

INSTITUT DE MICROTECHNIQUE
UNIVERSITÉ DE NEUCHÂTEL



Contributions to image processing algorithms for advanced 3D vision devices

Thèse présentée à la faculté des sciences
pour l'obtention du grade de docteur ès sciences

Par

James Mure-Dubois

Soutenue le 25 Août 2009

Acceptée sur proposition du jury:

Prof. Heinz Hügli (Université de Neuchâtel), directeur de thèse,
Dr. Alain Codourey (Asyrl SA, Villaz-St-Pierre),
Prof. Pierre-André Farine (Université de Neuchâtel),
Prof. Nadine Piat (École nationale supérieure de mécanique et des
microtechniques de Besançon).

IMPRIMATUR POUR LA THÈSE

Contributions to image processing algorithms for
advanced 3D vision devices

James MURE-DUBOIS

UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES

La Faculté des sciences de l'Université de Neuchâtel,
sur le rapport des membres du jury

Mme N. Le Fort-Piat (Besançon F),
MM. H. Hügli (directeur de thèse), P.-A. Farine
et A. Codourey (Asyril SA, Suisse)

autorise l'impression de la présente thèse.

Neuchâtel, le 27 août 2009

Le doyen :
F. Kessler

UNIVERSITE DE NEUCHATEL
FACULTE DES SCIENCES
Secrétariat - décanat de la faculté
Rue Emile-Argand 11 - CP 158
CH-2009 Neuchâtel
Felix Kessler

Summary

This thesis is a contribution to the wide and steadily evolving field of 3D vision. The presented work focuses on the role played by image processing algorithms employed in specific 3D vision systems and in challenging 3D applications. Two very different 3D imaging systems are considered: a 3D microscope based on the depth from focus principle, and a real-time 3D camera using time-of-flight (TOF) measurements. A concise presentation of the operating principle for recovery of depth information allows to highlight some critical aspects relating to the device performance and to propose new approaches and algorithms in challenging imaging applications.

The first part of the thesis is devoted to the design of a new 3D sensor for operation in an embedded vision system for micro-assembly. The principle of depth from focus is selected and implemented with a miniature microscopic device. The study shows that in this case, most of the technological limitations are related to the size of the optical components required for high measurement accuracy, while algorithms for image processing can easily be scaled to reach real-time operation.

A second part of the thesis is devoted to the study of time-of-flight cameras, with special emphasis on image processing algorithms for data error reduction. Several sources of errors are presented and analyzed. In contrast to previous studies devoted to error reduction at the device level, the present work analyzes commercialized hardware qualitatively and quantitatively, and aims to improve the measured data by image processing algorithms. The multiple sources of errors, like stochastic noise, multipath effects and scattering effects are considered. Then, a full chapter is devoted to a detailed analysis of the *scattering* phenomenon observed in TOF devices. This analysis allows to define a strategy for scattering compensation based on image filtering. Different implementations are compared, and a solution based on filtering in the Fourier domain is selected, as it provides superior speed for the compensation filter.

The last problem studied in this thesis concerns the simultaneous operation of multiple TOF cameras. Compared to a single camera, a multi-camera system allows to avoid occlusions in the acquired data, and provides generally an extended field of view. But the different views have to be registered in order to be exploited as a whole. This thesis includes a comparison of registration algorithms on real scenes acquired with two TOF cameras. Several state of the art registration methods are considered and evaluated experimentally. Medium to poor registration performance is observed and is clearly explained by the strong noise content of typical TOF images. Therefore, a best suited and simple registration method involving the extraction and matching of a plane region common in the two views is presented. The experiments performed verified that the proposed technique is well suited to current TOF acquisition devices.

Finally, a specific example among the new applications enabled by the combination of TOF cameras and advanced image processing algorithms is discussed: surveillance systems. Compared to conventional video data, range data allows for easier segmentation and scene interpretation. In various situations, scattering compensation allows to avoid erroneous measurements, while camera registration ensures a complete view of the scene under surveillance. Those improvements make surveillance by TOF cameras an attractive alternative to current systems based on conventional (i.e. 2D) imagers.

Keywords

3D vision, range images, 3D image processing, registration, range image registration, time-of-flight camera, TOF, depth from focus, microvision, scattering, scattering compensation, multi-camera systems, surveillance.

Résumé

Cette thèse est une contribution au champ vaste et dynamique de la vision 3D. Le travail présenté s'intéresse au rôle joué par les algorithmes de traitement d'image dans des systèmes de vision 3D spécifiques ainsi que dans des applications novatrices. Deux systèmes d'imagerie 3D très différents sont considérés : un microscope 3D basé sur le principe *depth from focus*, et une caméra 3D à temps réel utilisant des mesures temps de vol (*time-of-light* ou TOF). Une présentation concise du principe de récupération des informations de profondeur permet de souligner certains aspects critiques pour la performance des systèmes et de proposer de nouvelles approches et de nouveaux algorithmes pour des applications exigeantes.

La première partie de cette thèse est consacrée à la conception d'un nouveau capteur 3D destiné à un système de vision embarqué pour le micro-assemblage. Le principe *depth from focus* est sélectionné et implémenté à l'aide d'un dispositif microscope miniature. L'étude montre que dans ce cas, la plupart des limites technologiques sont liées à la taille des composants optiques requis pour des mesures à haute résolution, tandis que les algorithmes de traitement d'image sont facilement adaptés pour atteindre le fonctionnement en temps réel.

Une seconde partie est dédiée à l'étude des caméras temps de vol, en mettant l'accent sur les algorithmes de traitement d'image pour la réduction des erreurs de mesure. Plusieurs sources d'erreurs sont présentées et analysées. Alors que les précédentes études s'attaquaient à la réduction d'erreur au niveau du matériel, le présent travail analyse qualitativement et quantitativement un dispositif déjà commercialisé et cherche à améliorer les données mesurées à l'aide d'algorithmes de traitement d'image. Les différentes sources d'erreur, comme le bruit, les effets de réflexions multiples et la diffusion (*scattering*), sont prises en compte. Un chapitre entier est consacré à une analyse détaillée du phénomène de *scattering* observé avec les caméras TOF. Cette analyse nous permet de définir une stratégie pour la compensation du *scattering* basée sur le filtrage d'image. Différentes implémentations sont comparées ; c'est une solution basée sur le filtrage dans le domaine Fourier qui est finalement préférée, dans la mesure où elle permet de meilleures performances en termes de vitesse pour le filtre de compensation.

Le dernier problème étudié dans cette thèse concerne l'utilisation simultanée de plusieurs caméras TOF. Comparé à une caméra unique, un système multi-caméras permet d'éviter les occlusions dans les données acquises, et fournit généralement un champ de vision plus étendu. Mais les différentes vues doivent être alignées pour pouvoir être exploitées comme un tout. Cette thèse comporte une comparaison entre différents algorithmes d'alignement, basée sur des scènes réelles acquises par deux caméras TOF. Plusieurs méthodes courantes sont présentées et évaluées expérimentalement. Les performances observées restent médiocres, ce qui s'explique par le bruit important que comportent les images TOF. Pour y remédier, cette thèse propose une méthode d'alignement simple et mieux adaptée, basée sur l'extraction et l'appariement d'une région plane commune aux deux vues. L'expérimentation montre que cette technique est bien adaptée aux caméras TOF actuellement disponibles.

Finalement, nous discutons d'un exemple particulier parmi les nouvelles applications permises par la combinaison de caméras TOF et d'algorithmes de traitement d'image avancés : les systèmes de surveillance. Par rapport aux données vidéo conventionnelles, les images de profondeur permettent une segmentation et une interprétation de la scène plus aisée. Dans de nombreuses situations, la compensation du *scattering* permet d'éviter des mesures erronées, tandis que l'alignement de caméras garantit

une vue complète de la scène sous surveillance. Ces améliorations font de la surveillance par caméras TOF une alternative avantageuse aux systèmes actuels basés sur des caméras 2D conventionnelles.

Mots clés

3D vision, range images, 3D image processing, registration, range image registration, time-of-flight camera, TOF, depth from focus, microvision, scattering, scattering compensation, multi-camera systems, surveillance.

Remerciements

Arriver au bout de cette thèse n'a pas toujours été facile, cette entreprise n'aurait probablement pas été couronnée de succès sans la perspicacité, le soutien et la patience des personnes qui m'ont entouré.

Tout d'abord, je veux exprimer ma gratitude envers Heinz Hügli, qui m'a donné l'opportunité de rejoindre son équipe. Heinz a toujours été présent pour m'aider à faire ce qu'un doctorant doit fréquemment faire : mettre les choses en perspective. Ses suggestions, revenant souvent à l'analyse des questions essentielles, parfois basiques, m'ont évité de me perdre trop longtemps dans des voies sans issue. J'ai beaucoup apprécié de bénéficier des vastes connaissances techniques de Heinz, qui sont loin de se cantonner au seul domaine de la vision par ordinateur. La minutie avec laquelle Heinz relit les publications m'a aidé à sensiblement améliorer mon style d'écriture. De plus, avoir eu l'occasion d'assister Heinz pour les cours de traitement des images et de microprocesseurs m'a permis de saisir toute l'étendue de la différence entre la maîtrise technique d'un sujet d'une part, et la maîtrise de la transmission de ce savoir à une classe d'autre part. Merci Heinz pour ces quatre années inoubliables.

Je voudrais également remercier le professeur Pierre-André Farine, qui a eu une influence non négligeable sur l'orientation de ma carrière scientifique. Il m'a offert l'opportunité d'aller réaliser un stage chez Logitech pour mon travail de diplôme. C'est à partir de ce moment que ma spécialisation en traitement du signal s'est dessinée. Lorsqu'il a pris la tête du Parlab, pour ma dernière année de doctorat, le professeur Farine a su s'assurer que toutes les conditions soient requises pour me permettre de me concentrer sur mon travail de thèse. De plus, le professeur Farine a également eu la patience de lire et d'évaluer cette thèse, et a fourni des commentaires intéressants lors de la soutenance. Ma reconnaissance va aussi au professeur Nadine Piat, ainsi qu'au docteur Alain Codourey, qui ont accepté de bonne grâce de lire et d'évaluer cette thèse, même si certaines parties sortaient parfois du domaine de leurs intérêts de recherche. Je les remercie pour le temps investi ainsi que pour leurs questions et remarques pertinentes.

Les collaborations fructueuses autour des projets auxquels j'ai participé ont contribué à financer cette thèse, mais surtout à en enrichir le contenu. Je remercie Alexander Steinecker et Alain Codourey, à l'époque au CSEM Alpnach, pour leur apport concernant le côté robotique du projet CTI MiniVision. Ce projet a par la suite conduit à des échanges intéressants avec Guillaume Fortier, Soukalo Dembelé et Nadine Piat du Laboratoire d'Automatique de Besançon. Merci également à Claus Urban de Heliotis SA de m'avoir présenté les rudiments de l'imagerie pOCT. Au cours du projet CTI Perspass, Jan Holzbecher, Yannick Blondeau, Skye Legon et Philippe Niederhauser m'ont fourni un feedback précieux. Je remercie aussi toute l'équipe de Mesa Imaging SA, notamment Thierry Oggier, Thierry Zamofing et Bernhard Büttgen, qui a toujours bien accueilli mes questions concernant les caméras SwissRanger. Ce projet a également conduit à des échanges plus informels, mais toujours très enrichissants, avec Timo Kahlmann à l'ETHZ, ainsi que Trine Kirkhus et Tom Kavli de SINTEF. Finalement, Brienna Putz et Ken Weible de SÜSS MicroOptics SA m'ont permis de découvrir une application intéressante de l'imagerie microscopique pour l'inspection.

Au cours des années passées au Parlab, les collègues que j'ai côtoyés ont toujours maintenu une ambiance agréable et constructive, au travail comme en dehors. Nabil Ouerhani a contribué à souder l'équipe, par exemple en nous enrôlant dans les tournois de football en salle. L'exemple de Thierry Zamofing m'a beaucoup aidé pour progresser au point de vue de la programmation, il m'a incité à utiliser les techniques de

contrôle de source et de backup qui m'ont sauvé la mise plus d'une fois. Grâce à Iva Bogdanova, j'ai appris que les connaissances de mathématiques et de géométrie oubliées peuvent parfois ressurgir pour aider à saisir les subtilités du traitement d'images sur la sphère, ou sur des coniques (hyperboles ou paraboles), qui se ramènent adroitement au cas précédent.

Mention spéciale à Alexandre Bur, qui m'a supporté pendant plusieurs années comme collègue de bureau mais aussi comme colocataire. Alexandre m'a beaucoup fait courir : qui aurait cru que l'on pourrait me convaincre de faire non pas un demi-marathon, mais plusieurs ? Il a aussi eu la délicatesse de ne pas m'humilier trop souvent aux échecs, et se laissait volontiers distraire lorsque je lui proposais de boire une bière; ce qui ne l'empêchait pas du reste de me mettre mat six coups plus tard. Mais Alexandre est surtout un ami fiable et attentif (pas seulement visuellement), et ses conseils m'ont beaucoup aidé à des moments difficiles de la thèse. Je me réjouis de revoir Alexandre et Kim-Anne après leur aventure californienne, ou au cours de leurs pérégrinations futures.

Je n'oublie pas les collègues et amis des autres groupes à l'IMT. Les membres de l'Esplab ont toujours été accueillants et ouverts envers les membres du petit cousin Parlab. Je les remercie pour la bonne ambiance lors des apéros, des soirées pâtes et des week-ends de ski. Une pensée spéciale pour les *sportifs*: Sara, Davide, Roman, et Amadou, qui m'ont accompagné pour les courses en forêt; Davide, Sara, puis les autres *Courants Solaires* avec lesquels j'ai participé au tour du canton; Nabil, Alexandre, Davide, Christian, Piero, Grégory, qui ont été mes compagnons de dernière place pour le tournoi de football en salle; Vincent et Patrick qui ont compté les longueurs au Nid-du-Crô avec moi.

Je remercie également tous ceux qui facilitent la vie de tous les doctorants de l'IMT. Laurent, qui gère les ressources informatiques pour l'Esplab et le Parlab; Hassan, qui est toujours disponible si j'ai des problèmes avec le plotter ou les PC des étudiants, Claudine, Joëlle et Florence qui peuvent résoudre tous les problèmes administratifs, et finalement Martial, qui n'a pas hésité à faire contrôler les câblages de mon bureau quand j'ai grillé deux cartes-mères dans la même semaine.

