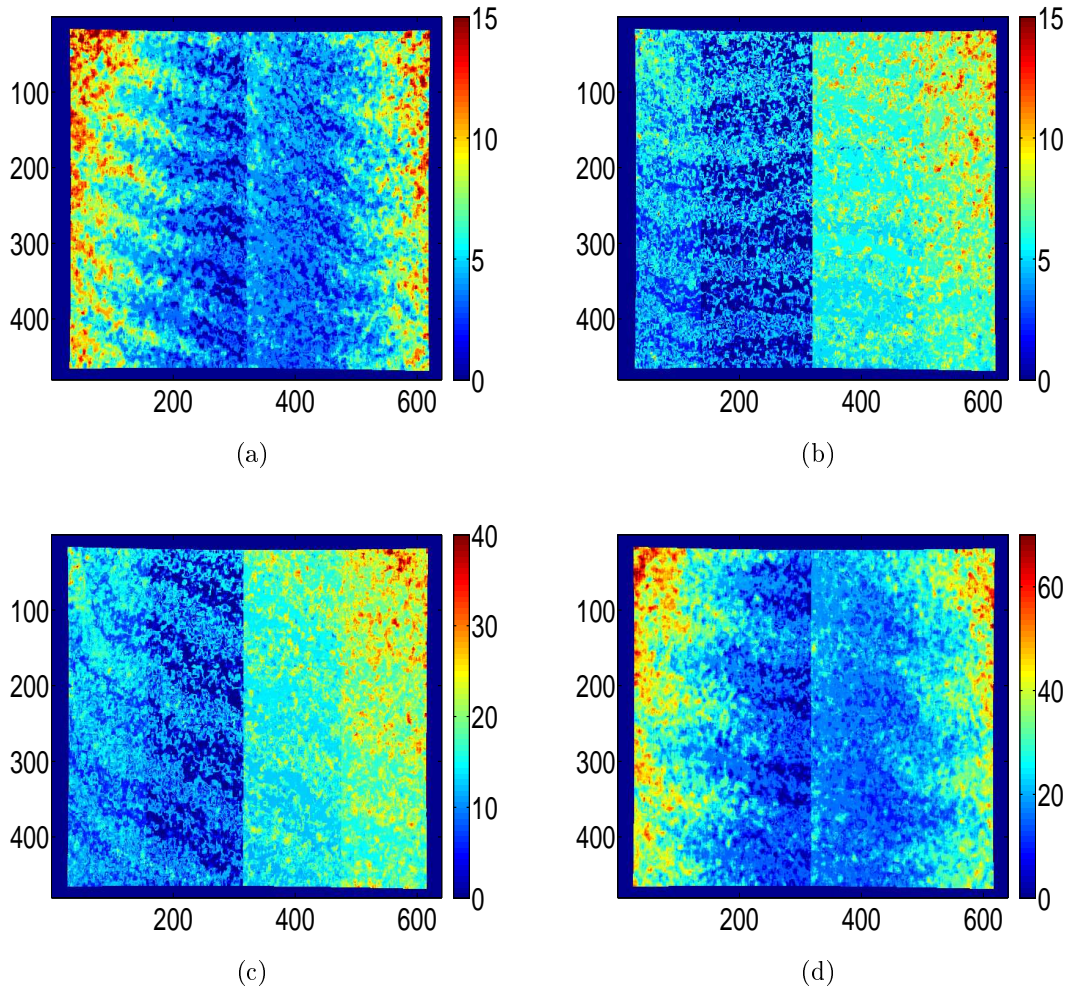


FIGURE A.1 – Écart-type σ (mm) d'une image de profondeur à une distance approximative de : (a) 1,5 m, (b) 2 m, (c) 3 m et (d) 3,5 m. Les formes des 4 figures sont différentes. Contrairement aux caméras Temps-de-Vol, la position du pixel dans l'image n'est pas un facteur de variation de σ pour les caméras Kinect.

l'écart type calculé pour 4 différentes images de profondeur. Nous constatons que les figures n'ont pas la même forme et sont très différents de ceux obtenus avec les caméras Temps-de-Vol (figure 3.6). Pour les caméras temps de vol, σ augmente du centre de l'image vers les bords de l'image, tandis que, pour les caméras Kinect, le bruit dépend de l'angle d'observation de l'objet. Nous nous intéressons maintenant à la variation du bruit en fonction de la profondeur. La figure A.2 présente un aperçu de la variation de σ en fonction de la profondeur pour une plage de distance allant de 1.25 m jusqu'à 3,5 m. Nous observons que σ augmente en fonction de la profondeur, comme c'est le cas des caméras Temps-de-Vol (figure 3.7). La valeur moyenne de σ , à une distance égale à 3,5 m, atteint approximativement 25 mm, tandis qu'elle ne dépasse pas 10 mm pour des *TIs* supérieur à 3,5 ms dans le cas des caméras Temps-de-Vol.

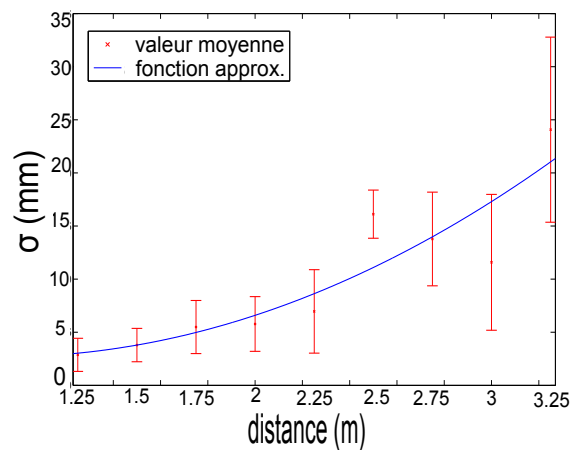


FIGURE A.2 – Écart-type σ (mm) correspondant à des images de profondeur d'un mur blanc tracé en fonction de la profondeur. La valeur moyenne de σ est tracée en points rouge et la fonction approximant ces données en ligne bleue. La valeur moyenne de σ augmente en fonction de la profondeur approximativement de 2,5 mm jusqu'à 25 mm.

ANNEXE B

Notions de base

Cette deuxième annexe expose les notions de base nécessaires à la lecture de cette thèse. Nous y décrivons tout d'abord des outils de vision par ordinateur. Nous présentons le modèle de caméra utilisé, ainsi que les méthodes de résolution des systèmes linéaires et les méthodes d'étalonnage des caméras et de calcul de pose. Ensuite, nous définissons des outils statistiques et mathématiques : méthode de propagation de l'incertitude, la fonction de type B-spline et la fonction spline de type plaque mince. Finalement, nous définissons la transformation de l'espace 2.5D vers l'espace 3D et la transformation entre les fonctions de correction de la distorsion de la mesure de profondeur définies dans ces deux espaces. Pour compléter la lecture de ce chapitre, nous renvoyons aux livres [Hartley & Zisserman 2004] et [Faugeras 1993].

B.1 Outils de vision par ordinateur

B.1.1 Modélisation de la caméra

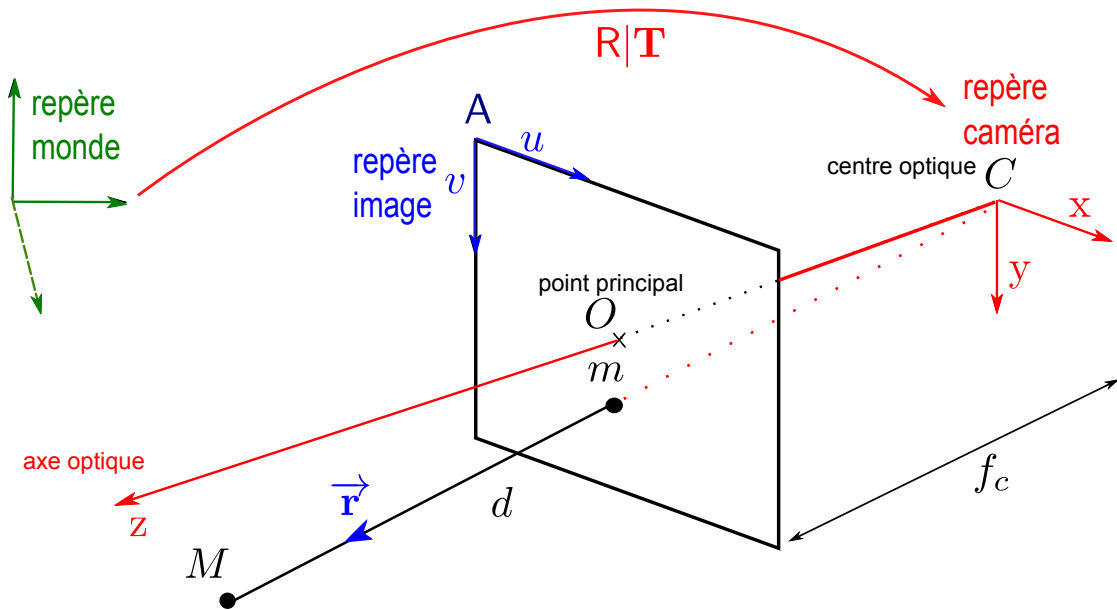


FIGURE B.1 – Modèle sténopé et repères associés. Le repère associé à la caméra est choisi de telle sorte que son centre soit placé en C , son axe z confondu avec l’axe optique et l’axe x parallèle aux bords horizontaux de l’image. Le repère image (dans lequel les coordonnées sont exprimées en pixels) a son centre A dans le coin supérieur gauche de l’image, ses axes u et v sont orientés respectivement de la gauche vers la droite et du haut vers le bas.