Mes amis et ma famille ont souvent subi mes plaintes lorsque tous ne se passaient pas comme je le souhaitais. Merci en particulier à mon père Gérard et à ma soeur Florence, qui ont sans relâche tenté d'améliorer mes capacités de communication, en me demandant des explications concernant l'avancement de cette thèse. Leurs efforts ne furent pas souvent couronnés de succès. Mais la confiance qu'ils m'ont témoignée m'a aidé à aller de l'avant. Merci aussi au groupe des gymnasiens pour les sympathiques soirées jeux, et les soupers de Noël en Août.

Finalement, je tiens à remercier celle qui m'a soutenu et encouragé au quotidien au cours de ces deux dernières années. Clémentine, tu as toujours su me faire sourire, même lorsque je me sentais totalement dépassé. Tu as insisté pour que je prenne du recul, souvent en m'emmenant en haut d'une montagne. Tu m'as aidé à contourner les difficultés que je trouvais insurmontables. Tu me faisais oublier la distance qui trop souvent nous séparait. Tu balayais mes soucis et mes préoccupations avec autant de force que tu balais les volants de badminton. Ni mes hésitations, ni les retards dans l'écriture ne sont parvenus à venir à bout de ta patience. J'espère pouvoir t'apporter moi aussi ce soutien quand tu en auras besoin. Et aussi te faire oublier tes difficultés grâce à de belles randonnées et de longs voyages à deux, tout autour du globe.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Scope of the thesis	2
1.3	Main contributions	2
1.4	Thesis outline	2
2	State of the art	4
2.1	High resolution 3D microscopy	4
2.1.1	Depth from structured light	4
2.1.2	Stereo vision	5
2.1.3	Parallel optical coherence tomography	7
2.1.4	Depth from focus	9
2.1.5	Comparison of 3D microscopy approaches	11
2.2	Time-of-flight cameras	13
2.2.1	TOF measurement principle	13
2.2.2	TOF cameras	14
2.2.3	TOF cameras applications	17
2.2.4	Comparison with mature 3D vision approaches	18
2.2.5	Reduction of TOF camera errors	22
2.2.6	Simultaneous operation of multiple TOF cameras	22
2.3	Range image registration techniques	23
2.3.1	Registration of intensity images for stereo vision	23
2.3.2	Iterative Closest Point registration	24
2.3.3	Registration based on geometric primitives	25
3	Depth from focus vision system for micro assembly	26
3.1	Motivations for 3D microscopy	26
3.1.1	Inspection and assembly applications	27
3.1.2	Embedded vision system requirements	27
3.2	Depth from focus measurement technique	29
3.2.1	Depth from focus system key components	29
3.2.2	Reference macroscopic implementation	30
3.2.3	Limitations	30
3.3	Image sensor for miniature system	32
3.4	Optics of depth from focus imaging	32
3.4.1	Available magnifications for miniature prototype	34
3.4.2	Comparison of macroscopic and miniature optics	36
3.5	3D microscope camera motion	36
3.5.1	Specifications for a micro-camera focus actuator	36
3.6	Depth from focus software processing	38

3.6.1	Sharpness operators	38
3.6.2	Comparison of sharpness operators	39
3.6.3	Multiresolution filtering	40
3.7	Miniature system - Experimental results	42
3.7.1	Sample images	42
3.7.2	Depth resolution	42
3.8	Miniature system - Perspectives	46
3.8.1	Low-mass depth from focus motor	46
3.8.2	High-frame rate imaging	46
3.9	Conclusion	47
4	Time-of-flight camera system	48
4.1	Characteristics of CW TOF cameras	48
4.1.1	Continuous wave TOF signal demodulation	48
4.1.2	Range accuracy limits for TOF cameras	49
4.1.3	From TOF range maps to cartesian coordinates	51
4.2	SR-3000 TOF camera characteristics	52
4.3	Noise in SR-3000 cameras	53
4.3.1	Average noise level	53
4.3.2	Amplitude dependent noise	57
4.3.3	Noise reduction	60
4.4	Deterministic error sources	61
4.4.1	Multipath	63
4.4.2	Scattering	63
4.5	Comparison of TOF error sources	68
4.6	Conclusion	71
5	Scattering compensation	72
5.1	Principle of compensation procedure	72
5.1.1	Simplified scattering model	73
5.1.2	Scattering point spread function	73
5.2	Scattering models	75
5.2.1	Space variant models	76
5.2.2	Space invariant models	77
5.3	Convolution based compensation	79
5.4	Compensation by Fourier division	80
5.4.1	Windowing function for FFT processing	81
5.5	Complexity comparison of scattering compensation techniques	81
5.6	Optimization of scattering model parameters	82
5.6.1	Family of models tested	83
5.6.2	Optimization experiments	83
5.7	Compensation results	83
5.8	Limitations of scattering PSF model	86
5.9	Conclusion	88
6	Registration of noisy range images	89
6.1	3D point clouds registration	89
6.2	Motivation	90
6.3	Registration based on intensity images : bundle adjustment	90
6.4	Matched set of reference points	92

Contents

6.4.1	Least squares transform for two matched point sets	92
6.4.2	5 spheres calibration object for SR camera	94
6.5	Iterative Closest Points (ICP) for registration	94
6.5.1	Algorithm principle	96
6.5.2	Limitations of ICP methods	96
6.6	Geometric primitives for registration	96
6.6.1	Geometric primitives and degrees of freedom	97
6.6.2	Random Sample Consensus (RANSAC) for plane primitive extraction	98
6.7	Registration from cube corner planes	99
6.7.1	Decomposition of rigid-body transformation	99
6.7.2	Rotation estimation	100
6.7.3	Translation estimation	101
6.8	Master plane point cloud alignment	101
6.8.1	Usage considerations	103
6.9	Conclusion	103
7	Registration experiments on noisy range images	105
7.1	Evaluation of registration error	105
7.1.1	Nearest neighbor distance metric	105
7.1.2	Illustration first order metric on depth-from-focus microscope datasets	106
7.2	TOF data registration based on 5 spheres calibration object	108
7.3	Large overlap between point sets	111
7.3.1	ICP registration	111
7.3.2	Registration based on cube corner planes	111
7.3.3	Master plane registration	114
7.3.4	Comparison of registration methods	114
7.4	Small overlap between point sets	116
7.4.1	Registration from bundle adjustment of checkerboard target	116
7.4.2	ICP registration	119
7.4.3	Master plane registration	119
7.4.4	Comparison of registration methods	119
7.4.5	Error estimation on segmented point subsets	122
7.5	Conclusion	126
8	Applications	127
8.1	Network of TOF cameras	127
8.2	Software implementation	127
8.3	Occlusion removal	130
8.4	Field of view extension	134
8.5	Conclusion	134
9	Conclusion	139
9.1	Depth from focus device for micro-assembly	139
9.2	TOF image error compensation	140
9.3	Registration of TOF images	141

A	Depth from focus - Optics	142
A.1	Image formation and depth of field definition	142
A.1.1	Available magnifications for miniature prototype	143
A.2	Relation between aperture and depth of field	144
A.3	Relative depth of field	144
A.4	Depth of field - Experiments	146
A.5	Telecentricity	149
A.5.1	Magnification of blurred regions	149
A.5.2	Telecentric objectives	151
B	Bibliography	152
B.1	Publications list	152
B.2	References	154

List of Tables

2.1	3D microscopy approaches comparison.	12
2.2	Comparison of sequential and parallel TOF devices	20
3.1	Expectations for local 3D vision sensor	29
3.2	Main components in macroscopic 3D depth from focus microscope	30
3.3	Characteristics of macroscopic depth from focus microscope system	31
3.4	Comparison: TM-1001 image sensor and CH-166 micro-camera head	33
3.5	Available magnifications for miniature imager prototype	35
3.6	Comparison: Leica MZ-12 optics and 15mm micro-camera objective	35
3.7	Specifications for micro-camera focus actuator	36
3.8	Comparison: sharpness operators for depth from focus	41
3.9	Main components in miniature depth from focus prototype	43
3.10	Resolution comparison: miniature imager and reference system	44
3.11	Summary: miniature depth from focus prototype	46
4.1	Time-of-flight camera average noise, example	53
4.2	Range standard deviation as function of amplitude	60
4.3	Multipath artifacts, cube corner example	63
4.4	Scattering artifacts: average background displacement	69
5.1	Comparison of deconvolution complexity for different scattering models	82
5.2	Average processing time for different scattering descriptors.	82
5.3	Geometric parameters of gaussian kernels used in optimization experiment.	83
5.4	Scattering compensation results for 3 simple scenes.	86
6.1	Expected success of registration techniques for SR data	103
7.1	Registration: 3D microscope dataset	108
7.2	Illustration of nearest neighbor distance metrics - Microscope dataset	109
7.3	5 spheres calibration object: results	110
7.4	Registration: large overlap dataset. Distance metrics	114
7.5	Large overlap dataset. Noise level	116
7.6	Intrinsic camera parameters for SR-3100 (sn097027)	116
7.7	Intrinsic camera parameters for SR-3000 (sn296012)	117
7.8	Registration: small overlap dataset. Distance metrics	120
7.9	Small overlap dataset. Noise level	120
7.10	Registration: small overlap dataset. Example 1	122
7.11	Registration: small overlap dataset. Example 2	126
7.12	Suitability of point cloud registration techniques to SR data	126

List of Tables

A.1 Available magnifications for miniature imager prototype 144

List of Figures

2.1	Illustration of four 3D microscopy approaches	5
2.2	Depth from structured light setup	6
2.3	Common main objective lens (CMO) stereo microscope	6
2.4	Stereo images of a microgripper	7
2.5	Parallel optical coherence tomography measurement setup	8
2.6	Parallel optical coherence tomography principle	8
2.7	Parallel optical coherence tomography example	9
2.8	Parallel OCT 3D images of a gearwheel	10
2.9	Depth from focus 3D microscopy: measurement principle	10
2.10	Depth from focus measurement of a 600 μ m gearwheel	11
2.11	MiCRoN on-board camera system prototype	12
2.12	Time-of-flight range imaging, principle of operation	13
2.13	Time-of-flight camera setup	15
2.14	Time-correlated single photon counting for range imaging	16
2.15	Shuttered time-of-flight depth measurement principle	16
2.16	Continuous wave time-of-flight cameras	17
2.17	Comparison: Stereo and time-of-flight range imaging	21
2.18	Structured light range imaging system	21
2.19	Noise reduction for CW TOF signal, face example	22
3.1	Real camera versus virtual camera in micromanipulation	28
3.2	Micro-assembly robot with embedded vision sensor	28
3.3	Depth from focus 3D microscope: Leica MZ-12 with Pulnix camera	31
3.4	Kappa CH-166 camera head with 15mm objective.	33
3.5	Depth of field expectation for miniature imaging system	35
3.6	Micro-camera displacement - Schematic	37
3.7	Comparison of sharpness processing time	40
3.8	Overview of multiresolution filtering for DFF imaging	43
3.9	Sample image for miniature system : screw tip	43
3.10	Sample image for miniature system : nails	44
3.11	Miniature system: resolution measurement	45
3.12	Resolution comparison: miniature imager and microscope system	45
4.1	4-taps demodulation of TOF return signal	49
4.2	Noise sources comparison for CCD imager	50
4.3	SR-3000 and SR-3100 time-of-flight cameras	52
4.4	Example SR-3000 data: range and amplitude maps	53
4.5	Example SR-3000 data: 3D rendering, amplitude colormap	54
4.6	3D rendering, depth colormap	55
4.7	Noise in TOF data: raw data vs time average	56

List of Figures

4.8	Noise in SR-3100 data: raw data vs time average	57
4.9	Noise in SR-3000 data: raw data vs time average	57
4.10	Noise in TOF data: comparison of amplitude and range error	58
4.11	Noise in SR-3000 and SR-3100 cameras: Amplitude vs range error	58
4.12	Noise in SR-3100 camera: Amplitude vs range error	59
4.13	Noise in SR-3000 camera: Amplitude vs range error	59
4.14	Saturation in SR-3100 data	62
4.15	Average of two TOF signals	62
4.16	Multipath in standard photography.	64
4.17	Multipath effects in TOF data: warped wall	64
4.18	Multipath effects in TOF data: distorted cube corner	65
4.19	Scattering in standard photography	65
4.20	Scattering example: wall moving behind a person	66
4.21	Scattering example: map of differences	68
4.22	Image segmentation for scattering measurement	69
4.23	Scattering measurement: example 2 (SR-3000)	70
4.24	Scattering measurement: example 3 (SR-3100)	70
5.1	Light scattering in TOF camera	74
5.2	Example of linear coupling between three measurement points	74
5.3	Distribution of scattered light for different target positions	76
5.4	Convolution scattering model	78
5.5	Schematic model of scattering compensation through convolution	79
5.6	Schematic model of scattering compensation in Fourier domain	80
5.7	Scattering PSFs generated by weighted sums of 3 gaussians	84
5.8	Scattering compensation: error minimization	84
5.9	Compensation: background vs foreground RMS displacements	85
5.10	Compensation: background vs foreground RMS displacements - detail	85
5.11	Scattering compensation results: example 1	87
6.1	Point clouds registration for field of view extension.	91
6.2	Point clouds registration for occlusions removal.	92
6.3	Simple calibration object and associated point clouds.	95
6.4	ICP algorithm workflow principle	97
6.5	Geometric shapes and corresponding undetermined DoFs	97
6.6	Workflow for master plane alignment procedure.	102
7.1	Microscope scene for synthetic registration experiment	107
7.2	Synthetic registration experiment	109
7.3	Registration based on calibration object	110
7.4	Scene with large overlap : corner of a room	112
7.5	Room corner scene: successful ICP registration	112
7.6	Plane primitives extraction with RANSAC	113
7.7	Room corner scene: successful cube corner planes registration	115
7.8	Room corner scene: successful master plane registration	115
7.9	Scene with small overlap : office door	117
7.10	Bundle adjustment calibration images	118
7.11	Bundle adjustment calibration toolbox results	120
7.12	Office door scene: invalid registration by bundle adjustment	121
7.13	Office door scene: invalid registration by ICP	122

List of Figures

7.14	Office door scene: successful master plane registration	123
7.15	Scene with person, example 1: bundle adjustment	124
7.16	Scene with person, example 1: master plane	124
7.17	Scene with person, example 2: bundle adjustment	125
7.18	Scene with person, example 2: master plane	125
8.1	Schematic of a multi-camera network	128
8.2	Screenshot of custom software for TOF camera network.	129
8.3	Access control: door scene - Empty scene	131
8.4	Access control: door scene - Door being opened	132
8.5	Access control: door scene - Standing in front of door	133
8.6	Occlusion removal: example scene	135
8.7	Occlusion removal: test	136
8.8	Field of view extension : access control, open door	137
8.9	Field of view extension : access control, moving person	138
A.1	Image formation system.	143
A.2	Depth of field variation with magnification M	145
A.3	Relative depth of field variation with lens diameter D	145
A.4	Relative depth of field variation with lens diameter D - detail	147
A.5	Relative depth of field for different focal lengths f	147
A.6	Depth of field determination experiment	148
A.7	Effect of $200\ \mu m$ camera displacement	149
A.8	Telecentricity issues: magnification varies when camera is moved	149
A.9	Telecentricity issues: 10% magnification change	150
A.10	Telecentricity issues: small scan depth	150
A.11	Telecentric objective	151

1

Introduction

The field of 3D vision is rapidly expanding, as many new imaging devices are introduced each year, in a market driven by increasingly demanding applications. When compared to conventional (2D) imaging, 3D imaging solutions are more difficult to characterize since a wide variety of approaches are used for measuring the depth information, each approach having its specific features, limitations and error sources. But in all cases, image processing algorithms can be used to increase the relevance of the measured information. This thesis presents original contributions for the improvement of two very different 3D imaging devices: 3D microscopy based on the depth from focus principle, and real-time 3D camera systems using time-of-flight (TOF) measurements.

1.1 Motivation

Vision systems are nowadays used in an extremely wide variety of applications, since current cameras are cheap and efficient, enabling low-cost non-contact measurements. In some applications, a 2D camera can not provide all the necessary information. In this case, 3D imaging solutions can be employed.

In the first part of this thesis, an illustration of the 3D vision approach selection process is presented for a real-world application: inspection and assembly of small ($< 10\text{mm}$) parts. In this application, the most important requirement is measurement resolution, while measurement speed comes in second place. Since currently available systems do not offer a convenient solution to this problem, the analysis is extended to propose and study the feasibility of a new system, based on the depth from focus (DFF) principle. The main issue considered is the potential for miniaturization, since a miniature DFF system would represent an adequate solution to the assembly problem.

In the second part of this thesis, the application of real time TOF cameras to surveillance applications is studied. Current systems are limited by camera noise and scattering effects. Since scattering effects are greater than camera noise in many experimental situations, we chose to study possible strategies for reducing scattering artifacts in TOF data. Another limitation to overcome is the limited field of view offered by a single de-

Chapter 1. Introduction

vice. Therefore, a study was performed to determine how to best register different views, in order to use all available cameras concurrently in a vision system.

1.2 Scope of the thesis

This thesis is a contribution to the wide and steadily evolving field of 3D vision. It aims to describe realistic uses of some currently available systems, by taking into account their present limitations. This study will emphasize the role played by state of the art image processing algorithms when 3D imaging devices are employed in challenging applications. The 3D systems under main consideration in this thesis are two very different 3D imaging devices: 3D microscopy system based on the depth from focus principle, and a real-time 3D camera system using time-of-flight (TOF) measurements. For each of these systems, the basic physical principles allowing the recovery of depth information will be reviewed, in order to highlight some critical aspects relating to the device performance and to propose new approaches and algorithms for their improvement in challenging imaging applications. An illustration of such improvements is presented for scattering errors in TOF devices.

Moreover, image processing offers interesting possibilities for aggregation of 3D data for noise reduction or shape inference purposes. Specifically, it allows conversion from low-level range information to higher level primitives (for example geometric primitives). Such primitives are then used as the elements of scene interpretation in complex systems, such as safety systems, or robotic guidance systems. In the present work, we will focus on the conversion of noisy range data into robust plane primitives. Those plane primitives are then used in registration tasks.

Finally, a new application enabled by the combination of several advanced 3D imaging devices and image processing methods is presented: a network of TOF cameras, used for surveillance applications.

1.3 Main contributions

The main contributions of this thesis are:

- a study of system miniaturization for embedded 3D microscopy, aimed at micro-assembly applications,
- a model for the systemic errors caused by light scattering in time-of-flight cameras, expressed as a filtering of the ideal TOF image,
- an efficient method of scattering compensation for TOF images and its optimal implementation for commercial TOF cameras,
- an alignment procedure for range images based on recognition of plane geometric primitives in the range data, used specifically in registration of concurrent TOF views, allowing to virtually extend the field of view of the 3D measurement device, and to avoid occlusions.

1.4 Thesis outline

Chapter 2 gives an overview of the state of the art relevant to the investigations described in this thesis. Chapter 3 describes the microscopic 3D imaging device used in

1.4. Thesis outline

this work, which is based on the *depth from focus* principle, and the feasibility study for its possible miniaturization.

Chapter 4 introduces 3D imagers based on the time-of-flight (TOF) measurement principle, and focuses more specifically on the SR-3000 device, which provided most of the TOF experimental data discussed in this thesis. Reduction of scattering effects observed with SR-3000 cameras is the topic of chapter 5. Chapter 6 presents image registration techniques, with special emphasis on techniques successfully applied to range images produced by TOF cameras. Experimental results are presented in chapter 7. An overview of the new applications possibilities offered by the algorithms developed in this thesis is given in chapter 8. Finally, chapter 9 summarizes the main conclusions drawn in the previous chapters, and includes a discussion of possible extensions to the present work.

2

State of the art

In recent years, devices for 3D imaging have evolved greatly. Many image processing techniques have been studied but new devices create new constraints and require to use new or more efficient algorithms.

This chapter presents an overview of the current state of the art along three main interest direction which were investigated in this thesis: devices for high resolution 3D microscopy, TOF cameras for surveillance applications, and range image registration for real-time operation.

2.1 High resolution 3D microscopy

In this section, four approaches to 3D microscopy are presented and compared. The key in selecting these approaches was the capability to produce range images. 3D microscopy based on sequential point measurements, such as chromatic aberration [70, 96] or confocal microscopy [5] were not considered: those methods are not compatible with real-time operation. The four approaches to 3D microscopy considered are (see fig. 2.1):

- Depth from structured light,
- Depth from stereo,
- Parallel optical coherence tomography (pOCT),
- Depth from focus.

The comparison analyzes the potential to realize *miniature* 3D microscopes, which could possibly be embedded in inspection or assembly systems.

2.1.1 Depth from structured light

In depth from structured light [13, 45], high resolution light patterns are projected on the object. A camera records the resulting images; their intensity distribution is

2.1. High resolution 3D microscopy

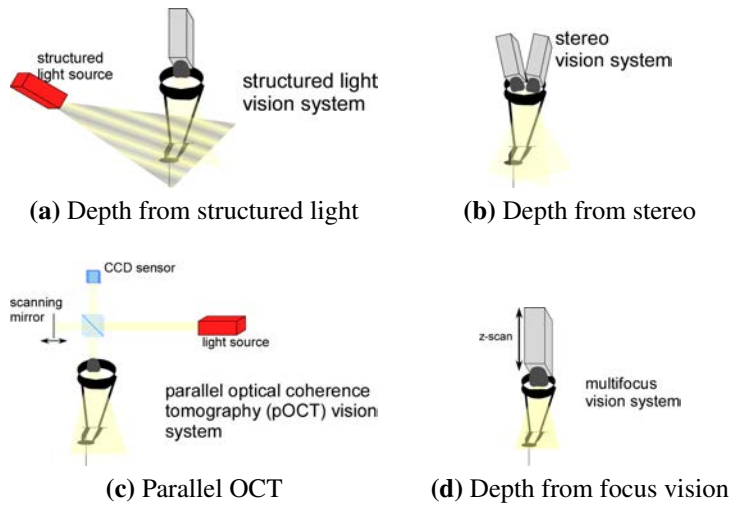


Figure 2.1: Illustration of four 3D microscopy approaches, on which an embedded 3D vision system could be based.

modified by the object shape. The images acquired are treated by software: algorithms analyze the intensity structures recorded for each point, allowing to determine a depth index.

From a detection point of view, structured light methods are very easy to implement: the hardware needed is a standard CCD camera (with appropriate optics). Their difficulty lies in the generation and projection of the appropriate illumination pattern. The simplest illumination scheme is a *laser line*. This technique is used in various commercial 3D scanners, but has the drawback that the object must be scanned laterally, either by moving the object or by the light source. In general, an optimum illumination pattern must be defined for each object to observe. A common solution to save some flexibility is to use a beamer to generate the illumination patterns [13, 47]; see fig 2.2. This also allows to scan the object at different z-resolutions.

In all cases the range image resolution is limited by the resolution of the projected patterns. To reach resolution in the μm range, high magnification optics are required. In that case, short focus depth limits sharpness to a very limited part of the whole scene range.

2.1.2 Stereo vision

The *stereo vision* process is based on the simultaneous acquisition of two images from two cameras. Note that binocular microscopes typically used for high-precision assembly in watchmaking are stereo microscopes. In machine vision applications, the oculars are replaced with CCD sensors (figure 2.3). Stereo vision algorithms are then employed to reconstruct 3D information from the disparity between the two images. A calibration is required to determine the epipolar constraints between the two image sensors [9, 11, 25, 26]. This step uses feature points on a calibration object, matched in the different views. Moreover, calibration allows to take into account distortions produced by the high magnification optics used. An example of a typical micro-assembly scene imaged with stereo-vision is provided in figure 2.4.

Chapter 2. State of the art

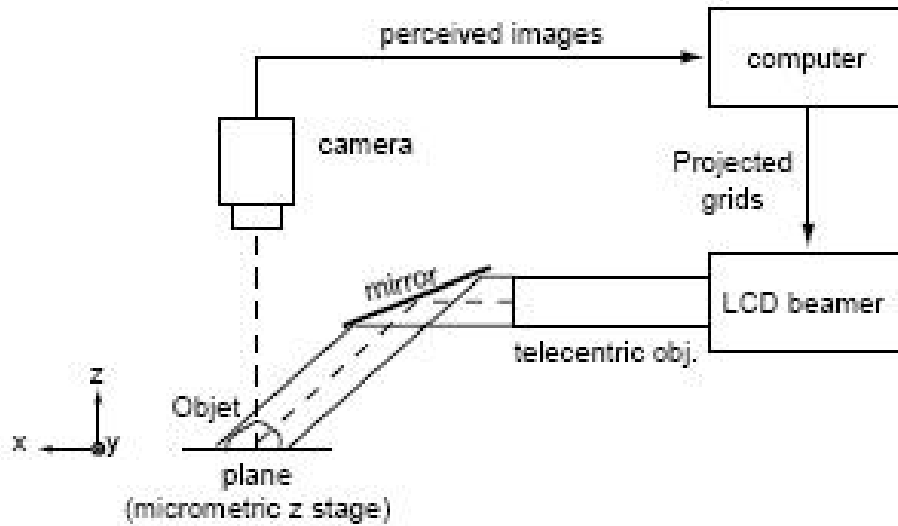


Figure 2.2: Depth from structured light setup, schematic representation (excerpt from [14])

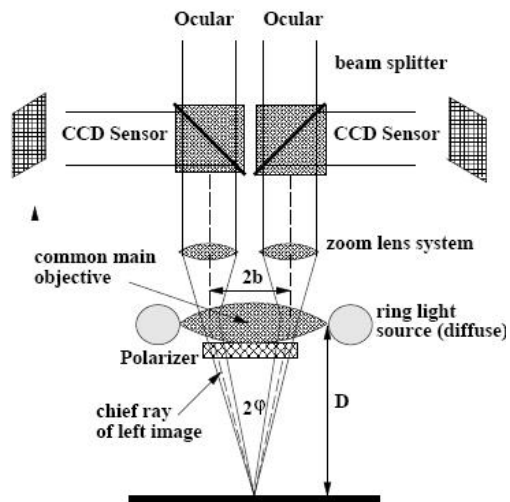


Figure 2.3: Cross section of the common main objective lens (CMO) type stereo microscope (excerpt from [25])

2.1. High resolution 3D microscopy

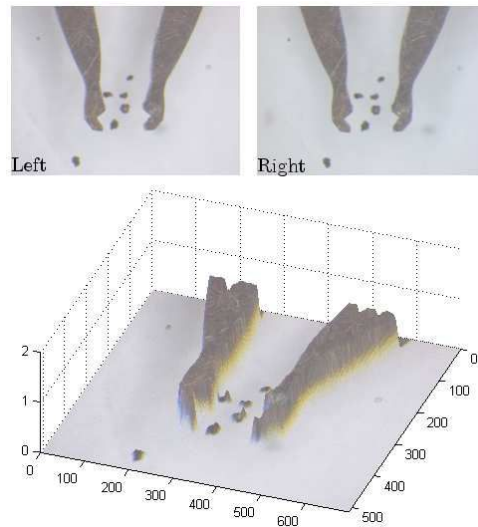


Figure 2.4: Top, the stereo images of a microgripper. Bottom, the surface registration. (Excerpt from Bert *et al.* [11]).

In contrast with depth from structured light, depth from stereo is a *passive* approach: no special illumination scheme is required. Stereo is also compatible with real-time operation, as algorithms for finding epipolar correspondences can be optimized to run in real-time. But one of the main limitations of stereo microscopy lies in the determination of epipolar correspondences: if the object imaged is textureless, or has a periodic texture, stereo matching is impossible. Another limitation is related to the short depth of field observed at high magnifications. The volume where both microscopes are in focus gets smaller as the magnification is increased. If a high lateral resolution is desired, only a slice of the object will be in focus. Finally, we note that the miniaturization potential of stereo system is limited by the constraint to have two independent imaging systems. This doubles the mass of the vision system when compared for example to a depth from focus design.

2.1.3 Parallel optical coherence tomography

Parallel optical coherence tomography (pOCT) is a specialized implementation of white light interferometry. White light interferometry [115] involves two light beams with limited coherence: the beams are not monochromatic, their power is spread among an interval of wavelengths $\Delta\lambda$. The reference beam is reflected on a scan mirror, while the probe beam is projected on the object; see figure 2.5. The optical path difference (OPD) between the two reflected beams produces an interference pattern (fig. 2.6, right) when the beams are recombined. Due to the low coherence of the light used, the interference signal envelope is limited [18, 101]. Displacement of the scan mirror allows to find the maxima of the interference signal. The scan positions of these maxima correspond directly to the range map of the scanned object.