Le modèle adopté pour modéliser une caméra est le modèle sténopé, appelé aussi *modèle trou d’épingle*, illustré sur la figure B.1. Ce modèle suppose qu’il existe un point C , appelé centre optique, par lequel passent tous les rayons lumineux. Ce centre optique est situé à une distance f_c (distance focale) du plan image. On appelle axe optique la droite orthogonale au plan image et passant par C . Le projeté O de ce point sur le plan image est appelé point principal de l’image.

B.1.1.1 Modèle sténopé

Projection perspective. La transformation géométrique liant les coordonnées 3D d’un point observé, dans un repère qui lui est lié (appelé repère monde), aux coordonnées dans l’image, exprimées en pixels, peut être décomposée en trois transformations :

- La première transformation correspond à un changement de repère entre le repère monde (de l'objet observé) et celui de la caméra. Elle est donc définie par la composée d'une rotation \mathbf{R} et d'une translation \mathbf{T} . Soit \mathbf{M} un point 3D observé, \mathbf{P} ses coordonnées dans le repère monde choisi et \mathbf{Q} ses coordonnées dans le repère caméra. Ces deux vecteurs sont liés par la relation :

$$\mathbf{Q} = \mathbf{R}\mathbf{P} + \mathbf{T}, \quad (\text{B.1})$$

En utilisant les coordonnées homogènes $\check{\mathbf{P}} = (U, V, W, 1)^T$ et $\check{\mathbf{Q}} = (X, Y, Z, 1)^T$ pour ces vecteurs, cette transformation peut être exprimée par une relation linéaire :

$$\check{\mathbf{Q}} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix} \check{\mathbf{P}}. \quad (\text{B.2})$$

Une autre représentation de $\check{\mathbf{Q}}$ en fonction de $\check{\mathbf{C}}$, les coordonnées homogènes du centre optique dans le repère monde, est sous la forme :

$$\check{\mathbf{Q}} = \mathbf{R}(\check{\mathbf{P}} - \check{\mathbf{C}}), \quad (\text{B.3})$$

avec

$$\check{\mathbf{C}} = -\mathbf{R}^{-1}\mathbf{T}, \quad (\text{B.4})$$

- La deuxième transformation est une projection de centre C et faisant passer les coordonnées d'un point dans le repère caméra aux coordonnées métriques du point image qui lui est correspondant. Les coordonnées millimétriques $\mathbf{m} = (x, y)^T$ du point 2D m , projeté de \mathbf{Q} dans l'image, dans le repère (O, x, y) , sont :

$$\begin{aligned} x &= f_c \frac{X}{Z} \\ y &= f_c \frac{Y}{Z} \end{aligned} \quad (\text{B.5})$$

soit :

$$\check{\mathbf{m}} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim \begin{pmatrix} f_c X \\ f_c Y \\ Z \end{pmatrix}$$

Le symbole \sim est utilisé puisque l'égalité vectorielle est définie à un facteur scalaire près, ce qui donne :

$$\check{\mathbf{m}} \sim \begin{pmatrix} f_c & 0 & 0 & 0 \\ 0 & f_c & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \check{\mathbf{Q}}. \quad (\text{B.6})$$

- La dernière transformation est un changement de coordonnées dans le repère image. Les points image sont en effet mesurés en pixels dans le repère (A, u, v) associé à l'image. En notant d_u et d_v les tailles horizontale et verticale d'un pixel (en général $d_u = d_v$ car les pixels sont carrés), et (c_u, c_v) les coordonnées du point principal dans le repère image, les coordonnées notées \mathbf{p} du point m dans le repère image sont données par :

$$\mathbf{p} = \begin{pmatrix} \frac{1}{d_u} & \gamma \\ 0 & \frac{1}{d_v} \end{pmatrix} \mathbf{m} + \begin{pmatrix} c_u \\ c_v \end{pmatrix} \quad (\text{B.7})$$

où γ est un terme traduisant l'éventuelle non orthogonalité des axes du repère de l'image, en général très proche de zéro. En coordonnées homogènes nous obtenons :

$$\check{\mathbf{p}} = \begin{pmatrix} \frac{1}{d_u} & \gamma & c_u \\ 0 & \frac{1}{d_v} & c_v \\ 0 & 0 & 1 \end{pmatrix} \check{\mathbf{m}}. \quad (\text{B.8})$$

La relation globale entre $\check{\mathbf{p}}$ et $\check{\mathbf{P}}$ s'écrit finalement :

$$\check{\mathbf{p}} \sim \begin{pmatrix} \frac{1}{d_u} & \gamma & c_u \\ 0 & \frac{1}{d_v} & c_v \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f_c & 0 & 0 & 0 \\ 0 & f_c & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix} \check{\mathbf{P}} \quad (\text{B.9})$$

$$\sim \begin{pmatrix} f_u & \gamma & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{T} \end{bmatrix} \check{\mathbf{P}}, \quad (\text{B.10})$$

où $f_u = \frac{f_c}{d_u}$ et $f_v = \frac{f_c}{d_v}$ sont les composantes de la focale effective exprimée en pixels.

En général $f_u = f_v$ car les pixels sont carrés. En notant $\mathbf{K} = \begin{pmatrix} f_u & \gamma & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{pmatrix}$ la matrice des paramètres intrinsèques de la caméra, il vient donc :

$$\boxed{\check{\mathbf{p}} \sim \mathbf{K} \begin{bmatrix} \mathbf{R} & \mathbf{T} \end{bmatrix} \check{\mathbf{P}}}. \quad (\text{B.11})$$

Les 6 paramètres de position de la caméra par rapport au repère du monde qui sont les 3 paramètres de \mathbf{T} et les 3 paramètres de la matrice \mathbf{R} sont appelés les paramètres extrinsèques. $\check{\mathbf{p}}$ peut aussi être exprimé par :

$$\boxed{\check{\mathbf{p}} \sim \mathbf{K} \mathbf{R} \begin{bmatrix} \mathbf{I} & -\check{\mathbf{C}} \end{bmatrix} \check{\mathbf{P}}}. \quad (\text{B.12})$$

La matrice de projection de la caméra est notée \mathbf{P} :

$$\mathbf{P} \sim \mathbf{K} \begin{bmatrix} \mathbf{R} & \mathbf{T} \end{bmatrix} \quad (\text{B.13})$$

$$\sim \mathbf{K} \mathbf{R} \begin{bmatrix} \mathbf{I} & -\check{\mathbf{C}} \end{bmatrix}. \quad (\text{B.14})$$

Rétroprojection perspective. La *rétroprojection* peut être vue comme l'opération inverse de la projection. Son but est d'inférer la ligne de vue \mathbf{r} passant par le point 3D \mathbf{M} de coordonnées \mathbf{P} à partir de son observation 2D \mathbf{m} (de coordonnées \mathbf{p}) dans l'image. (voir figure B.1) :

$$\mathbf{r} = \mathbf{R}^{-1}\mathbf{K}^{-1}\mathbf{x}. \quad (\text{B.15})$$

Ce rayon n'est pas normalisé. Une normalisation avec $\sqrt{\mathbf{p}^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{p}}$ est parfois nécessaire. La position du point 3D \mathbf{P} est exprimée en fonction sa profondeur d :

$$\begin{aligned} \mathbf{P} &= d \frac{\mathbf{r}}{\sqrt{\mathbf{p}^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{p}}} + \mathbf{C} \\ &= d \frac{\mathbf{R}^T\mathbf{K}^{-1}\mathbf{p}}{\sqrt{\mathbf{p}^T\mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{p}}} + \mathbf{C}. \end{aligned} \quad (\text{B.16})$$

B.1.1.2 Modèle de distorsion optique

En pratique, le modèle sténopé peut être insuffisant. Des modèles plus complexes intégrant par exemple les distorsions optiques radiales et tangentielles ont été proposés. Par exemple, le terme de distorsion radiale suivant :

$$\begin{aligned} \Delta u_r &= u(k_1 r^2 + k_2 r^4 + k_3 r^6) \\ \Delta v_r &= v(k_1 r^2 + k_2 r^4 + k_3 r^6), \end{aligned}$$

où $r = \sqrt{(u^2 + v^2)}$ est la distance au point principal de l'image et k_1 , k_2 et k_3 sont 3 coefficients. Le terme de distorsion tangentielle peut être donné par :

$$\begin{aligned} \Delta u_t &= \pi_1(r^2 + 2u^2) + 2\pi_2 uv \\ \Delta v_t &= \pi_2(r^2 + 2u^2) + 2\pi_1 uv, \end{aligned}$$

où π_1 et π_2 sont deux coefficients.

Ces distorsions correspondent à une modification du plan image et peuvent être facilement ajoutées au modèle sténopé déjà étudié. Plus de détails peuvent être trouvés dans le livre [Hartley & Zisserman 2004]. Les effets des distorsions radiales sont souvent représentés par des images déformées en barillet ou coussinet. En pratique, comme ces distorsions s'appliquent uniquement au niveau du plan image, il est possible de rectifier les images afin d'obtenir les images dans le plan image d'une caméra sténopé sans distorsion.

Notons que toutes les images utilisées dans ce mémoire sont rectifiées (la distorsion optique est corrigée). La figure B.2 présente un exemple d'image issue de la

caméra Temps-de-Vol et celle obtenue après la correction de la distorsion optique. Après correction, les bords du damier d'étalonnage ne sont plus courbés.