To achieve real-time operation, the scan mirror is kept in periodic motion, at relatively high speed: 10mm/s. The solid state sensor used must thus be capable to operate at high frame rates, in order to keep track of changes in the interference pattern during the scan mirror motion. A system involving a dedicated sensor developed at CSEM

Chapter 2. State of the art

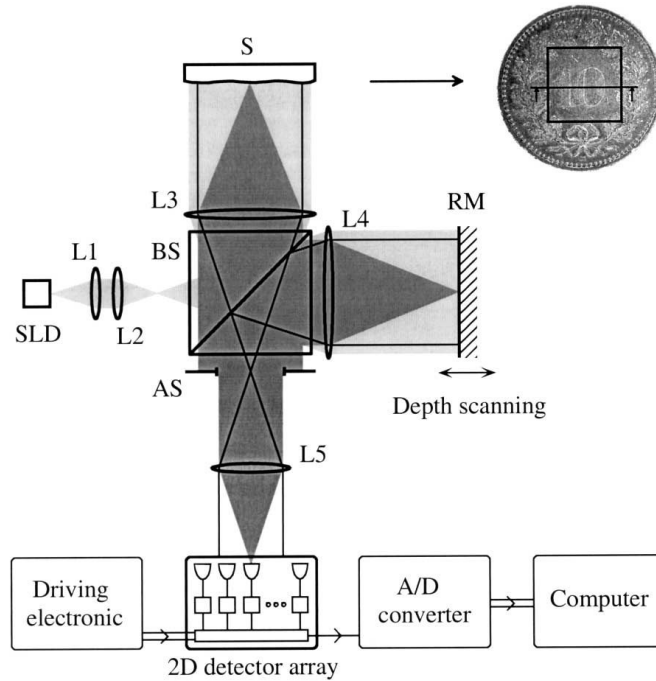


Figure 2.5: Parallel OCT measurement setup [18]. Low coherence light emitted by a superluminescent diode (SLD) is separated by a beam-splitter (BS). The reference beam goes to a mirror (RM), while the probing beam goes to the sample (S). The two reflected beams are combined in the beam splitter, and the resulting intensity is measured by a solid-state detector.

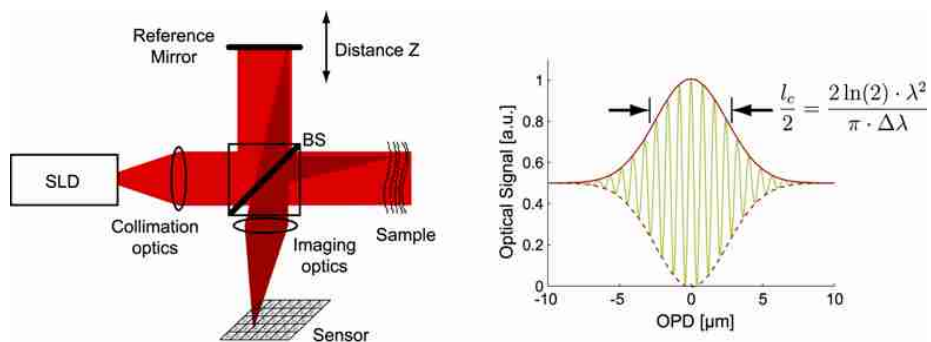


Figure 2.6: Left: parallel OCT schematic. Right: optical signal amplitude as a function of the optical path difference. [48]

2.1. High resolution 3D microscopy

Zürich is commercialized by Heliotis AG [48]; see fig. 2.7. The sensor resolution is

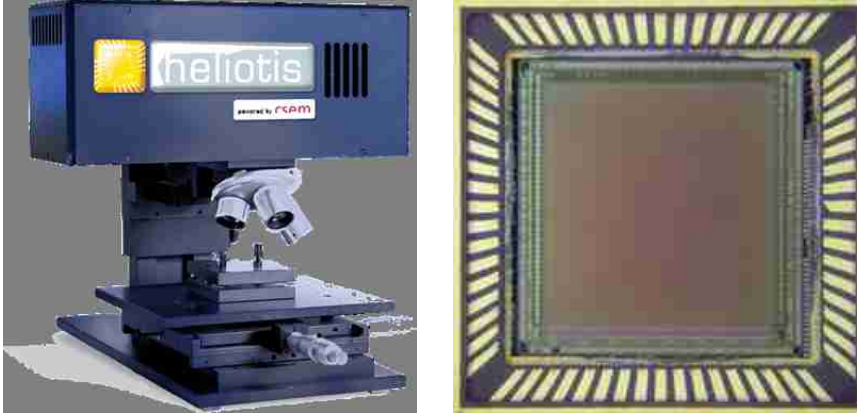


Figure 2.7: Left: Heliotis M1 pOCT microscope. Right: pOCT image sensor. [48]

small: 144×90 pixels, but the frame rate is very high: up to 5000 fps (2D), and 10 fps (3D).

Note that the pOCT technique is referred to as *tomography* due to its application in the field of biology. At the infrared wavelength used (820nm), light can penetrate a few micrometers into biological tissues, and then be reflected to the sensor. This allows to obtain measurements of a *volume* of biological samples. When metallic parts are imaged, the light is reflected in the frontmost surface, and depth maps are obtained. Figure 2.8 provides an example of a gearwheel imaged with the Heliotis M1 microscope.

Technology currently puts a limit on the sensor size used for data readout. Considering miniaturization, difficulties would be encountered in selecting appropriate mechanical and optical components. Moving the mirror element requires small, but high speed and high precision motors. For optical components, the complete optical system with light source, beam-splitter, collimation optics and imaging optics should be miniaturized.

2.1.4 Depth from focus

The principle of depth from focus 3D measurement is described in Ens [28], Geissler & Dierig [40]. A microscope with a short depth of field, is used to acquire a series of images $I_i(x, y)$ at different elevations z_i relative to the object (figure 2.9). After transformation of the images into associated sharpness images $S_i(x, y)$, the object depth for any pixel in the image is the depth associated to the image of maximum sharpness among the stack.

$$Z(x, y) = z_{\hat{i}(x, y)} \quad \text{where} \quad \hat{i}(x, y) = \operatorname{argmax}(S_i(x, y)) \quad (2.1)$$

Note that depth from focus is a passive approach: no active illumination is required. But, in order to perform the sharpness analysis, the sample must have a contrast texture.

The extent of the depth of field puts a higher limit to the achievable depth resolution. Microscope systems with a short depth of field can reach micrometer resolution [83]. Note that the limited depth of field in microscopy is a problem for many applications, for example biological imaging. In this case, depth from focus techniques are

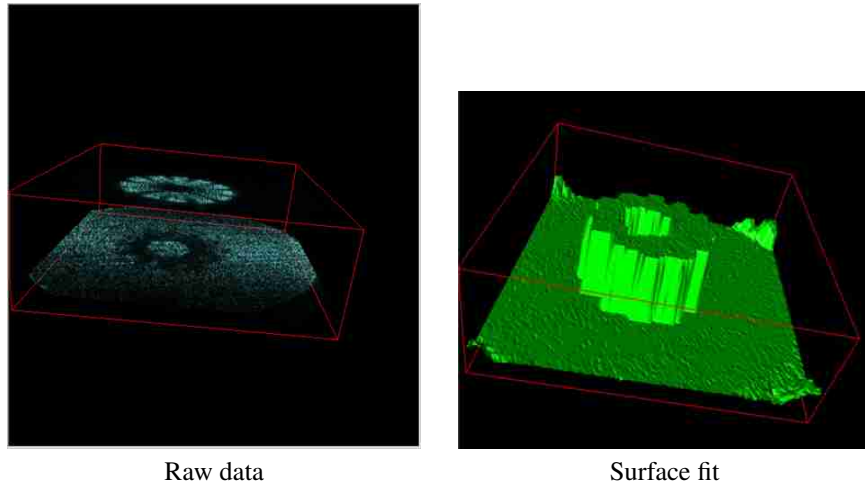


Figure 2.8: Parallel OCT 3D images of a gearwheel [48]

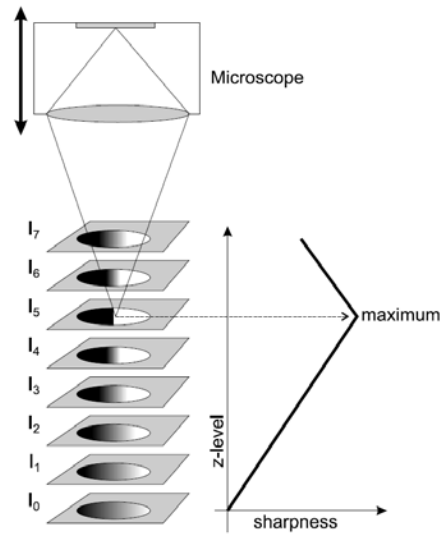


Figure 2.9: Depth from focus 3D microscopy: measurement principle

2.1. High resolution 3D microscopy

developed as a necessary step in the recovery of a well-focused image [2]. Various options for the computation of the sharpness maps have been proposed. Zamofing & Hügli [125] use variance or a Laplacian filter, Boissenin *et al.* [16] use horizontal and vertical gradients, and Aguet *et al.* [2] use a wavelet decomposition of each image in the stack.

Considering miniaturization, the difficulties to overcome are related to the camera motion, and to the requirement to reach high magnifications with small sized optical components. If the magnification is not high enough, the depth of field becomes large and the z-resolution of the system is reduced. Moreover, suppression of out of focus objects requires telecentric lenses when the magnification is small (see sec A.5).

Figure 2.10 provides an example of depth from focus imaging of a gearwheel. Note that this image was produced during the MiCRoN project. This project funded by the European Community under the "Information Society Technologies" program "*aimed at the development of a new Microrobot system based on flexible mobile, 1cm3 sized robots acting autonomously*"([29]). This project required the development of a local vision sensor. In 2006, such a sensor was realized, based on the depth from focus principle; see fig. 2.11. The magnification reached was $5\times$ and the lateral resolution was $1\mu\text{m}$. Unfortunately, no information was published concerning the vertical resolution of this sensor.

2.1.5 Comparison of 3D microscopy approaches

Of the four methods mentioned above, two require a mechanical depth scan (depth from focus, parallel optical coherence tomography), two require active illumination (parallel OCT, structured light), and one requires a correspondence matching algorithm (stereo vision). Key characteristics of the 3D microscopy methods discussed above are presented in table 2.1. For an embedded implementation, depth from structured light and parallel OCT are ruled out by the mass requirements (mass of the active illumination setup). Stereo vision is limited by the resolution/depth of field tradeoff: the high optical magnification required for high resolution imaging reduces the depth of field which, for stereo, limits the depth where correspondences can be found between image pairs. Contrasting with stereo, depth from focus benefits from the limited depth of field when working at high magnification. This passive technique seems therefore most appropriate when trying to produce high resolution 3D images with a miniature system. Chapter 3 presents a feasibility study for a miniature depth from focus system.

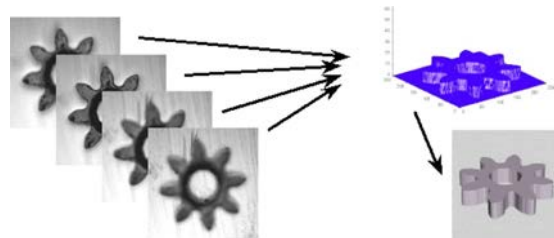
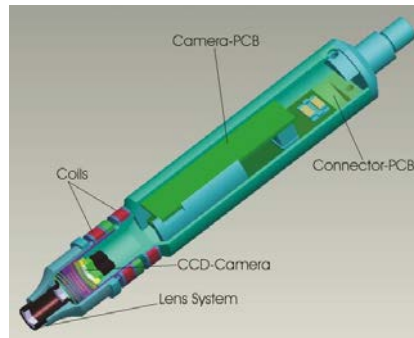
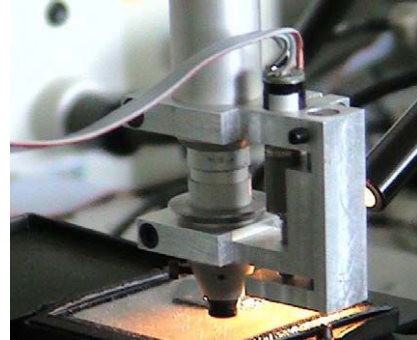


Figure 2.10: Depth from focus measurement of a $600\mu\text{m}$ gearwheel [16].

Chapter 2. State of the art



On-board camera system schematics.



On-board camera system in operation.

Figure 2.11: MiCRoN on-board camera system prototype. [29]

Table 2.1: 3D microscopy approaches comparison.

Method	Advantages	Penalties	Miniaturization issues
Depth from structured light	+ low complexity processing + simple imaging hardware	- active illumination required - complex pattern projection optics	- light source miniaturization - high res. pattern proj. over large scene
Depth from stereo	+ passive system + no moving parts	- correspondence problem - small depth of view	- complex optics
Parallel optical coherence tomography	+ high accuracy + low computational cost when using smart pixel sensors	- mechanical scan required - limited depth range	- complex optics - active illumination required
Depth from focus	+ Passive system + Easy parametrization of z-resolution	- mechanical depth scan required - performance is sample dependent	- increased depth of field for miniature system - camera motion required

2.2 Time-of-flight cameras

Time-of-flight (TOF) cameras are range cameras in which the range measurement is based on the propagation time of an active light signal. In the the last decade, TOF cameras moved from the laboratory [71] to commercial products [20, 82, 113]. Those new devices are expected to fulfill the need for a “*low-priced off-the-shelf system [...] which provides full-range, high resolution distance information in real-time*” (Kolb *et al.* [69]). Considered applications range from cultural heritage to vehicle safety systems, from surveillance to human machine interaction and gaming.

The TOF depth measurement principle is presented in section 2.2.1. Section 2.2.2 reviews implementations of TOF cameras, while an overview of the field of applications of this new technology is given in section 2.2.3; those applications are centered around real-time measurements. A comparison of TOF cameras with existing technologies for real-time range imaging is provided in section 2.2.4. Considerations of TOF cameras measurement accuracy and its possible improvements are reviewed in section 2.2.5. Finally, developments related to usage of TOF cameras in multi-cameras systems are discussed in section 2.2.6.

2.2.1 TOF measurement principle

Time-of-flight cameras involve active illumination, and deliver range (or depth) data by measuring the time needed for a light signal to travel from the camera light source to the scene and back to the camera sensor, as illustrated in figure 2.12. Two methods of TOF measurement are currently used in TOF imaging systems. These methods are often referred to as the *pulsed* and the *continuous wave (CW)* methods [61].

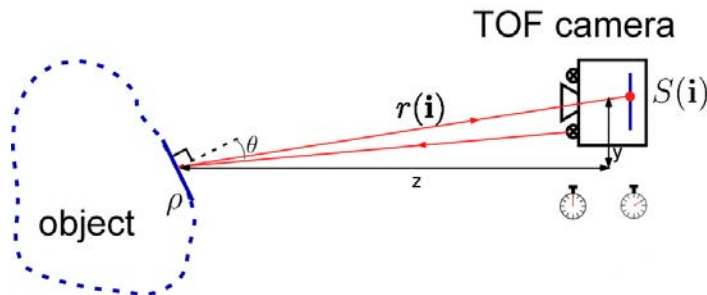


Figure 2.12: Time-of-flight range imaging, principle of operation. A light signal is emitted by the device, reflected on the object, and finally collected on the sensor, producing the electrical signal S . The delay between light collection and emission is proportional to the propagation distance r .

2.2.1.1 Pulsed time-of-flight method

This method was the first historically in commercial TOF acquisition devices, and is still dominant for sequential point scanners. In the TOF device, a pulse of light is generated by a high intensity source (usually a laser). The time interval Δt between pulse emission and return pulse detection allows to compute the distance D between

Chapter 2. State of the art

the targeted spot \mathbf{i} and the device as :

$$D(\mathbf{i}) = \frac{c \cdot \Delta t(\mathbf{i})}{2} \quad (2.2)$$

where $c = 2.99892458 \cdot 10^8 \text{ m/s}^{-1}$ is the speed of light. The measurement range of devices based on pulsed time-of-flight measurement will be constrained by the maximum time interval set to wait for the returning signal.

2.2.1.2 Continuous wave (CW) time-of-flight method

This time-of-flight range measurement method uses the continuous emission of a periodic signal. The distance is derived from the phase shift of the returning signal. The periodic signal $S(\mathbf{i})$ received at the sensor is described by its amplitude $A(\mathbf{i})$ and its phase difference $\Delta\varphi(\mathbf{i})$ with the original signal. The range r is directly proportional to the phase difference. Assuming a sinusoidal signal with frequency f , and with c as the speed of light, we can write :

$$S(\mathbf{i}) = A(\mathbf{i}) \cdot e^{j\Delta\varphi(\mathbf{i})} \quad (2.3)$$

$$r(\mathbf{i}) = \frac{c}{4\pi f} \cdot \Delta\varphi(\mathbf{i}) \quad (2.4)$$

It shall be noted that since the phase difference $\Delta\varphi(\mathbf{i})$ is constrained in the range $[0, 2\pi]$ there is a possible ambiguity in the distance measurement. Typically, the frequency of operation f is chosen such that the ambiguity distance interval R_0 is larger than the required measurement range. R_0 is obtained by inserting the maximum phase difference 2π in eq. 2.4:

$$R_0 = \frac{c}{2f} \quad (2.5)$$

2.2.2 TOF cameras

In TOF cameras, distance measurement is performed in parallel for all the pixels on an imaging sensor. This innovation allows fast acquisition of range images, and represents a major evolution when compared to the established technology of sequential TOF acquisition (see sec. 2.2.4.1). However, parallel imaging requires to spread the illumination on the full field of view: see fig 2.13. This generally results in a low signal to noise ratio (SNR) for the TOF signal, making pulsed TOF measurement difficult. This explains why CW cameras were the first type of TOF cameras developed: the SNR at the operation frequency is much higher, so that the TOF signal can be recovered. However, this type of operation requires ad-hoc sensor, capable of demodulating high frequency signals independently for each pixels.

Recently, prototypes cameras based on pulsed TOF have emerged, but such developments are either still at the research level Niclass *et al.* [91], or largely unpublished due to ongoing proprietary product development [1]. This discussion of the state of the art presents two implementations of pulsed TOF cameras. But the implementations that will be considered throughout this thesis are based on the comparatively more mature CW TOF camera operation principle (see chapter 4); all commercial TOF cameras available today belong to this category.

2.2. Time-of-flight cameras

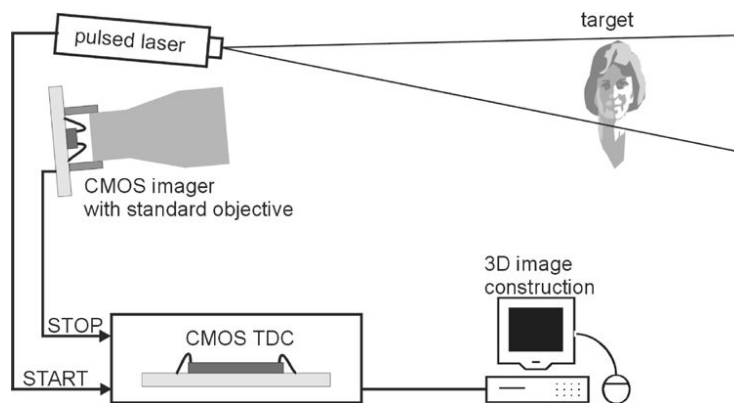


Figure 2.13: TOF camera setup [90]. Note that the illumination must be spread over the whole field of view. Here, the illumination uses a non-collimated pulsed laser.

2.2.2.1 Pulsed TOF cameras

For pulsed TOF cameras, the challenge lies in ensuring that the measurement corresponds to the active illumination signal, and not to noise, either in the measurements system or in the scene (other light sources). Two approaches can be distinguished: Niclass *et al.* [90] use a ultrafast and highly sensitive sensor based on single photon detectors, while Iddan & Yahav [54] use a standard sensor, but with an ultrafast shutter kept open only at times where the return signal dominates. Those two approaches are briefly illustrated below.

Single photon detector TOF Niclass [90, 91] realized a prototype TOF camera based on single photon avalanche diodes (SPAD). Technology is not yet mature for producing large arrays of SPADs; the camera resolution was therefore limited to 32×32 pixels. Using a single photon detector allows a TOF camera to operate in *Geiger mode*. The first detected photon is assumed to be part of the TOF return signal, allowing to compute a range value. One significant advantage of this approach is that the measurement is digital: it isn't affected by noise from in-pixel analog amplifiers, so device that mismatch isn't a significant error source.

However, since *Geiger mode* measurements are largely affected by the sensor's dark count rate, Niclass *et al.* [90] also proposed a more robust method, based on the histogram of arrival time for repeated measurements; see fig. 2.14. The motivation for this processing is that noise is expected to produce dark counts at random times, while the illumination signal will produce strongly correlated TOF measurements. This correlation manifests as a peak in the TOF histogram. This technology still in development could reach very high accuracies. Niclass *et al.* [91] expect the distance accuracy to be under 1cm for ranges up to 7.5m. Moreover, Kahlmann [61] notes that optical crosstalk effects, caused by multipath and scattering (see sec. 4.4), do not affect the depth measurement in this system. The main drawbacks of this technology is that SPAD arrays are difficult to produce, and that data readout requires a large number of high resolution timers (50ps [91]); interfacing and data readout is therefore expensive for such sensors.

Shuttered TOF This variant of the pulsed time-of-flight method is based on a fast shutter allowing to integrate the light pulse [54] only on a short time gate. The principle

Chapter 2. State of the art

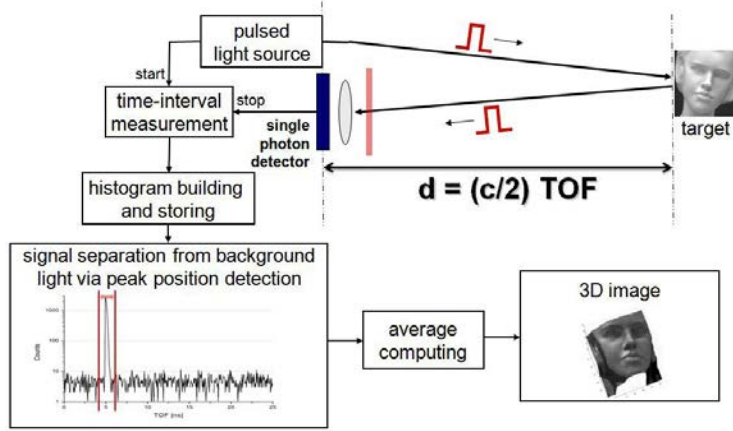


Figure 2.14: 3D imaging setup based on time-correlated single photon counting (TC-SPC) [91]. The peak in the histogram allows a more reliable range determination.

of operation can be summarized as follows:

- A precisely timed light pulse covering the whole field of view is emitted: fig. 2.15 left.
- The reflected signal is integrated for a short period of time only: fig. 2.15, middle.
- The fraction of the light signal returned is interpreted as a range map: fig. 2.15, right

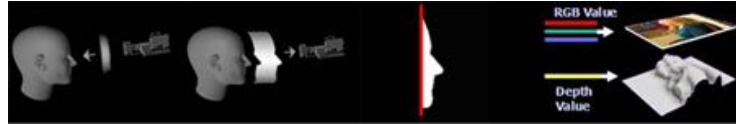


Figure 2.15: Shuttered time-of-flight depth measurement principle [1].

In this system, the distance range of interest $[r_{min}, r_{max}]$ determines the opening of a time gate $\tau = [t_{min}, t_{max}]$ by the fast shutter placed in front of the CCD sensor. The fraction α of light measured during the time gate I_{gate} compared to the total light amount I_{total} varies linearly with the distance to the object [69]:

$$r(i, j) = r_{min} + \alpha(i, j) \cdot (r_{max} - r_{min}) \quad \alpha(i, j) = \frac{I_{gate}(i, j)}{I_{total}(i, j)} \quad (2.6)$$

If the desired measurement range $[r_{min}, r_{max}]$ is too wide to fit in a single exposure, several exposures with different gating parameters can be performed [44]. One advantage of this approach is that standard image sensors (with high resolution) can be used. The hardware complications are related to illumination and shutter components. Developments towards a commercial product were carried out by 3DV systems [1, 44, 124]. However, no implementation is publicly available, so that the characteristics of this sensor aren't well known.

2.2. Time-of-flight cameras

Conclusion Although pulsed TOF cameras have been demonstrated, there is currently no *off-the-shelf* camera available for testing. This technology is nevertheless promising, as Niclass *et al.* [91] showed that high resolution measurements is possible, while 3DV systems [1] suggests that low-cost TOF cameras could be produced in a very near future. Note that even if such a development occurs, higher priced CW TOF cameras would still be useful in applications where pulsed TOF isn't adequate, such as simultaneous acquisition with multiple TOF cameras.

2.2.2.2 CW TOF cameras

Similarly to Niclass *et al.* [91] pulsed camera, CW TOF cameras require special pixel sensor designs, in order to allow for demodulation of the periodic light signal. Lange [71] reports on developments leading to the realization of such sensors. Oggier *et al.* [93] presents the first family of commercial TOF cameras. Büttgen [19] proposed various improvements to CW TOF cameras, including active background light suppression at the pixel level, and pseudo-noise modulation for simultaneous multi-camera operation. Kahlmann [61] investigated methods for CW TOF cameras calibration, and proposed the addition of an optical reference path to reduce the effects of temperature variations on TOF sensors [59].

Mesa-imaging [82], PMDtec [113] and Canesta [20] offer commercial CW TOF products. Two cameras are illustrated in 2.16: a Swissranger SR-4000 and a CamCube. For indoor cameras, a non-ambiguity range of 7.5m is often chosen, resulting in a modulation frequency f of 20MHz (see eq. 2.5). The parallel demodulation of the light signal allows to obtain range maps in a very short time (typically, at 20 frames/s). In these cameras, the lateral resolution is limited by the sensor size, which is small in current commercial products (ranging from 64×48 to 204×204 [69]), since all pixels must incorporate signal demodulation electronics. Future generations of TOF cameras are expected to reach resolution similar to video cameras (VGA).

More details on CW TOF camera operation are presented in section 4.1.

2.2.3 TOF cameras applications

Usage of range sensors touches a wide variety of applications, going from cultural heritage[122] to vision feedback for robot control [68], from safety systems[68] to human computer interfaces. Kolb *et al.* [69] recently produced an overview of published literature concerning TOF cameras. TOF cameras have been proposed for all tasks mentioned above. Research on TOF cameras has been stimulated by different research projects. The "Action Recognition and Tracking based on Time-of-Flight Sensors"



Figure 2.16: Left: PMDTec/ifm electronics CamCube sensor; Middle: MESA SR4000 sensor; Right: The ToF phase-measurement principle. - Excerpt from Kolb *et al.* [69].

Chapter 2. State of the art

(ARTTS) project, funded by the European Union (www.artts.eu) investigates “*algorithms for tracking and action recognition with a focus on multi-modal interfaces and interactive systems*”([69]), that is, human machine interaction based on TOF cameras. The key characteristic of TOF cameras appealing for human machine interaction is the production of range maps in real-time, at with a low processing cost. Du *et al.* [27] propose a virtual keyboard where the user’s fingers are tracked by a TOF camera. At a larger scale (1.5m × 1m), Oggier *et al.* [94] presents a touchscreen application where a TOF camera track the user’s hand to control a computer display.

In the surveillance context, Kahlmann *et al.* [60] proposed to use a particle filter on TOF point clouds for tracking people or other moving objects. This approach has also been followed by Hansen *et al.* [46]. Jensen *et al.* [56] use a TOF camera for gait analysis when tracking walking persons. In the field of video content generation, TOF range information allows an easier registration of high resolution video sequences. Guan & Pollefeys [42] use TOF cameras to enhance the robustness of shape from shading 3D reconstruction. For augmented reality applications, where computer generated models are added to video sequences, range information from TOF sensors enables a correct ordering of the different depth layers [44, 44, 54]. The EU-project 3D4YOU (www.3d4you.eu) encourages further development in this direction, with the final aim of enabling real-time 3D acquisition for consumer 3D television systems.

TOF cameras can also be used for non-contact sensing in robotics. Kohoutek [68] presents an application of TOF cameras for collision avoidance with approaching humans. Fuchs & May [39] use TOF camera vision feedback for controlling a robot grasping different objects. TOF cameras can also be used for mobile robot guidance [120]. Finally, specialized long range versions of TOF cameras have been proposed for collision avoidance in automotive applications [61, 94]. Note again that the main advantage expected of TOF cameras over other vision approaches is the high frequency at which range images are delivered, enabling real time applications.

2.2.4 Comparison with mature 3D vision approaches

In order to evaluate the improvements brought by TOF cameras over previous 3D measurement methods, a brief comparison of TOF cameras with mature 3D measurement technologies is provided.

2.2.4.1 TOF point scanners

In these scanners, the TOF measurements are sequential : each point of the acquired 3D point cloud is measured separately. Commercial devices [31, 72] use a laser source for signal emission, with rotating mirrors to provide lateral scanning. TOF point scanners were historically the first TOF imaging devices, and their operation was until recently based on the pulsed time-of-flight measurement method.

Since the full power of the light signal is available for each measurement point, such scanners have a large measurement range (up to 300m) and a good accuracy. The main drawbacks of these systems are slow acquisition time and high cost. The sequential acquisition scheme prohibits real-time data acquisition with such systems. Nevertheless, TOF point scanners are widely used in architecture, geodesy or cultural heritage applications.