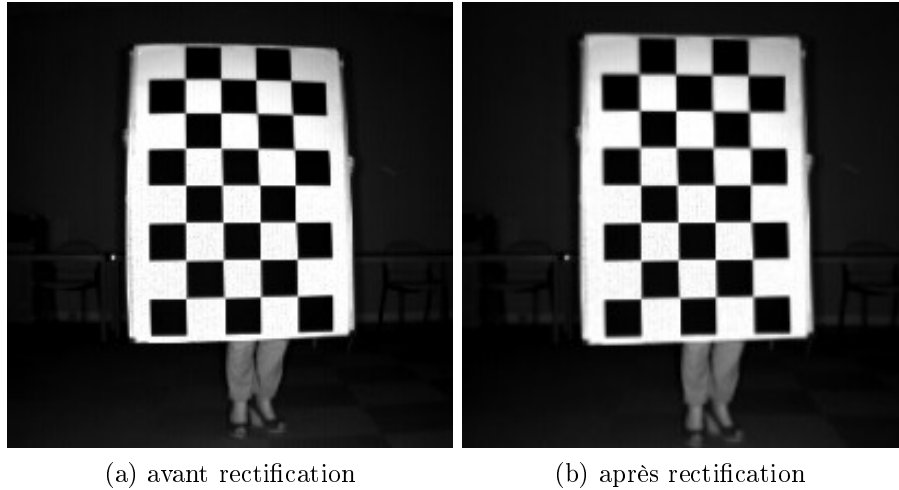


FIGURE B.2 – Image issue de la caméra Temps-de-Vol utilisée dans ce mémoire : (a) avant et (b) après correction de la distorsion optique.

B.1.2 Résolution de systèmes linéaires

B.1.2.1 Décomposition en valeurs singulières

La décomposition en valeurs singulières (en anglais SVD, "Singular Value Decomposition") est un outil permettant, entre autres, d'obtenir la solution d'un système linéaire exact, de trouver la solution au sens des moindres carrés d'un système linéaire sur-contraint et de donner une famille de solutions pour un système linéaire sous-contraint. La décomposition en valeurs singulières donne aussi un diagnostic de la situation. Dans cette thèse, elle est utilisée pour la résolution au sens des moindres carrés de systèmes linéaires sur-contraints. Nous nous contentons de la présenter brièvement.

La décomposition en valeurs singulières est basée sur le théorème de l'algèbre linéaire suivant. Toute matrice A de taille $(r \times c)$ peut être décomposée comme le produit des trois matrices suivantes :

$$A_{r \times c} = U_{r \times r} D_{c \times c} V_{c \times c}^T, \quad (\text{B.17})$$

où U a des colonnes orthogonales, V est orthogonale et contient les vecteurs singuliers de A et D est diagonale et contient les valeurs singulières de A . Lorsque le rang de A n'est pas plein, les dernières colonnes de V forment une base orthogonale

de l'espace nul de \mathbf{A} . Supposons $r \geq c$, comme c'est le cas lorsque \mathbf{A} représente un système linéaire exact ou sur-contraint. Notons $a = \text{rang}(\mathbf{A}) \leq c$ le rang de \mathbf{A} . La dimension de l'espace nul de \mathbf{A} est donné par $g = c - a$. Les matrices \mathbf{D} et \mathbf{V} peuvent être réécrites comme suit :

$$\mathbf{D}_{c \times c} = \text{diag}(\sigma_1, \dots, \sigma_a, \underbrace{0, \dots, 0}_g) \quad (\text{B.18})$$

$$\mathbf{V}_{c \times c} = \left(\underbrace{\mathbf{V}_1}_{c \times a} \quad \underbrace{\mathbf{V}_2}_{c \times g} \right) \quad (\text{B.19})$$

où les σ_i sont les valeurs singulières. Les colonnes de la matrices \mathbf{V}_2 forment une base orthogonale de l'espace nul de \mathbf{A} , c'est-à-dire :

$$\mathbf{A}\mathbf{V}_2 = \mathbf{0}. \quad (\text{B.20})$$

B.1.2.2 Pseudo-inverse

Certains problèmes étudiés dans nos travaux conduisent à la résolution de systèmes non homogènes de la forme :

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{b} \neq \mathbf{0}. \quad (\text{B.21})$$

Un tel système a une solution unique lorsque le rang de \mathbf{A} est plein. La solution au sens des moindres carrés minimise le critère d'erreur suivant :

$$C_A = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2. \quad (\text{B.22})$$

Elle est donnée par la résolution des équations induites par le système suivant :

$$\begin{aligned} \mathbf{A}^T \mathbf{A} \mathbf{x} &= \mathbf{A}^T \mathbf{b} \\ \mathbf{x} &= \underbrace{(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T}_{\mathbf{A}^\dagger} \mathbf{b}, \end{aligned} \quad (\text{B.23})$$

où \mathbf{A}^\dagger est appelée une pseudo-inverse de \mathbf{A} .

B.1.3 Étalonnage et calcul de pose

B.1.3.1 Étalonnage standard d'une caméra

Nous utilisons le terme étalonnage standard pour ne pas le confondre avec l'étalonnage de la mesure de profondeur. L'étalonnage standard (en anglais "calibration") est l'estimation des paramètres intrinsèques de la caméra (les coefficients

de la matrice \mathbf{K} introduite dans la section B.1.1.1) ainsi que les paramètres de distorsion optique (introduits dans la section B.1.1.2). Il existe plusieurs méthodes d'étalonnage standard. Nous présentons, dans cette section, la méthode que nous avons utilisée pour étalonner les caméras utilisées dans nos travaux.

Nous avons utilisé un outil existant pour estimer les paramètres du modèle de caméra. Il s'agit de "Calibration Camera Toolbox for Matlab" ([Bouguet 2010]). La méthode d'étalonnage "Direct Linear Transforma" (DLT) est utilisée pour obtenir une première initialisation des paramètres du modèle. Ensuite, un raffinement de la solution par optimisation non linéaire est effectuée.

"Direct Linear Transform" (DLT). C'est une méthode permettant d'estimer directement la matrice de projection \mathbf{P} à partir de l'observation des projections sur l'image des points 3D dont la position spatiale est connue. Cet ensemble de points 3D constitue ce que l'on appelle une mire d'étalonnage (présentée plus en détail dans la suite). Chaque point 3D M_i observé fournit une équation liant ses coordonnées 3D \mathbf{P}_i , dans un repère qui lui est lié, à son projeté dans l'image $\mathbf{p}_i = (u_i \ v_i)^\top$:

$$\check{\mathbf{p}}_i \sim \mathbf{P}\check{\mathbf{P}}_i. \quad (\text{B.24})$$

qui se développe en :

$$\begin{pmatrix} \mathbf{P}_1^\top & p_{14} \\ \mathbf{P}_2^\top & p_{24} \\ \mathbf{P}_3^\top & p_{34} \end{pmatrix} \check{\mathbf{P}}_i = \begin{pmatrix} w_i u_i \\ w_i v_i \\ w_i \end{pmatrix}$$

En passant aux coordonnées réelles, nous obtenons :

$$u_i = \frac{\mathbf{P}_1^\top \mathbf{P}_i + p_{14}}{\mathbf{P}_3^\top \mathbf{P}_i + p_{34}} \quad \text{qui se développe en } \mathbf{P}_1^\top \mathbf{P}_i - u_i \mathbf{P}_3^\top \mathbf{P}_i + p_{14} - u_i p_{34} = 0.$$

$$v_i = \frac{\mathbf{P}_2^\top \mathbf{P}_i + p_{24}}{\mathbf{P}_3^\top \mathbf{P}_i + p_{34}} \quad \text{qui se développe en } \mathbf{P}_2^\top \mathbf{P}_i - v_i \mathbf{P}_3^\top \mathbf{P}_i + p_{24} - v_i p_{34} = 0.$$

Ce sont en fait deux équations de plan. Avec un minimum de six points, les douze équations nécessaires à l'estimation de la matrice de projection \mathbf{P} sont obtenues. Pour se faire, il faut représenter le système sous la forme suivante : $\mathbf{A}\mathbf{X} = 0$ où $\mathbf{X} = (p_{11}, \dots, p_{34})^\top$ regroupe les différents coefficients de la matrice \mathbf{P} . La méthode SVD (présentée en détail dans la section B.1.2.1) est utilisée pour résoudre le système. La solution d'un système $\mathbf{A}\mathbf{X} = 0$ au sens des moindres carrés forme une base de l'espace nul de \mathbf{A} défini par les colonnes de la matrice \mathbf{V}_2 (les colonnes de \mathbf{V} correspondantes aux valeurs singulières nulles, voir équation (B.19) pour plus

de détail). Une fois la solution du système linéaire ci-dessus obtenue, sa décomposition sous la forme décrite dans l'équation (B.9) est effectuée. Cette méthode est détaillée dans le livre [Faugeras 1993] et [Hartley & Zisserman 2004].

Raffinement par optimisation non linéaire. Il n'existe pas, en général, de matrice P permettant d'aligner parfaitement le modèle aux données pour plus de cinq points. Nous cherchons donc le meilleur jeu de paramètres afin de minimiser une erreur quadratique : l'erreur de reprojection (exprimée en pixels) exprime pour un point M_i la distance géométrique dans l'image entre le point \widehat{m}_i obtenu par le modèle paramétrique et le point m_i réellement observé dans l'image. Ayant à disposition k observations de points, la matrice P optimale est celle qui minimise le critère :

$$\sum_{i=1}^k d(\widehat{m}_i, m_i) \quad (\text{B.25})$$

Pour minimiser ce critère en considérant une distance géométrique, une méthode linéaire telle que la DLT n'est pas suffisante et il faut utiliser des outils d'optimisation non linéaire. Partant de la solution fournie par la DLT, l'optimisation non linéaire permet de raffiner d'une manière itérative le jeu de paramètres afin de minimiser le critère.