2.2. Time-of-flight cameras

2.2.4.2 Comparison of TOF point scanners and TOF cameras

TOF point scanners - based on sequential acquisition - and TOF cameras - based on parallel imaging - are compared in table 2.2. The main drawback of TOF cameras when compared to TOF points scanners is the low accuracy and limited range. This is mainly explained by the fact that the energy of the light signal emitted by the TOF device must be spread to cover the whole field of view of all pixels. The main drawbacks of TOF point scanners are:

- complexity and high cost: the scanning mechanism required to move the light beam across the scene involves moving parts (mirrors). Therefore, the price of TOF point scanners is high.
- low acquisition speed: although each new version of commercial devices increases the faster data acquisition rate, the sequential operation of point scanners makes them significantly slower than TOF cameras, where the acquisition is parallel.

This comparison confirms that sequential point scanners are not a valid option for real-time 3D acquisition. Therefore, they will not be discussed further in this dissertation. Instead, our attention will be focused on TOF cameras using parallel imaging.

2.2.4.3 Stereo vision

Stereo vision is historically the first approach to 3D computer vision. Many authors proposed and developed algorithms for stereo data registration and processing. The setup of a stereo system is not straightforward, since calibration must be performed. But this issue is so standard that generic toolbox help with this step [17, 123]. Off-the-shelf systems are now available. For example BumbleBee [95] for computer vision, or the inexpensive Minoru webcam [97], aimed at the consumer market.

While stereo systems can provide range information in real-time, over a wide field of view, they suffer from limitations related to the computation of the epipolar correspondences. Those correspondences cannot be determined in uniform image areas. For those areas, the range data is undetermined. This phenomenon is illustrated in figure 2.17. The gray areas in the stereo disparity images correspond to regions where the range measurement failed. In comparison, a TOF camera can easily image such zones.

2.2.4.4 Structured light

The general availability of cameras and light projectors make this method very affordable. Laser line scanners are widely used for range sensing in the industry. However, laser line scanners require a mechanical motion to sweep the line across the camera field of view. They are therefore not well adapted to the quick acquisition of an arbitrary scene. More flexible structured light systems are based on computer controlled digital light projectors. A schematic diagram of a real-time structured light acquisition system is reproduced in figure 2.18. Open source tools [30] can be used for the calibration of a custom setup. Constrained by the type of camera and the power of light source used, the acquisition can nevertheless be fast. Zhang & Huang [128] report 2D data acquisition at 120Hz for grayscale images and 26.8Hz for color images (532×500 pixels). 3D shape reconstruction can be performed in 24.2ms, so that this system can produce 3D data in real-time, and capture scenes with motion. The range accuracy is

Chapter 2. State of the art

Table 2.2: Comparison of sequential and parallel TOF devices

System	Sequential TOF point scanner	Parallel TOF camera
Data acquisition rate	Slow typically measured in points/s. State of the art scanners provide 120000 points/s	High typically measured in frame/s (fps). State of the art cameras provide 30+ fps
Measurement range	High from 20m in high-res mode up to 300m	Small typically < 8m
Range data accuracy	High typically mm for small distance (< 20m), cm for larger distances	Low typically cm
Lateral resolution	High some vendors provide up to 470000 × 40000 points for a 360° scan	Low limited by sensor size, currently 176 × 144 for SR-3000 cameras, should be improved in the near future
Mode of operation	Pulsed TOF: Leica[72] CW TOF: Faro[31]	CW TOF: SR[24], PMD[113], Canesta[20] Pulsed TOF: Niclass[90], 3DV[1]
Applications	Architecture, Geodesy, Cultural heritage, Forensics, ...	Human-Computer interface, Surveillance, Robot control, Industrial inspection, ...
Cost	High > 50000 CHF	High to low > 5000CHF for current small series, could reach < 100USD if mass-produced

2.2. Time-of-flight cameras

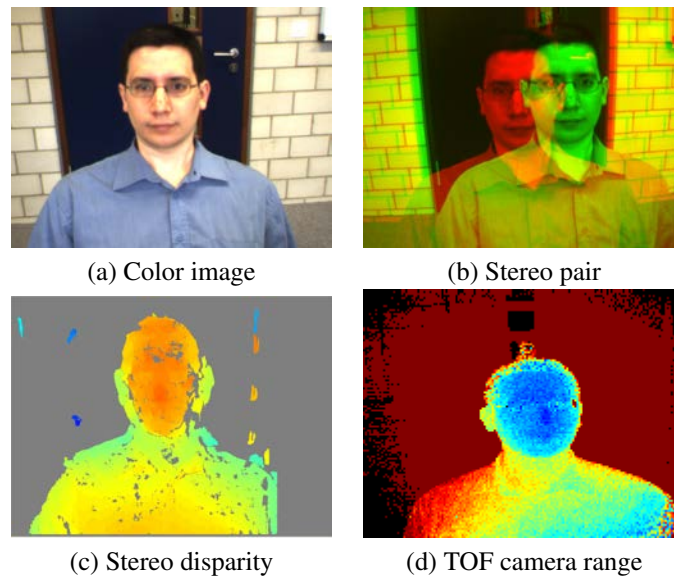


Figure 2.17: Comparison of stereo and TOF range imaging. Stereo processing cannot determine the range information for regions with no texture, or periodic textures. TOF cameras, employing active illumination, aren't affected by textures in the scene.

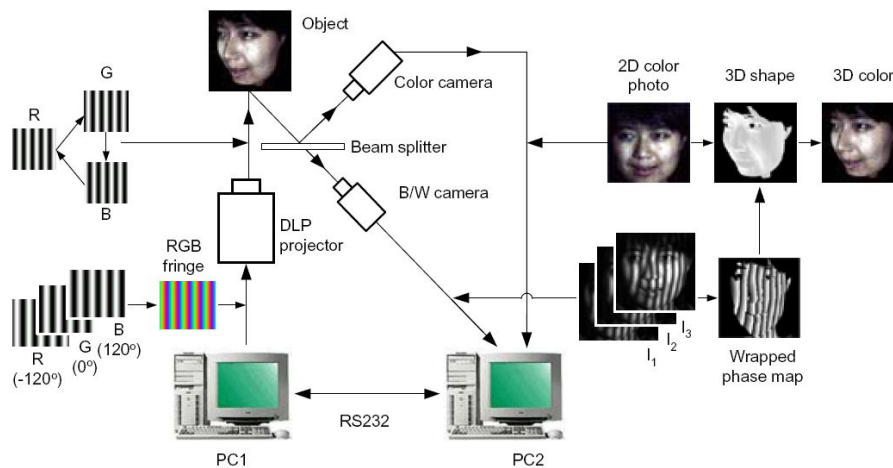


Figure 2.18: Zhang & Huang [128]: Schematic diagram of a real-time 3D shape acquisition system. A color fringe pattern is generated by a PC and is projected onto the object by a DLP video projector (Kodak DP900). A high-speed B/W CCD camera (Dalsa CA-D6-0512W) synchronized with the projector is used to capture the images of each color channel. Then image processing algorithms are used to reconstruct the 3D shape of the object. A color CCD camera (UniqVision UC-930) also synchronized with the projector and aligned with the B/W camera is used to capture color images of the object for texture mapping.

Chapter 2. State of the art

generally in the millimeter range, and seems to be limited by lens distortion. Huang & Han [51] reports RMS errors of 0.35 mm for a plane object measured with structured light.

The main drawback of structured light systems, when compared to TOF cameras, lies in the requirement to conserve a sharp light structure from the light source to the scene, and then from the scene to the camera. The light structure contrast must be significantly stronger than the background illumination, and the focus of the illumination must be adjusted so that blur stays small within the whole depth of field. To reach high accuracy and high acquisition frequencies, the field of view is typically limited to a 200mm \times 200mm area [51]. In comparison, TOF cameras offer a much larger field of view.

2.2.5 Reduction of TOF camera errors

Büttgen [19], Kahlmann [61] already studied noise reduction, but considered mostly stochastic noise source, such as photon shot noise and thermal noise, and proposed solutions involving hardware improvements. Kahlmann [59] introduced an optical feedback loop allowing to compensate for sensor thermal drift. Mure-Dubois & Hügli [85] proposed a scattering compensation algorithm based on a sum of convolution with scattering functions described as sum of gaussians. Kavli *et al.* [64] experimented with scattering compensation involving a space variant scattering model. Recently, Bohme *et al.* [15] proposed to use shading constraints from a Lambertian reflection model to reduce the influence of noise in TOF measurements; see figure 2.19 for an illustration.

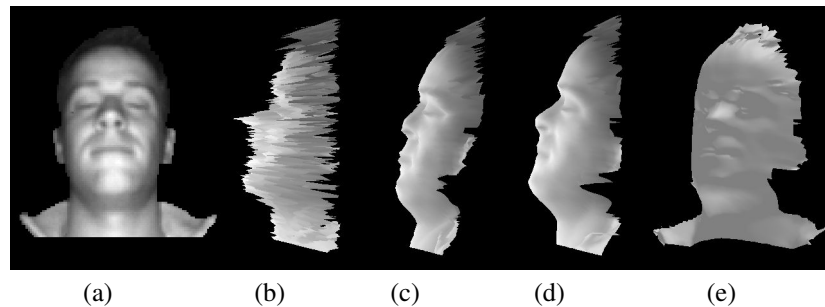


Figure 2.19: Bohme *et al.* [15]: 3D reconstruction of a human face. The manually segmented intensity image is given in (a). A lateral view of the surface as measured by the SR3000 TOF camera is shown in (b). Figure (c) presents the reconstructed surface using the global albedo algorithm. Lateral and frontal views of the reconstructed surface based on the local albedo algorithm are shown in (d) and (e), respectively. The surface in (b), (c), and (d) was textured using the intensity image to facilitate a qualitative comparison.

2.2.6 Simultaneous operation of multiple TOF cameras

TOF cameras use active illumination. This can lead to problems when multiple cameras are used simultaneously to image the same scene. Strategies to avoid interference between TOF devices are discussed in detail in [19]. For pulsed TOF cameras the only option is to have a common trigger signal sent to all devices to avoid simultaneous pulse emissions. For CW TOF cameras, Büttgen [19] presents an original approach for

2.3. Range image registration techniques

code division multiple access (CDMA), based on pseudo-random sequences. Careful selection of the codes used allow for simultaneous usage of a high (100+) number of cameras, the only adverse consequence being an increase in the background illumination level BG . Unfortunately, and although CDMA was demonstrated for SR-3000 cameras, no implementation of this technique is available to camera end-users at the time of writing.

Currently, simultaneous acquisition with different CW TOF camera implies using different modulation frequencies for each camera used. Lange [71] pointed out that the attenuation between signals at different frequencies when using 4 tap sampling depends on the frequency difference Δf and on the integration time T_{int} . When operation frequencies are separated by 1 MHz, the crosstalk between two devices is less than -40 dB for integration times higher than $T_{int,min} = 100 \mu s$.

2.3 Range image registration techniques

Range image registration is used in the production of 3D models. Typical applications include safety systems [68], cultural heritage [122], city modeling [112] and localisation in a context of mobile robots [120]. Many of the techniques used for range images registration are derived from methods used in the registration of conventional 2D images. Therefore, 2.3.1 briefly discusses the techniques used in 2D images registration, with special focus on the techniques used to calibrate a stereo pair. Section 2.3.2 then presents the iterative closest point registration. Section 2.3.3 discusses registration based on geometric primitives, which can be interpreted as an extension of 2D registration methods based on features.

2.3.1 Registration of intensity images for stereo vision

Registration methods are often divided in two categories: correlation based registration, and feature based registration. Correlation based registration is most appropriate when the majority of the image contents is similar in both images. This is the case for example when the overlap between the two views is large. In that case, the correlation between the intensity distribution of the target image and the intensity distribution of the source image has a well defined maximum when the two images are registered. Unfortunately, when a stereo pair is calibrated, there is no guarantee that the overlap between views will be large. In that case, feature based method provide better results. The workflow of feature based registration is the following:

- Definition of feature points in each view. Feature points are generally defined by looking at significant intensity variations. The Harris corner detector is often used.
- Matching of feature points across the views. This step can be based simply on the geometric distance between feature points, but more advanced matching techniques exist, based on feature properties.
- Validation of matches: aberrant feature point pairs are suppressed. aberrant feature point pairs are suppressed.
- Evaluation and iteration: an error metric on the feature points distance is measured. If the error is high, the process is repeated. The registration is stopped when the error falls below a pre-determined threshold.

Chapter 2. State of the art

This technique is often referred to as bundle adjustment, as the registration parameters are iteratively adjusted until all feature points are matched.

For registration of images from stereo cameras Tsai [118] proposed to use a planar calibration target in the field of view to reliably define feature points. Torr & Murray [116] analyzed algorithm robustness, with special focus on solutions to avoid aberrant pairs; least median squares (LMedS) and random sample consensus (RANSAC) are among the most efficient strategies for increased robustness. Zhang [130] used the planar target proposed by Tsai [118], adding procedures to estimate and compensate for lens distortions. This approach is widely used in computer vision research for calibration of a stereo camera registration setup, since implementations of the required algorithms are widely diffused [17, 123]. Note that in some situations, the planar target approach isn't practical, so that different target patterns are used [11]. Lindner *et al.* [76] demonstrated in 2007 that this technique could be used to register a TOF camera with a standard color camera. This approach has been largely followed [66, 100]. Unfortunately, registration of two TOF cameras seems inaccurate, due to the low image resolution of current TOF sensors [61, 88]. Note that in 2008 Kim *et al.* [66] have presented a system where 3 high-resolution CCD cameras and three TOF camera were calibrated using feature points extracted from intensity images.

2.3.2 Iterative Closest Point registration

Since its introduction in the early 1990s (Besl & Mckay [12], Chen & Medioni [21]), the Iterative Closest Points (ICP) algorithm has been widely used to register data produced by 3D scanners [102]. The procedure provides a solution to the registration problem of two point clouds under the assumption of a rigid body transformation, and is based on minimization of an error function defined from distance from point in one cloud to their closest counterpart in the other cloud.

Many variants have been proposed to increase the robustness and speed of ICP algorithms. Rusinkiewicz & Levoy [102] present a good review of ICP algorithms, in which they identify 6 main directions for ICP method optimizations :

1. **Selection** of sets of points from \mathcal{P}_0 and \mathcal{P}_1 .
2. **Matching (pairing)** points between the sets.
3. **Weighting** the point pairs.
4. **Rejection** of outlier point pairs.
5. **Error metric** assignment based on the point pairs.
6. **Minimization** of error metric.