En pratique, le modèle complet de caméra intègre les distorsions optiques (présentées dans la section B.1.1.2). Notons qu'il est possible de supposer les distorsions optiques nulles dans un premier temps afin d'estimer les paramètres de la matrice de projection P par DLT puis d'estimer l'intégralité des paramètres du modèle lors de l'optimisation non linéaire.

Mire d'étalonnage. Supposons que tous les points M_i soient disposés dans un unique plan. Le système linéaire à résoudre par DLT est alors dit mal conditionné. Cela signifie qu'il n'est pas possible d'estimer les paramètres du modèle dans cette configuration. Il faut donc utiliser un ensemble de points n'appartenant pas à un même plan. En pratique, une telle mire est assez difficile à réaliser précisément et l'on préfère utiliser plusieurs mires planes ou une mire plane observée dans différentes configurations.

B.1.3.2 Calcul de pose d'une caméra

L'estimation de la pose d'une caméra consiste à déterminer la position et l'orientation de la caméra à partir de points 3D déjà reconstruits et leurs coordonnées 2D dans l'image courante. Dans cette thèse, elle est utilisée pour calculer la vérité

terrain de la première méthode d'étalonnage de la mesure de profondeur présentée dans le chapitre 5.

Pour estimer la pose d'une caméra, les paramètres intrinsèques de la caméra doivent être connus. Il s'agit donc de déterminer la pose $[\mathbf{R} \ \mathbf{T}]$ (\mathbf{R} rotation et \mathbf{T} translation) à partir de correspondances de points 3D/2D. Cela est réalisé par minimisation de l'erreur de reprojection de plusieurs correspondances de points 2D/3D. L'erreur de reprojection consiste à mesurer la distance entre l'observation d'un point 3D dans l'image (représenté par le point m de coordonnées \mathbf{p} sur la figure B.3) et la projection du point 3D M de coordonnées $\check{\mathbf{P}}$ reconstruit dans cette même image (noté \hat{m} de coordonnées $\check{\mathbf{P}}$) où \mathbf{P} représente la matrice de projection de la caméra (plus de détails dans la section B.1.1). \mathbf{P} est la matrice de projection contenant les paramètres intrinsèques de la caméra \mathbf{K} et la pose \mathbf{R} et \mathbf{T} (voir l'équation (B.13)). Notons que la caméra est déjà étalonnée : les paramètres intrinsèques (\mathbf{K}) sont connus. Ayant à disposition k observations de points, la pose de la caméra \mathbf{R} et \mathbf{T} est celle qui minimise le critère :

$$\sum_{i=1}^k d(\hat{m}_i, m_i). \quad (\text{B.26})$$

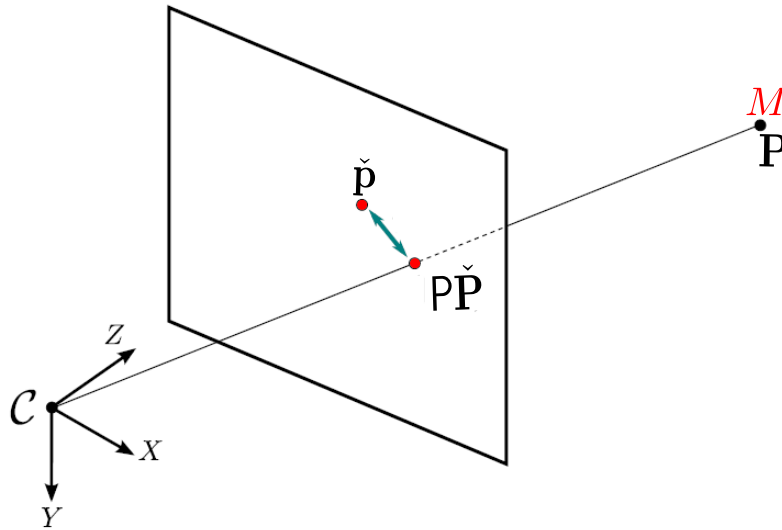


FIGURE B.3 – Erreur de reprojection : distance entre l'observation 2D (m de coordonnées \mathbf{p}) d'un point 3D (M de coordonnées \mathbf{P}) et sa projection \hat{m} dans l'image de coordonnées $\check{\mathbf{P}}$.

B.1.3.3 Étalonnage stéréo

Un capteur stéréoscopique est constitué de deux caméras qui observent une même scène. Il permet d'obtenir un couple d'images représentant deux projections différentes de la scène. Dans nos travaux, nous avons utilisé un système stéréoscopique pour calculer la vérité terrain utilisée dans la première méthode d'étalonnage de la mesure de profondeur présentée dans le chapitre 4. Pour se faire, nous avons besoin de connaître la transformation rigide liant les deux caméras, c'est ce que nous déterminons par ce que nous appelons étalonnage stéréo.

Dans un système stéréoscopique, pour un point de la scène, il existe deux projections du point dans deux images différentes. Nous parlons d'image gauche issue de la caméra gauche et d'image droite issue de la caméra droite. Un repère est associé également à chacune des caméras : repère caméra gauche et repère caméra droite (voir figure B.4).

Une mire d'étalonnage (voir section B.1.3.1) est utilisée pour effectuer l'étalonnage stéréo. Chaque point 3D M de la mire d'étalonnage se projette dans l'image gauche en m_L et dans l'image droite en m_R . Notre objectif est de déterminer les paramètres extrinsèques du capteur stéréoscopique *i.e.* la matrice de rotation (notée $\mathbf{R}_{L \rightarrow R}$) et le vecteur de translation (notée $\mathbf{T}_{L \rightarrow R}$). Nous avons choisi de calculer la transformation de la caméra gauche vers la caméra droite. Les paramètres stéréo sont définis de telle sorte que si nous considérons un point 3D M , ses coordonnées \mathbf{Q}_L et \mathbf{Q}_R respectivement dans le repère de la caméra gauche et droite sont reliées par la transformation rigide donnée par :

$$\mathbf{Q}_R = \mathbf{R}_{L \rightarrow R} \mathbf{Q}_L + \mathbf{T}_{L \rightarrow R}. \quad (\text{B.27})$$

Nous utilisons aussi le même outil "Calibration Camera Toolbox for Matlab" pour l'étalonnage stéréo. Nous supposons que les paramètres intrinsèques des deux caméras sont connus (déjà estimés par la méthode présentée dans la section B.1.3.1). Le meilleur jeu de paramètres stéréo est choisi de telle sorte qu'il minimise une erreur quadratique : l'erreur de reprojection (exprimée en pixels) pour un ensemble de points. L'erreur de reprojection pour un point M consiste à mesurer la distance entre la projection du point m_L dans l'image droite notée \widehat{m}_R et le point m_R réellement observé dans l'image ainsi que la distance entre la projection du point m_R dans l'image gauche notée \widehat{m}_L et le point m_L réellement observé dans l'image. Les coordonnées du point \widehat{m}_R sont $(\mathbf{K}_R (\mathbf{R}_{L \rightarrow R} \mathbf{K}_L^{-1} \mathbf{P}_L \check{\mathbf{P}} + \mathbf{T}_{L \rightarrow R}))$ avec $\check{\mathbf{P}}$ les coordonnées du point M dans le repère monde, \mathbf{P}_L la matrice de projection de la caméra gauche intégrant les paramètres intrinsèques \mathbf{K}_L et la pose \mathbf{R}_L et \mathbf{T}_L (obtenue par correspondance de points 2D/3D comme décrit dans la section B.1.3.2) et \mathbf{K}_R la matrice des paramètres intrinsèques de la caméra droite. Les coordonnées

du point \widehat{m}_L sont calculées de la même façon. Ayant à disposition k observation de points, les paramètres stéréo $\mathbf{R}_{L \rightarrow R}$ et $\mathbf{T}_{L \rightarrow R}$ sont ceux qui minimisent le critère :

$$\sum_{i=1}^k d(\widehat{m}_{R_i}, m_{R_i}) + d(\widehat{m}_{L_i}, m_{L_i}), \quad (\text{B.28})$$

B.2 Outils statistiques et mathématiques

B.2.1 Propagation d'incertitude

La notion d'incertitude est à considérer dès que nous utilisons des données mesurées ou estimées. En effet, la seule donnée de la valeur mesurée est insuffisante