While the original ICP algorithm uses all points for estimation of the transformation, several authors have proposed to select only a subset of points, either chosen randomly [119], or chosen according to their intensity or color (when available) [58]. Bae *et al.* [6] proposed to consider features such as surface normal vector, change of curvature, and variance angle defined at each point to improve the pairing quality. The *k-D tree* technique [36] is used in many ICP variants to speed-up the matching stage [12, 57, 102]. This technique allows to speed up the nearest neighbors search for a target point cloud \mathcal{P}_0 . Other improvements to ICP use the assumption that the points are arranged into a mesh, where the points are the corners of triangular surface patches. When such

2.3. Range image registration techniques

surfaces patches are defined, Chen & Medioni [21] propose to replace the point to point distance by point to plane distances.

To avoid erroneous registration due to aberrant point pairs, the contribution of each pair to the final error metric can be weighted. The weights are generally chosen inversely proportional to the point pair distance [106]. Trucco *et al.* [117] use least median squares (LMedS) to filter outlier point pairs, while Fitzgibbon [35] use a variation of the Levenberg-Marquardt algorithm. Nevertheless, since ICP methods are based on pairing of data in the two sets to register, they tend to fail when the overlap between the datasets is low. Chetverikov *et al.* [22] report that advanced ICP algorithm still require at around 50% overlap for successful registration.

2.3.3 Registration based on geometric primitives

Registration from geometric primitives is an extension of feature based registration to range images. The geometric primitive extraction procedures can simply be interpreted as feature detectors. A simplest primitive extraction procedure is RANSAC: Tarsha-Kurdi *et al.* [112] uses RANSAC to identify plane regions corresponding to roofs in aerial TOF laser scanner data. Von Hansen [121] proposes a framework for the grouping of such plane patches. Rabbani & van den Heuvel [99] and [104] extract plane, but also sphere and cylinder primitives from TOF laser scanner data. This procedure is applied to 3D city digitizing for cultural heritage [105]. Mure-Dubois & Hügli [88] use a combination of RANSAC extracted planes and conventional 2D point features to register range images produced by two real-time TOF cameras. Guan & Pollefeys [42] use a sphere calibration target to calibrate a system containing four cameras: two conventional CCD cameras and two real-time TOF cameras.

3

Depth from focus vision system for micro assembly

This chapter is devoted to the design of a new 3D sensor for operation in an embedded vision system for micro-assembly. System requirements for an embedded sensor are presented. A prototype implementation of a miniature microscope device allowing depth from focus measurements is proposed. Future steps in development of a real-time embedded depth from focus sensor are presented, with a discussion of the most critical tradeoffs.

Section 3.1.2 introduces the main characteristics to consider in the design of an embedded, 3D vision sensor. Section 3.2 discusses the key components in a depth from focus system. For each component, the relevance of miniaturization constraints is analyzed (sections 3.3 to 3.6). In section 3.7, 3D measurements obtained with a prototype miniature imaging system are presented and compared to the performance of a reference, high-resolution imaging system. Finally, section 3.8 includes a summary of the performance attained with a miniature depth from focus system. This summary is supplemented with a list of future developments required to realize a high performance 3D local vision sensor.

3.1 Motivations for 3D microscopy

Machine vision plays an important role in automated assembly. However, present vision systems are not adequate for robot control in an assembly environment where individual components have sizes in the range of 1 to 100 micrometers, since current systems do not provide sufficient resolution in the whole workspace when they are fixed and they are too bulky to be brought close enough to the components.

This chapter provides a feasibility study of a 3D vision sensor easily embedded in a micro-assembly robot. A small-size 3D vision system is expected to provide two decisive advantages: high accuracy and high flexibility. In planar assembly tasks, an embedded camera can provide vision feedback [49]. More general assembly tasks will

3.1. Motivations for 3D microscopy

require 3D sensing. Bert *et al.* [10] notes that 3D vision can be used for synthesis of a virtual camera view, allowing to provide visual feedback from a position where a camera could not be placed, for example because this position is occupied by the robot; see figure 3.1.

3.1.1 Inspection and assembly applications

In micro-assembly, a robot is used to manipulate the micro-sized parts. Accuracy requirements are such that assembly can not be realized through open-loop robot command. In order to close the loop, either for teleoperation [34, 110], semi-autonomous [8, 23] or autonomous [50, 62, 111] robot operation, vision sensors are used to provide feedback on the relative positions of the effector and parts to assemble. Most vision systems employed in automated micro-assembly are based on bulky microscopes [23, 32, 34, 62, 65, 110] with a fixed field of view. Typically, such systems feature fixed robot effectors, centered within the microscope field of view. Parts are brought for assembly by a motorized stage moving under the microscope. This mode of operation is slow since a large mass must be set in motion for each new part introduced. In contrast with this approach, parallel robot systems [4] use low-mass effectors capable of being moved quickly over a large assembly workspace, allowing for much faster assembly. However, in this situation, a vision system covering all possible positions of the effectors, i.e. the whole assembly workspace, is not accurate enough (see figure 3.2). Therefore, for closed-loop operation, a high-resolution, embedded vision system must be moved with the effectors.

3.1.2 Embedded vision system requirements

In this section, we present a list of the desired properties for an embedded 3D vision sensor. The tradeoffs associated with each of these properties are also briefly discussed.

3.1.2.1 Mass

While sensor mass is generally not considered important in computer vision applications, it is, however, of key relevance for embedded system design. As a principle for a local sensor, the volume imaged is only a small fraction of the assembly workspace. This implies that the embedded sensor will be moved with the robot active systems (grippers, actuators) during assembly operations. Fast motion is possible only if sensor mass is low. As a guideline for a practical application, we set the constraint : $m \leq 100 \text{ g}$. We will see that this constraint is of critical importance, since it prohibits the use of high performance optics, such as bulky microscope objectives. This in turn limits the lateral and vertical resolution that can be attained. It also constrains the choice of the imager device (the pixel size must be small, in order to provide high resolution images with a small optical magnification).

3.1.2.2 Resolution - Field of view

In order to reach assembly tolerances, the spatial resolution (r_x, r_y, r_z) must be as high as possible. Furthermore, it is desirable to have a volume of view (L_x, L_y, L_z) as large as possible, in order to include all the relevant parts present in the workspace into the local 3D image. As a target for system design, we specify a range for the volume of view varying between 1 mm^3 and 1 cm^3 . Since the number of pixels in a standard camera

Chapter 3. Depth from focus vision system for micro assembly

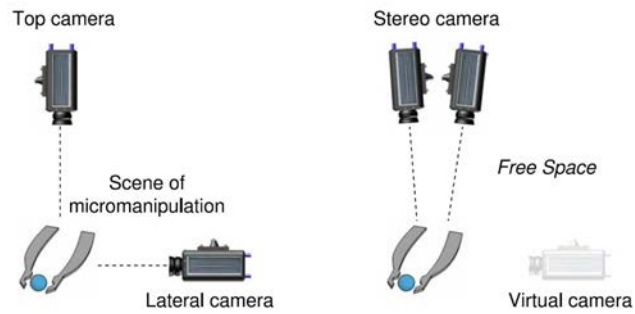


Figure 3.1: Real camera versus virtual camera in micromanipulation [10]. Using 3D vision, in this case by stereo, allows to provide visual feedback from a position where a camera could not be placed.

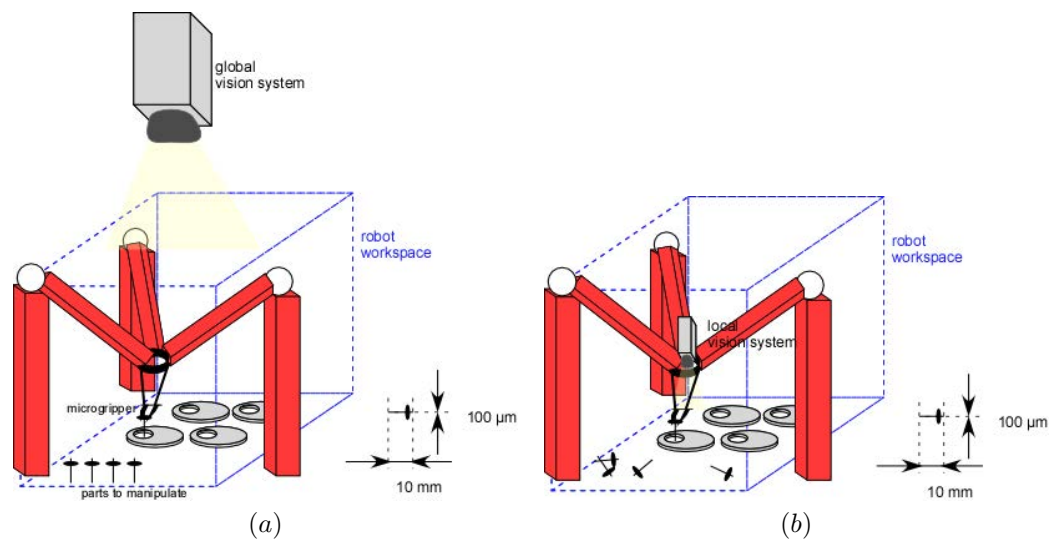


Figure 3.2: Micro-assembly robot with global (a) or embedded (b) vision sensor systems.

3.2. Depth from focus measurement technique

rarely exceeds 1000×1000 , planar resolution will be, at best, limited to one thousandth of the volume of view $(0.1 \mu m^{-1}, 0.1 \mu m^{-1}) \leq (r_x, r_y) \leq (1 \mu m^{-1}, 1 \mu m^{-1})$. A compromise must be found between resolution and volume of view, depending on the target assembly application.

3.1.2.3 Frame rate

When used in a production environment, the 3D sensor must provide data in real-time. As a target value for a practical application, we specify that the data should be produced at $R = 10 \text{ fps}$ (allows real-time teleoperation, and more efficient autonomous operations). Depending on the final application, a lower frame-rate could be accepted, especially in applications where high-precision is more critical than high-speed operation.

3.1.2.4 Summary

The requirements exposed above are very different in their nature. It may show difficult to reach all target values simultaneously. Therefore, we need to set priority rules between the different requirements. Table 3.1 summarizes the expectations for a local 3D sensor, and exposes the main penalty if the expected values cannot be reached.

When aiming for an embedded application, the highest priority shall be set to the compliance with the mass requirement. Next in order of priority comes the spatial resolution. Volume imaged and frame rate share the third level of priority.

3.2 Depth from focus measurement technique

Section 2.1.5 provided a comparison of 3D microscopy methods. The comparison aimed to identify the best candidates for miniaturization. The *depth from focus* principle was selected: for high magnifications, the short depth of field is an advantage, when it is an issue limiting the work volume for stereo vision and structured light.

3.2.1 Depth from focus system key components

Based on the measurement principle presented in section 2.1.4, we can distinguish 4 main components in a depth from focus 3D imaging system:

Table 3.1: Expectations for local 3D vision sensor

Property	Ideal case	Minimal expectation	Penalty
Mass	As low as possible	$m \leq 100 \text{ g}$	Embedment in robot impossible
Spatial resolution	As high as possible	$r_x, r_y, r_z \geq 0.1 \mu m^{-1}$	Assembly impossible (not accurate enough)
Volume imaged	As high as possible	$L_x, L_y, L_z \geq 1 \text{ mm}$	Not enough information (local scan required)
Frame rate	As high as possible	$R \geq 10 \text{ fps}$	Low assembly speed

Chapter 3. Depth from focus vision system for micro assembly

- Opto-electronic component : image sensor.
- Optical component : image formation system.
- Mechanical component : vertical translation mechanism (z -motor).
- Software component : control of camera displacement, sharpness maximization and depth determination algorithms.

The next section presents a reference depth from focus system, which is too bulky for embedded applications. Therefore, sections 3.3 to 3.6 present an evaluation of the feasibility of an *embedded* 3D microscope based on the depth from focus principle [83]. The discussion presented below is focused on the effect of embedment constraints (principally mass) on the optical and electronic component, since those hardware components are the most critical with respect to 3D imaging performance.

3.2.2 Reference macroscopic implementation

As a reference in the work on miniaturization, we consider a macroscopic 3D microscope. The key elements of this system are summarized in table 3.2. The hardware, illustrated in figure 3.3, is based on a Leica MZ-12 stereo microscope [73], with a phototube supporting a C-mount Pulnix TM-1001 camera [98]. The depth scanning motion is provided by a Märzhäuser focus motor [78]. A custom software [125] developed for the Windows operating system, handles microscope z and sample xy motions, and automates the image stack acquisition. Several sharpness function are implemented to compute the DfF range maps. The software also includes multi-resolution filtering for noise reduction [126], and a 3D rendering module for output data visualization.

The main characteristics of this system are summarized in table 3.3. Note that it is well suited for the inspection of micro-mechanical parts with dimensions between $200\mu\text{m}$ up to 10mm . But, since the hardware used is bulky and massive, it cannot be embedded in an assembly system for visual feedback.

3.2.3 Limitations

The depth from focus method described above isn't appropriate in all situations. In particular, since the depth determination is based on local sharpness, the scenes mea-

Table 3.2: Main components in macroscopic 3D depth from focus microscope [125].

Element	Description
Image sensor	Pulnix TM 1001 progressive-scan CCD camera 997×1016 , monochrome, 1 inch sensor.
Image formation system	Leica MZ-12 stereo-microscope , planapo objective, variable zoom: $1\times$ to $10\times$, trinocular video/phototube with C-mount camera adapter.
Vertical motion actuation	Märzhäuser focus motor, MC-2000 controller with serial interface to PC.
Software	Custom software for Windows OS, controls microscope z and sample xy motions, five sharpness function are implemented, includes multi-resolution filtering for noise reduction.

3.2. Depth from focus measurement technique



Figure 3.3: Depth from focus 3D microscope: Leica MZ-12 with Pulnix TM-1001 camera.

Table 3.3: Characteristics of macroscopic depth from focus microscope system

Operation	Range map size	Field of view	Best case accuracy	Typical accuracy	Frame rate
Zoom 1×	997 × 1016	10 × 10 mm	50 μm	250 μm	0.02 fps
Zoom 10×	997 × 1016	1 × 1 mm	5 μm	25 μm	0.04 fps

Chapter 3. Depth from focus vision system for micro assembly

ured are required to have local intensity structures. Highly polished surfaces, such as mirrors, cannot be measured directly with depth from focus; if possible such objects should be coated prior to measurement. Uniformly colored objects are also problematic to acquire with depth from focus imaging: the sharpness isn't defined for uniform regions and this results in random depth values. In that case, projecting a light structure on this object improves the measurement [74, 92]. But this approach involves practical difficulties since it requires a light projection pattern with resolution similar to the desired range accuracy.