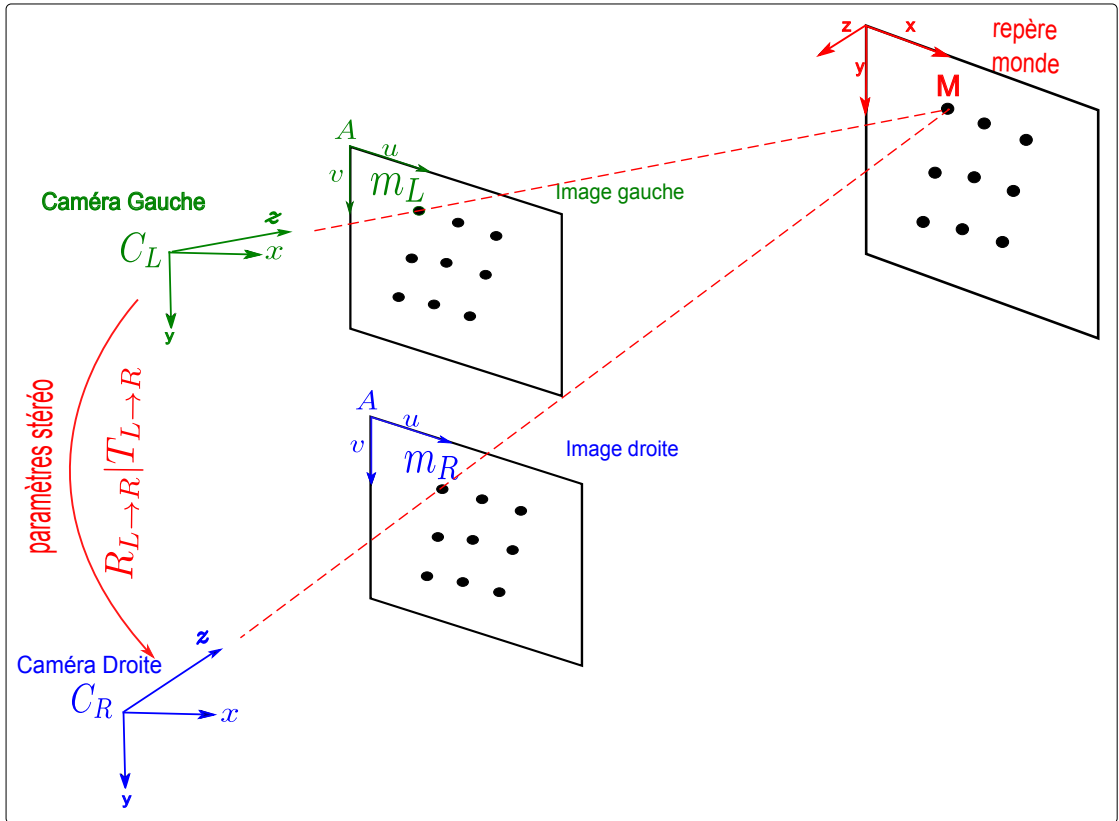


FIGURE B.4 – Étalonnage stéréo : utilisation d'une mire d'étalonnage plane. Les paramètres stéréo $\mathbf{R}_{L \rightarrow R}$ et $\mathbf{T}_{L \rightarrow R}$ sont déterminés en minimisant l'erreur de projection δ . δ exprimée pour un point M est la distance géométrique dans l'image entre la projection du point m_L dans l'image droite notée \widehat{m}_R et le point m_R réellement observé dans l'image.

et peut conduire à de mauvaises interprétations. Il est nécessaire de fournir, en plus de la valeur mesurée, son incertitude. Pour cela, une des solutions utilisées est la propagation d'incertitude. Nous proposons d'utiliser cette technique, dans ce mémoire, pour estimer l'incertitude des mesures 3D (*e.g.* distance entre deux points 3D) connaissant le modèle de bruit des points 2.5D mesurés par la caméra Temps-de-Vol.

le but de la technique de propagation d'incertitude ou d'erreur est de déterminer l'incertitude des sorties d'un processus en connaissant celle des données d'entrée. Le problème se pose comme suit :

- Soit \mathbf{x} , les données d'entrée du système, un vecteur aléatoire de moyenne $\bar{\mathbf{x}}$ et de matrice de covariance $\Sigma^{\mathbf{x}}$.
- Soit \mathbf{y} , les données de sortie du système, un vecteur aléatoire de moyenne $\bar{\mathbf{y}}$ et de matrice de covariance $\Sigma^{\mathbf{y}}$.
- Soit T la fonction modélisant le système et permettant de transformer \mathbf{x} en \mathbf{y} .

$$\mathbf{y} = T(\mathbf{x}) \quad (\text{B.29})$$

Connaissant la distribution des variables d'entrée \mathbf{x} , il nous faut déterminer la distribution des variables de sorties \mathbf{y} . Les variables d'entrée et de sortie sont supposées Gaussiennes. Il existe deux familles de méthodes pour résoudre ce problème : les méthodes de simulation et les méthodes analytiques. Dans la première famille, le système est considéré comme une boîte noire. A partir de plusieurs entrées connues, la distribution de sortie est déterminée. Ce sont des méthodes assez précises, cependant, le temps de calcul est très élevé. Il est nécessaire d'effectuer un nombre important de tirages pour obtenir un résultat fiable et le temps pour chaque tirage dépend essentiellement du temps de traitement de T . La deuxième famille suppose le modèle du système connu et s'en sert pour propager analytiquement la distribution d'entrée vers la distribution de sortie. Les méthodes de cette famille ne sont pas forcément très exactes car des approximations sont souvent effectuées. Cependant, une fois la formule de propagation établie, elles permettent de propager très rapidement l'incertitude.

Dans notre cas, le modèle du système est connu. Nous avons donc choisi d'utiliser une méthode analytique pour limiter le temps de calcul. Nous nous contentons de présenter, dans ce paragraphe, la méthode analytique que nous utilisons dans ce mémoire, appelée delta en statistique. Elle consiste à linéariser au premier ordre le modèle du système pour effectuer la propagation.

La propagation d'erreur au premier ordre est une méthode permettant de propager les erreurs des variables d'entrée vers les variables de sortie dans le cas où le modèle du système T est connu, même si celui-ci est non-linéaire ou implicite.

Le développement de Taylor de la fonction T au premier ordre autour du point étudié $(\bar{\mathbf{x}})$ est utilisé :

$$T(\mathbf{x}) = T(\bar{\mathbf{x}}) + \mathbf{J}_T(\mathbf{x} - \bar{\mathbf{x}}) + O(\|\mathbf{x} - \bar{\mathbf{x}}\|^2), \quad (\text{B.30})$$

avec \mathbf{J}_T la matrice jacobienne de T par rapport à \mathbf{x} évaluée à $\bar{\mathbf{x}}$. La moyenne est donnée par :

$$\begin{aligned} \bar{\mathbf{y}} &= \mathbf{E}(\mathbf{y}) \\ &= \mathbf{E}(T(\mathbf{x})) \\ &\approx \mathbf{E}(T(\bar{\mathbf{x}}) + \mathbf{J}_T(\mathbf{x} - \bar{\mathbf{x}})) \\ &\approx \mathbf{E}(T(\bar{\mathbf{x}})) + \mathbf{J}_T(\mathbf{E}(\mathbf{x}) - \bar{\mathbf{x}}) \\ &\approx T(\bar{\mathbf{x}}) \end{aligned} \quad (\text{B.31})$$

car $\mathbf{E}(\mathbf{x}) - \bar{\mathbf{x}} = \bar{\mathbf{x}} - \bar{\mathbf{x}} = 0$ et $\mathbf{E}(T(\bar{\mathbf{x}})) = T(\bar{\mathbf{x}})$.

La covariance $\Sigma^{\mathbf{y}}$ est donnée par :

$$\begin{aligned} \Sigma^{\mathbf{y}} &= \Sigma^{T(\mathbf{x})} \\ &\approx \Sigma^{T(\bar{\mathbf{x}}) + \mathbf{J}_T(\mathbf{x} - \bar{\mathbf{x}}) / T(\bar{\mathbf{x}}) + \mathbf{J}_T(\mathbf{x} - \bar{\mathbf{x}})} \\ &\approx \Sigma^{T(\bar{\mathbf{x}}) / T(\bar{\mathbf{x}})} + \Sigma^{T(\bar{\mathbf{x}}) / \mathbf{J}_T(\mathbf{x} - \bar{\mathbf{x}})} \\ &\quad + \Sigma^{\mathbf{J}_T(\mathbf{x} - \bar{\mathbf{x}}) / T(\bar{\mathbf{x}})} + \Sigma^{\mathbf{J}_T(\mathbf{x} - \bar{\mathbf{x}}) / \mathbf{J}_T(\mathbf{x} - \bar{\mathbf{x}})} \\ &\approx \mathbf{J}_T \Sigma^{\mathbf{x} - \bar{\mathbf{x}}} \mathbf{J}_T^{\top} \\ &\approx \mathbf{J}_T \Sigma^{\mathbf{x}} \mathbf{J}_T^{\top} \end{aligned} \quad (\text{B.32})$$

Et ainsi :

$$\Sigma^{\mathbf{y}} \approx \mathbf{J}_T \Sigma^{\mathbf{x}} \mathbf{J}_T^{\top} \quad (\text{B.33})$$

B.2.2 Fonction de type *B-spline*

Les *B-splines* sont des courbes polynomiales par morceaux, construites à partir d'un ensemble de points de contrôle. Ces fonctions paramétriques sont utilisées dans les méthodes de l'état de l'art des approches d'étalonnage de la mesure de profondeur. Nous les présentons brièvement, plus de détails peuvent être trouvés dans le livre [BOOR 2001].

Soit p le degré des polynômes utilisés et $n + 1$ le nombre de points de contrôle. Le i ème polynôme s'écrit : $L_i(t) = B_0(t)\mathbf{c}_i + B_1(t)\mathbf{c}_{i+1} + \dots + B_p(t)\mathbf{c}_{i+p}$, avec $i \in [0, n - p]$ et $t \in [0, 1]$. Les $B_i(t)$ forment une base de polynômes de degré p . Cette base doit respecter certaines conditions : les courbes doivent se rejoindre

au point de raccordement tout en assurant la continuité de leurs dérivées jusqu'à l'ordre $p - 1$. De plus la somme des coefficients $\sum_{i=0}^p B_i(t)$ est égale à 1 $\forall t \in [0, 1]$. Ces conditions permettent d'estimer les polynômes $B_i(t)$. Pour $p = 3$, on obtient après calcul la base de polynômes suivante :

$$\begin{aligned} B_0(t) &= \frac{1}{6}(-t^3 + 3t^2 - 3t + 1) & B_1(t) &= \frac{1}{6}(3t^3 - 6t^2 + 4) \\ B_2(t) &= \frac{1}{6}(-3t^3 + 3t^2 + 3t + 1) & B_3(t) &= \frac{1}{6}t^3. \end{aligned} \quad (\text{B.34})$$

Ce type de fonction a une influence spatiale spatiale. Cela est illustré par un exemple de B-spline 2D obtenu par produit tensoriel extrait de [Gay-Bellile 2008]. Un exemple synthétique est généré illustré par la figure B.5a. Il consiste à appliquer une rotation de 45 degrés (dans le sens direct), à un carré de dimension 2×2 d'une grille régulière de taille 10×10 . Les coins de la grille sont, quant à eux, fixes. La figure B.5b montre la déformation de la grille par ce modèle de déformation synthétique. L'ensemble des nœuds constituant la grille est utilisé comme points de contrôle illustré par la figure B.5c. La région d'intérêt est déformée en utilisant une fonction de type B-spline. La figure B.5d montre que la déformation engendrée est locale : seule la région où les points de contrôle sont déplacés ainsi que son voisinage proche sont déformés.

B.2.3 Fonction spline de type plaque mince TPS 3D

La fonction TPS-3D $\mathbb{R}^3 \rightarrow \mathbb{R}$ est une fonction lisse ([May et al. 2009]) connue pour être une bonne approximation pour plusieurs déformations qui minimise l'énergie de courbure ("bending energy").

Cette fonction est flexible, contrôlée par l centres 3D \mathbf{c}_k ($\mathbf{c}_k \in \mathbb{R}^3, k = 1, \dots, l$) qui peuvent être placés n'importe où dans l'espace et est basée sur l'association des valeurs cibles α_k aux l centres 3D \mathbf{c}_k . Elle est généralement paramétrée par un vecteur de $l + 4$ coefficients $\mathbf{h}^\top = (\mathbf{w}^\top \mathbf{a}^\top)$ calculé à partir des valeurs cibles α (détaillé dans la suite) et par un paramètre de régularisation $\lambda \in \mathbb{R}^+$. Il y a l coefficients dans \mathbf{w} et 4 coefficients dans \mathbf{a} . Les coefficients dans \mathbf{w} doivent satisfaire la condition suivante $\tilde{\mathbf{P}}^\top \mathbf{w} = \mathbf{0}$, où la k^{me} ligne de $\tilde{\mathbf{P}}^\top$ est donnée par $(\mathbf{c}_k^\top \mathbf{1})$. Ces quatre "conditions aux bords" permettent d'assurer que les équations possèdent des dérivées secondes intégrables deux fois. Soit :

$$\ell_{\mathbf{q}}^\top = ((d(\mathbf{q}, \mathbf{c}_1)) \cdots (d(\mathbf{q}, \mathbf{c}_l)) \mathbf{q}^\top \mathbf{1}), \quad (\text{B.35})$$

où d est la distance euclidienne, la fonction TPS 3D à un point \mathbf{q} (3 composantes)

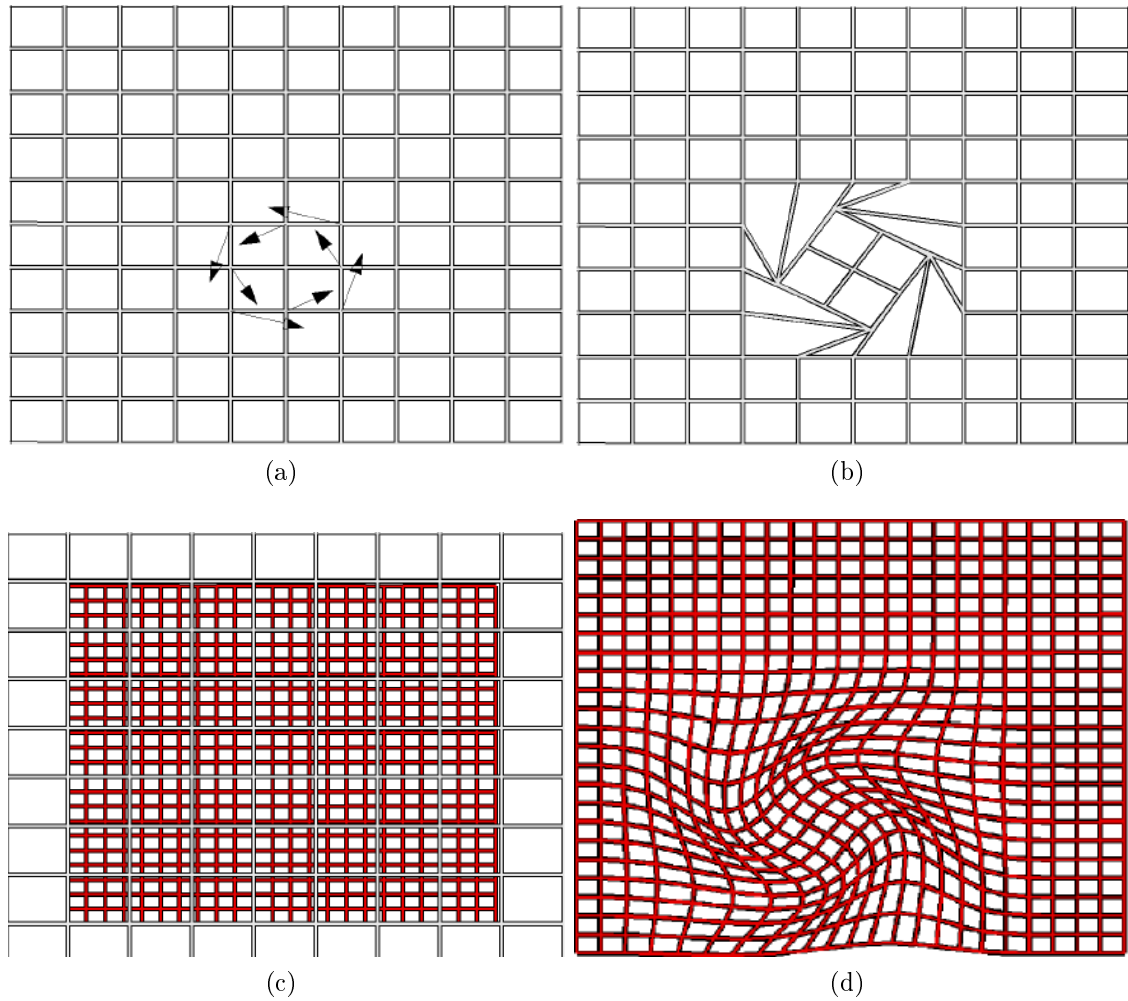


FIGURE B.5 – Exemple de déformation par fonction de type B-spline de dimension 2D. (a) Modèle de la déformation synthétique : rotation d’une région d’intérêt de taille 2×2 d’une grille dont les coins sont fixés. (b) Déformation de la grille par ce modèle synthétique. (c) Déformation estimée avec la fonction B-spline. (*extrait de [Gay-Bellile 2008]*)

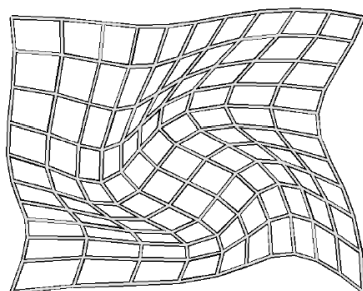
est donnée par :

$$\omega(\mathbf{q}, \mathbf{h}) = \ell_{\mathbf{q}}^{\top} \mathbf{h} = \left(\sum_{k=1}^l \mathbf{w}_k d(\mathbf{q}, \mathbf{c}_k) \right) + \mathbf{a}^{\top} \check{\mathbf{q}}. \quad (\text{B.36})$$

Ce type de fonction a une influence spatiale globale. Cela est illustré par un exemple de la fonction TPS 2D extrait de [Gay-Bellile 2008]. Plus de détails sur

les TPS 2D peuvent être trouvés dans [Bookstein 1989]. L'exemple de la figure B.5 est repris sur la figure B.6. Les centres de la fonction TPS 2D sont les 8 points constituant le carré et les 4 coins de la grille. La figure B.5b montre la déformation de la grille par ce modèle le déformation synthétique. La déformation générée par la fonction TPS 2D est illustré par la figure B.6a. Elle déforme l'ensemble de la région d'intérêt, c'est-à-dire la région mise en mouvement ainsi que son voisinage proche.

Les fonction de type plaque mince, à influence globale, est utilisé pour l'étude de surfaces telles que les feuilles de papiers, les vêtements, les tissus [Bartoli *et al.* 2007, Bartoli & Zisserman 2004, Gay-Bellile *et al.* 2007, Gay-Bellile *et al.* 2006, Vedaldi & Soatto 2006] où sa propriété de déformation intrinsèquement lisse est bien adaptée. Il est aussi employé en imagerie médicale pour le recalage d'images du cerveau [Johnson & Christensen 2002]. Les fonctions plaque mince sont utilisées à deux reprises dans ce mémoire. La première fois dans le chapitre 3 pour modéliser le bruit des caméras Temps-de-Vol. La deuxième fois dans le chapitre 5 dans la deuxième méthode d'étalonnage de la mesure de profondeur.



(a)

FIGURE B.6 – Exemple de déformation par fonction de type plaque mince de dimension 2D. (*extrait de [Gay-Bellile 2008]*)

B.3 Transformations utilisées

Dans ce mémoire, deux méthodes d'étalonnage de la mesure de profondeur sont proposées. Notons $\mathbf{q} = (u \ v \ d)$ un point 2.5D où $(u \ v)$ représente les coordonnées pixélique du point et d sa profondeur et $\mathbf{Q} = (X \ Y \ Z)$ le point 3D correspondant. La première méthode proposée opère dans l'espace 2.5D, c'est-à-dire que la FCD (Fonction de Correction de la Distorsion) associée à chaque point \mathbf{q} la correction correspondant, tandis que, la deuxième méthode opère dans

l'espace 3D. Dans cette section, nous montrons que les deux espaces 2.5D et 3D sont équivalents.

B.3.1 Transformation de l'espace 2.5D vers l'espace 3D

La transformation d'un point \mathbf{q} vers un point \mathbf{Q} dans l'espace métrique, connaissant les paramètres intrinsèques de la caméra est donnée par la relation suivante (comme le montre la figure B.7) :

$$\begin{aligned} u &= \frac{f_c}{d_u} \frac{X}{Z} + c_u, \\ v &= \frac{f_c}{d_v} \frac{Y}{Z} + c_v, \\ d &= Z \sqrt{\frac{f_c^2 + ((u-c_u)d_u)^2 + ((v-c_v)d_v)^2}{f_c^2}}, \end{aligned} \quad (\text{B.37})$$

où $(c_u \ c_v)$ est le centre optique dans le plan image, f_c est la distance focale de la caméra et $(d_u \ d_v)$ la taille d'un pixel respectivement dans les directions u, v . Cette transformation est exprimée en négligeant la distorsion optique. De même la transformation entre \mathbf{Q} et \mathbf{q} est donnée par :

$$\begin{aligned} Z &= d \frac{f_c}{\sqrt{f_c^2 + ((u-c_u)d_u)^2 + ((v-c_v)d_v)^2}}, \\ X &= Z \frac{(u-c_u)d_u}{f_c}, \\ Y &= Z \frac{(v-c_v)d_v}{f_c}. \end{aligned} \quad (\text{B.38})$$

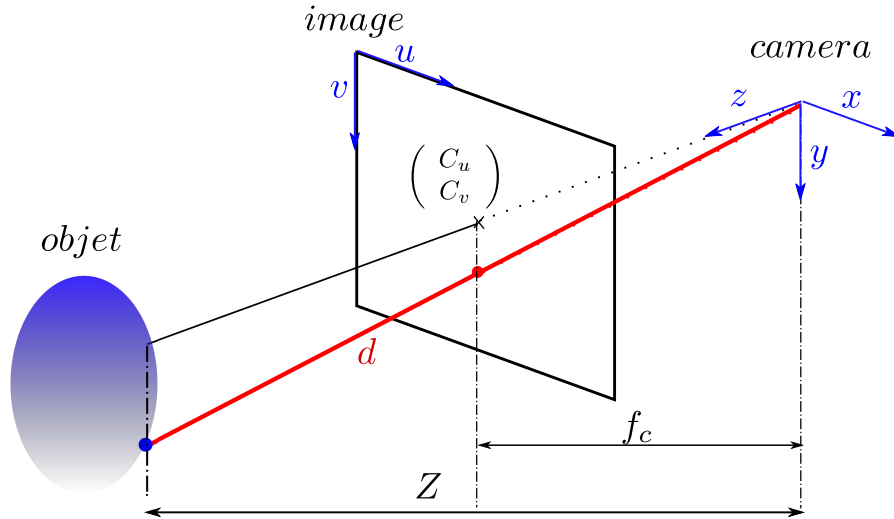


FIGURE B.7 – Transformation d'un point 2.5D \mathbf{q} vers un point 3D \mathbf{Q} .

B.3.2 Transformation entre les fonctions de correction f et F

La fonction de correction de la distorsion de la mesure de profondeur des caméras Temps-de-Vol est généralement modélisée par une fonction $f : \Omega \rightarrow \mathbb{R}$, $\Omega \subset \mathbb{R}^3$ qui associe à chaque point 2.5D \mathbf{q} de l'espace la correction correspondante telle que : $f(\mathbf{q}) = c$ où Ω est un sous ensemble de \mathbb{R}^3 : $\Omega = [u_{min}; u_{max}] \times [v_{min}; v_{max}] \times [d_{min}; d_{max}]$ et c est un scalaire est présente la correction de la distorsion. D'une autre façon, $d + c$ est la valeur de la profondeur après correction. La méthode proposée dans la section 5.2 du chapitre 5 estime une fonction équivalente que nous notons F qui opère sur les points 3D \mathbf{Q} . Nous avons démontré dans la section B.3.1 qu'il existe une transformation entre les deux espaces 2.5D et 3D. En utilisant cette transformation, estimer la fonction f ou la fonction F est équivalent puisque c peut être obtenu à partir de $F(\mathbf{Q})$ par une simple transformation :

$$c = f(\mathbf{q}) = F(\mathbf{Q}) \frac{f_c}{\sqrt{f_c^2 + ((u - c_u)d_u)^2 + ((v - c_v)d_v)^2}}. \quad (\text{B.39})$$

Bibliographie

- [Bartoli & Zisserman 2004] A. Bartoli et A. Zisserman. *Direct Estimation of Non-Rigid Registrations*. In British Machine Vision Conference, 2004.
- [Bartoli *et al.* 2007] A. Bartoli, M. Perriollat et S. Chambon. *Generalized Thin-Plate Spline Warps*. In Computer Vision and Pattern Recognition, 2007.
- [Belhedi *et al.* 2012a] A. Belhedi, A. Bartoli, S. Bourgeois, K. Hamrouni, P. Sayd et V. Gay-Bellile. *Noise Modelling and Uncertainty Propagation of a TOF camera*. In European Conference on Computer Vision Workshop on The Quest for 3D, 2012.
- [Belhedi *et al.* 2012b] A. Belhedi, A. Bartoli, V. Gay-Bellile, S. Bourgeois, P. Sayd et K. Hamrouni. *Depth correction for depth cameras from planarity*. In British Machine Vision Conference, 2012.
- [Belhedi *et al.* 2012c] A. Belhedi, S. Bourgeois, V. Gay-Bellile, P. Sayd, A. Bartoli et K. Hamrouni. *Non-parametric depth calibration of a TOF camera*. In International Conference on Image Processing, 2012.
- [Biswas & Veloso 2012] J. Biswas et M. Veloso. *Depth camera based indoor mobile robot localization and navigation*. In International Conference on Robotics and Automation, pages 1697–1702, 2012.
- [Bookstein 1989] F.L. Bookstein. *Principal Warps : Thin-Plate Splines and the Decomposition of Deformations*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 6, pages 567–585, June 1989.
- [BOOR 2001] C. DE BOOR. A practical guide to splines. Springer, 2001.
- [Bouguet 2010] Jean-Yves Bouguet. *Camera Calibration Toolbox for Matlab*, 2010.

- [Castaneda *et al.* 2011] V. Castaneda, D. Mateus et N. Navab. *SLAM combining TOF and high-resolution cameras*. In Proceedings of the 2011 IEEE Workshop on Applications of Computer Vision, WACV '11, 2011.
- [Chambers *et al.* 1983] J. M. Chambers, W. S. Cleveland, B. Kleiner et P. A. Tukey. *Graphical methods for data analysis*. Wadsworth, 1983.
- [Cho *et al.* 2007] J.H. Cho, I.Y. Chang, S.M. Kim et K.H. Lee. *Depth Image Processing Technique for Representing Human Actors in 3DTV using Single Depth Camera*. In 3DTV, 2007.
- [Choi & Lee 2012] O. Choi et S. Lee. *Wide range stereo Time-Of-Flight camera*. In International Conference on Image Processing, 2012.
- [Choi *et al.* 2010] O. Choi, H. Lim, B. Kang, Y. S. Kim, K. Lee, J. D. K. Kim et C.Y. Kim. *Range unfolding for Time-Of-Flight depth cameras*. In International Conference on Image Processing, 2010.
- [Dong & Kamat 2010] S. Dong et V. R. Kamat. *Resolving incorrect visual occlusion in outdoor augmented reality using TOF camera and OpenGL frame buffer*. In International Conference on Construction Applications of Virtual Reality, pages 55–63, 2010.
- [Droeschel *et al.* 2010a] D. Droeschel, D. Holz et S. Behnke. *Multifrequency phase unwrapping for Time-Of-Flight cameras*. In Intelligent Robots and Systems, 2010.
- [Droeschel *et al.* 2010b] D. Droeschel, D. Holz et S. Behnke. *Probabilistic phase unwrapping for Time-Of-Flight cameras*. In Joint International Symposium on Robotics and German, 2010.
- [Edeler *et al.* 2010] T. Edeler, K. Ohliger, S. Hussmann et A. Mertins. *Time-Of-Flight depth image denoising using prior noise information*. In In Circuit Serial Programming, 2010.
- [Falie & Buzuloiu 2007] D. Falie et V. Buzuloiu. *Noise Characteristics of 3D Time-Of-Flight Cameras*. In International Symposium on Signals, Circuits and Systems, 2007.
- [Faugeras 1993] O. Faugeras. *Three-dimensional computer vision : a geometric viewpoint*. MIT Press, 1993.
- [Garro *et al.* 2009] V. Garro, C. Dal Mutto, P. Zanuttigh et G. M. Cortelazzo. *A novel interpolation scheme for range data with side information*. In Conference for Visual Media Production, 2009.
- [Gay-Bellile *et al.* 2006] V. Gay-Bellile, M. Perriollat, A. Bartoli et P. Sayd. *Image Registration by Combining Thin-Plate Splines With a 3D Morphable Model*. In International Conference on Image Processing, 2006.

- [Gay-Bellile *et al.* 2007] V. Gay-Bellile, A. Bartoli et P. Sayd. *Feature-Driven Non-Rigid Image Registration*. In British Machine Vision Conference, 2007.
- [Gay-Bellile 2008] V. Gay-Bellile. *Contribution au recalage et à la reconstruction 3D de surfaces déformables*. PhD thesis, Université Blaise Pascal, France, 2008.
- [Guömundsson *et al.* 2007] S. Á. Guömundsson, H. Aanæs et R. Larsen. *Environmental effects on measurement uncertainties of Time-Of-Fight cameras*. In International Symposium on Signals, Circuits and Systems, 2007.
- [Guömundsson *et al.* 2008] S. Á. Guömundsson, H. Aanæs et R. Larsen. *Fusion of stereo vision and Time Of Flight imaging for improved 3D estimation*. Int J. Intelligent Systems Technologies and Applications, vol. 5, no. 3, pages 425–433, 2008.
- [Hartley & Zisserman 2004] R. Hartley et A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, second édition, 2004.
- [Huhle *et al.* 2007] B. Huhle, S. Fleck et A. Schilling. *Integrating 3D Time-Of-Flight camera data and high resolution images for 3DTV applications*. In 3DTV, 2007.
- [Huhle *et al.* 2010] B. Huhle, T. Schairer, P. Jenke et W. Strasser. *Fusion of range and color images for denoising and resolution enhancement with a non-local filter*. In Computer Vision and Image Understanding, 2010.
- [Johnson & Christensen 2002] H. J. Johnson et G. E. Christensen. *Consistent Landmark and Intensity-Based Image Registration*. IEEE Transactions on Medical Imaging, vol. 21, no. 5, pages 450–461, May 2002.
- [Jutzi 2009] B. Jutzi. *Investigations on ambiguity unwrapping of range images*. In International Archives of Photogrammetry and Remote Sensing Workshop on Laserscanning, 2009.
- [Kahlmann *et al.* 2006] T. Kahlmann, F. Remondino et H. Ingensand. *Calibration for increased accuracy of the range imaging camera SwissRangerTm*. In Image Engineering and Vision Metrology, 2006.
- [Karel *et al.* 2007] W. Karel, P. Dorninger et N. Pfeifer. *In situ determination of range camera quality parameters by segmentation*. In Optical 3-D Measurement Techniques, 2007.
- [Kim *et al.* 2006] S. M. Kim, J. Cha, J. Ryu et K. H. Lee. *Depth Video Enhancement for Haptic Interaction Using a Smooth Surface Reconstruction*. IEICE Transaction on Information and Systems, vol. E89-D, pages 37–44, 2006.

- [Kim *et al.* 2010] S.Y. Kim, J.H. Cho, A. Koschan et M. Abidi. *Spatial and Temporal Enhancement of Depth Images Captured by a Time-Of-Flight Depth Sensor*. In International Conference on Pattern Recognition, 2010.
- [Kuhnert & Stommel 2006] K. D. Kuhnert et M. Stommel. *Fusion of stereo-camera and PMD-camera data for real-time suited precise 3D environment reconstruction*. In International Radar Symposium, 2006.
- [Lange 2000] R. Lange. *3D Time-Of-Flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology*. PhD thesis, University of Siegen, Germany, 2000.
- [Lindner & Kolb 2006] M. Lindner et A. Kolb. *Lateral and Depth calibration of PMD-distance sensors*. In International Symposium on Visual Computing, 2006.
- [Lindner & Kolb 2007] M. Lindner et A. Kolb. *Calibration of the intensity-related distance error of the PMD TOF-Camera*. In Intelligent Robots and Computer Vision, 2007.
- [Lindner *et al.* 2010] M. Lindner, I. Schiller, A. Kolb et R. Koch. *Time-Of-Flight sensor calibration for accurate range sensing*. Computer Vision and Image Understanding, vol. 114, pages 1318–1328, 2010.
- [May *et al.* 2009] S. May, D. Droeschel, D. Holz et S. Fuchs. *Three-dimensional mapping with Time-Of-Flight cameras*. J. Field Robot, vol. 26, pages 934–964, 2009.
- [McClure *et al.* 2010] S. H. McClure, M. J. Cree, A. A. Dorrington et A. D. Payne. *Resolving depth measurement ambiguity with commercially available range imaging cameras*. In Image Processing : Machine Vision Applications III, 2010.
- [Mutto *et al.* 2010] C. Dal Mutto, P. Zanuttigh et G. M. Cortelazzo. *A probabilistic approach to TOF and stereo data fusion*. In 3D Data Processing, Visualization, and Transmission, 2010.
- [Nguyen *et al.* 2012] V. C. Nguyen, S. Izadi et D. Lovell. *Modeling Kinect Sensor Noise for Improved 3D Reconstruction and Tracking*. In 3D Imaging, Modeling, Processing, Visualization and Transmission, 2012.
- [Park *et al.* 2011] J. Park, H. Kim, Y.W. Tai, M.S. Brown et I. Kweon. *High quality depth map upsampling for 3D-TOF cameras*. In International Conference on Computer Vision, 2011.
- [Payne *et al.* 2009] A.D. Payne, A.P.P. Jongenelen, A.A. Dorrington, M.J. Cree et D.A. Carnegie. *Multiple frequency range imaging to remove measurement ambiguity*. In Optical 3-D Measurement Techniques, 2009.

- [PMDTechnologies 2013] PMDTechnologies. *PMD[vision] CamBoard pico*. www.pmdtec.com/news_media/news/pico_meets.php, 2013.
- [Poppinga & Birk 2009] J. Poppinga et A. Birk. *A Novel Approach to Efficient Error Correction for the SwissRanger Time-Of-Flight 3D Camera*. In RoboCup, 2009.
- [Prasad *et al.* 2006] M. Prasad, A. Zisserman et A. Fitzgibbon. *Single view reconstruction of curved surfaces*. In Computer Vision and Pattern Recognition, 2006.
- [PrimeSense 2013] PrimeSense. *Kinect*. www.primesense.com/solutions/sensor/, 2013.
- [Schiller *et al.* 2008] I. Schiller, C. Beder et R. Koch. *Calibration of a PMD-Camera using a Planar Calibration Pattern together with a Multi-Camera Setup*. In International Society of Photogrammetry and Remote Sensing, 2008.
- [Spirig *et al.* 1995] T. Spirig, P. Seitz, O. Vietze et F. Heitger. *The lock-in CCD-two-dimensional synchronous detection of light*. IEEE Journal of Quantum Electronics, vol. 31, no. 9, pages 1705–1708, 1995.
- [Vedaldi & Soatto 2006] A. Vedaldi et S. Soatto. *Local Features, All Grown Up*. In Computer Vision and Pattern Recognition, 2006.
- [Weyer *et al.* 2008] C. A. Weyer, K. H. Bae, K. Lim et D. D. Lichti. *Extensive metric performance evaluation of a 3D range camera*. In International Society of Photogrammetry and Remote Sensing, 2008.
- [Yang *et al.* 2007a] Q. Yang, R. Yang, J. Davis et D. Nistér. *Spatial-Depth Super Resolution for Range Images*. In Computer Vision and Pattern Recognition, 2007.
- [Yang *et al.* 2007b] Q. Yang, R. Yang, J. Davis et D. Nistr. *Spatial-depth super resolution for range images*. In Computer Vision and Pattern Recognition, 2007.
- [Yang *et al.* 2010] Q. Yang, K.H. Tan, B. Culbertson et J. Apostolopoulos. *Fusion of active and passive sensors for fast 3D capture*. In Multimedia Signal Processing, 2010.
- [Zhang *et al.* 2012] L. Zhang, J. Sturm, D. Cremers et D. Lee. *Real-time human motion tracking using multiple depth cameras*. In International Conference on Intelligent Robots and Systems, pages 2389–2395, 2012.
- [Zhu *et al.* 2008a] J. Zhu, L. Wang, R. Yang et J. Davis. *Fusion of Time-Of-Flight depth and stereo for high accuracy depth maps*. In Computer Vision and Pattern Recognition, 2008.

- [Zhu *et al.* 2008b] J. Zhu, L. Wang, R. Yang et J. Davis. *Fusion of Time-Of-Flight depth and stereo for high accuracy depth maps*. In Computer Vision and Pattern Recognition, 2008.
- [Zhu *et al.* 2010] J. Zhu, L. Wang, J. Gao, R. Yang et J. Davis. *Spatial-temporal fusion for high accuracy depth maps using dynamic MRFs*. IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 32, no. 5, pages 899–909, 2010.
- [Zhu *et al.* 2011] J. Zhu, L. Wang, J. E. Davis et Z. Pan. *Reliability fusion of Time-Of-Flight depth and stereo geometry for high quality depth maps*. IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 33, no. 7, pages 1400–1414, 2011.