Finally, the depth from focus approach is valid only if the magnification of the microscope system is high. When magnification is small, the depth of field becomes large, so that the vertical resolution of the system is reduced. Moreover, if a large depth range is scanned, artifacts due to out of focus objects may appear. These artifacts are removed when telecentric lenses are used (see sec A.5), or when the magnification is increased.

3.3 Image sensor for miniature system

The image sensor is the first part to take into account when designing a miniature depth from focus system. The system optics are then matched to the sensor. The relevant parameters are : sensor lateral dimension L , amount of pixels, light sensitivity, electronic noise, and maximum sustainable frame rate. Miniature image sensors is a rapidly evolving field, as the huge cell phone camera market leads to strong competition. However, commercial systems available in low volumes are still rare. For our prototype system[83], a Kappa CH-166 micro camera head was used [63]. Its properties are compared to the image sensor from our reference system in table 3.4. The choice of this particular sensor head is motivated by its small dimensions and weight, and also by its good signal to noise ratio. Note that the number of pixels is lower than in the reference system. The technological limitation in this case is light diffusion inside the semiconductor material of the CCD device [114]: for pixels smaller than $3\mu\text{m}$, this diffusion would result in blur, as photo-electrons could be collected by a gate different from the photon entry region. The light sensitivity is also lower than in the reference case; this is expected since the pixel size is lower for the micro-camera.

The camera solution chosen for the miniature prototype, shown on figure 3.4, is a good compromise: its PAL resolution (752×582) is adequate for DfF processing, its video frame rate (25fps) allows for quick data acquisition, and its size and weight are consistent with embedment requirements. Unfortunately, the choice of optics for this camera head is limited. The shortest objective focal length available was $f = 15\text{mm}$, for an objective diameter of 7mm. As we will see in the next section, this set of optics isn't optimal for DfF imaging.

3.4 Optics of depth from focus imaging

Good optics is key to superior image quality, which is a requirement in computer vision applications. The accuracy in depth determination will be at best of the same order of magnitude as depth of field. Depth of field is defined as the maximum displacement in depth for an object while its image blur stays confined within one pixel of the sensor.

3.4. Optics of depth from focus imaging

Table 3.4: Comparison between reference image sensor and micro-camera head used in prototype miniature system.

Property	Reference: TM-1001	Prototype: CH-166
Photosensitive area [mm]	9.1 × 9.2	2.5 × 1.8
Pixel count $c \times r$ [1]	1008 × 1018	752 × 582
Pixel size [μm]	9.0 × 9.0	3.0 × 3.0
Light sensitivity [lux]	1.0	2.67
S/N ratio [dB]	50	50
Frame rate [fps]	15	25
Size [mm]	44 × 48.5 × 136	30 × 7 \emptyset
Mass [g]	330	12



Figure 3.4: Kappa CH-166 camera head with 15mm objective.

Chapter 3. Depth from focus vision system for micro assembly

Using a simple, single lens model, depth of field DoF can be expressed [125] as :

$$DoF = \frac{2 \cdot \left(\frac{N_s \cdot \epsilon}{X} + 1 \right) \cdot \epsilon \cdot f \cdot D}{D^2 \cdot \left(\frac{N_s \cdot \epsilon}{X} \right)^2 - \epsilon^2} \quad (3.1)$$

where:

- f is the focal length of the optical system,
- D is the optical system entrance pupil diameter,
- ϵ is the imaging sensor pixel pitch,
- N_s is the lateral extension of the sensor (number of pixels),
- X is the lateral extension of the image field.

Note that the term $\frac{N_s \cdot \epsilon}{X}$ is simply the optical magnification M .

Equation 3.1 clearly shows that a short depth of field is obtained with short focal length, high magnification, and large entrance pupil diameter. Depth from focus measurement requires having the shortest possible depth of field. But the entrance pupil diameter is limited by weight considerations in a local sensor. Similarly, reducing the focal length will reduce working distance for the sensor. The curve in figure 3.5 shows predicted depth of field a miniature imaging system when N_s , ϵ , D and f are fixed, and the optical magnification $M = \frac{N_s \cdot \epsilon}{X}$ is varied to accommodate for different object size into the field of view X . The parameters chosen for this plot were: $N_s = 752$, $\epsilon = 3\mu\text{m}$, $D = 7\text{mm}$, $f = 15\text{mm}$. Those values correspond to micro camera system introduced in section 3.3, while the range for field dimension corresponds to the requirements in section 3.1.2. Three representative magnification values ($M = 0.3, 1.0, 3.0$) are reported on the curve. In figure 3.5, asterisks represent the results of experimental accuracy measurements (see section 3.7.2). Those measurements are in good agreement with the theoretical expectations, indicating that the single lens model is indeed a valid description of our optical system.

3.4.1 Available magnifications for miniature prototype

As mentioned above, the single objective fitting for microscopic imaging with the Kappa CH-166 camera head has a focal length of 15mm. The standard position of this objective relative to the camera results in a magnification $M = 0.36$. Since higher magnification are desired, 3 special spacer elements were produced, to be added between the objective and imager. Adding a spacer element increases the image distance d_i ; since the focal length f is fixed, the image formation equation requires a reduction in object distance d_o . This results in an effective magnification increase, coming at the cost of a reduced working distance (between sample and objective). The attained magnifications with the 3 spacer elements are reported in table 3.5. The 4 discrete magnification values go from 0.36 to 1.85. Note that, the largest field of view is 6.7×5.1 mm. In the situation with highest magnification, the object distance d_o is relatively short: 40 mm. Higher magnifications would require to put the objective very close to the sample.

3.4. Optics of depth from focus imaging

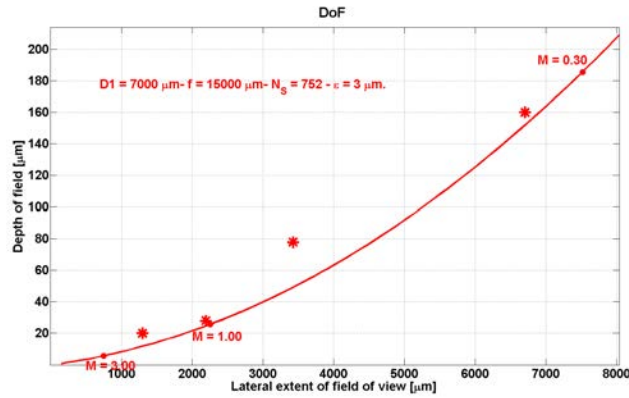


Figure 3.5: (Plain curve) Depth of field expectation for miniature imaging system - (Asterisks) Measured depth accuracy (see section 3.7.2)

Table 3.5: Available magnifications for miniature imager prototype

Spacers	0	1	2	3
Field of view [mm]	6.7×5.1	3.4×2.6	2.2×1.7	1.3×1.0
Magnification	0.36	0.72	1.10	1.85

Table 3.6: Comparison between reference optics and objective used in prototype miniature system.

Property	Reference: Leica MZ-12	Prototype: 15mm objective
Entrance pupil \emptyset [mm]	45	7
Magnification	from $1\times$ to $10\times$	$0.36\times, 0.72\times, 1.10\times, 1.85\times$
Max. depth of field [μm]	100	160
Min. depth of field [μm]	1	10
Size [mm]	$220 \times 280 \times 400$	$15 - 40 \times 7 \emptyset$
Mass [g]	3000	3 - 5

Chapter 3. Depth from focus vision system for micro assembly

3.4.2 Comparison of macroscopic and miniature optics

The main characteristics of the optics used in the prototype miniature system[83] are compared reference system's optics in table 3.6. See appendix A for a more detailed comparison. Note that the small size optics result in a significantly larger depth of field, which is a severe penalty for depth from focus imaging. Moreover, the prototype system is less flexible than the macroscopic system, since only 4 discrete magnification values are available. But these optics fit into the size and mass requirements for an embedded system.

3.5 3D microscope camera motion

In a depth from focus system, the focus is changed by moving the camera relative to the scene. Here, three conflicting requirements must be taken into account. In order to have a high frame rate, the motion should be fast, so that the whole scene is scanned in a short amount of time. But the motion system must provide high resolution positioning, to reliably produce depth differences between each image in the image stack. Finally, the camera motion system should be small enough to be embedded.

For the prototype miniature system, priority was given to motion accuracy, and the same motors than in the reference macroscopic system were used. Unfortunately, this choice is incompatible with the mass and speed requirements for an embedded system. In the next paragraph, a set of specification is provided for an actuation system that would comply with those requirements.

3.5.1 Specifications for a micro-camera focus actuator

The purpose of the actuator is to move the micro-camera relative to the object under observation, in the vertical direction. The desired effect is a controlled change in acquired image focus. An actuator fitting those specifications would enable to build a depth from focus system adapted to a work volume of $10\text{mm} \times 10\text{mm} \times 10\text{mm}$, where this volume would be scanned once per second. The system could thus deliver a new range image each second, which is sufficient for robot control in many assembly operations.

Table 3.7: Specifications for a micro-camera focus actuator; see fig 3.6

Actuator load (micro-camera with objective) :	
mass :	$5.0 \text{ g} < m_c < 10 \text{ g}$
diameter :	$d = 7 \text{ mm}$
length :	$40 \text{ mm} < L < 55 \text{ mm}$
Stroke :	$10 \text{ mm} < \Delta z < 20 \text{ mm}$
Precision :	$\epsilon_z < 5 \text{ }\mu\text{m}$
Velocity (on stroke Δz) :	$v_z > 10 \text{ mm/s}$
Actuator mass :	$m_a < 80 \text{ g}$

3.5. 3D microscope camera motion

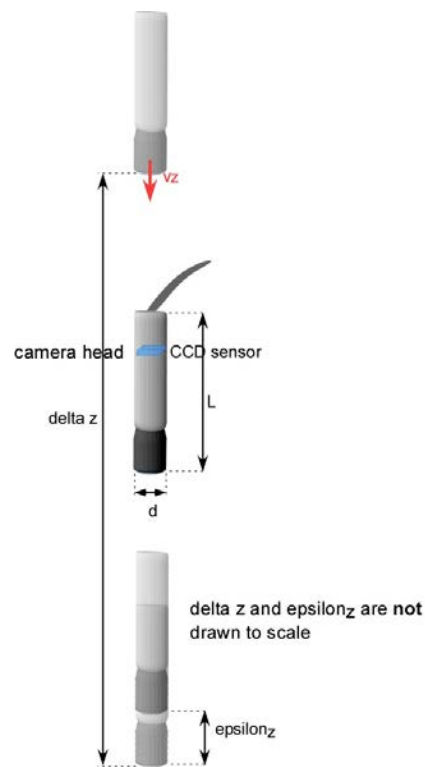


Figure 3.6: Micro-camera displacement - Schematic

3.6 Depth from focus software processing

The key element in depth from focus software processing is sharpness analysis. According to equation 2.1, the locations of sharpness maxima across the image stack can directly be reported as the depth image. The following sections are devoted to the presentation of various sharpness operators, their comparison and the presentation of a multiresolution filter procedure for reducing depth image noise. The implementation chosen is identical to the one previously developed [126] for the macroscopic system (see sec 3.2.2).

3.6.1 Sharpness operators

The quality of the range map obtained in depth from focus measurements depends critically on the sharpness operator used. Sharpness is measured by integrating intensity variations over a certain neighborhood, of support w .

3.6.1.1 Difference filter

To isolate high frequency components from depth from focus processing, the absolute of the difference is taken between the original image $I(x, y, z)$ and a low-pass version of this image $\mathbf{I} * \mathbf{K}_{lp}$, where $*$ denotes a 2D convolution and \mathbf{K}_{lp} is a low-pass filtering kernel. In the DFF software implementation used, it was chosen as:

$$\mathbf{K}_{lp} = \frac{1}{16} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} \quad (3.2)$$

The sharpness map \mathbf{S}_{dif} is then given by :

$$\mathbf{S}_{dif}(\mathbf{I}) = |\mathbf{I} - (\mathbf{I} * \mathbf{K}_{lp})| \quad (3.3)$$

Note that in terms of processing time, the computation of the sharpness map depends mostly on the speed of the convolution operation. Different image processing libraries [55, 80] offer convolution functions, with varying degrees of optimization.

3.6.1.2 Variance filters

Sharpness can be estimated by computing the grayscale variance over a neighborhood $B_x \times B_y$ of size $w \times w$:

$$\mathbf{S}_{var}(\mathbf{I}) = \frac{1}{w^2} \sum_{x' \in B_x} \sum_{y' \in B_y} |I(x', y') - \mu(x, y)| \quad (3.4)$$

where

$$\mu(x, y) = \frac{1}{w^2} \sum_{x' \in B_x} \sum_{y' \in B_y} I(x', y') \quad (3.5)$$

Zamofing & Hügli [125] proposed an efficient implementation of variance computation based on convolutions with a constant box filter \mathbf{H} of size $w \times w$:

$$\mathbf{S}_{var}(\mathbf{I}) = \mathbf{I}^2 * \mathbf{H} - (\mathbf{I} * \mathbf{H})^2 \quad (3.6)$$

3.6. Depth from focus software processing

where the square operation is performed elementwise. This approach allows to compute variance in a time nearly independent of w , for reasonable values of the neighborhood size: $w < 25$. In the software implementation used, 3 different values can be selected for w : 3, 5 or 21.

3.6.1.3 Laplacian filters

The discrete Laplacian filter produces an estimation of second derivative of an image. Taking the absolute value of this second derivative gives a sharpness estimate :

$$S_{lap}(\mathbf{I}) = |\mathbf{I} * \mathbf{K}| \quad (3.7)$$

where \mathbf{K} is for example a Laplacian kernel . The software implementation includes a 3×3 kernel \mathbf{K}_3 and a 5×5 kernel \mathbf{K}_5 :

$$\mathbf{K}_3 = \begin{pmatrix} 1 & 2 & 1 \\ 2 & -12 & 2 \\ 1 & 2 & 1 \end{pmatrix} \quad \mathbf{K}_5 = \begin{pmatrix} 1 & 2 & 3 & 2 & 1 \\ 2 & 4 & 8 & 4 & 2 \\ 3 & 8 & -84 & 8 & 3 \\ 2 & 4 & 8 & 4 & 2 \\ 1 & 2 & 3 & 2 & 1 \end{pmatrix} \quad (3.8)$$

A 21×21 Laplacian filter is also included. In the implementation, this filter is emulated by a DOB filter where the difference is taken from images convolved with filters expressed as separable gaussian kernels [127]:

$$\mathbf{K}_{21} = (\mathbf{g}_1^t * \mathbf{g}_1 - \mathbf{g}_2^t * \mathbf{g}_2) \quad (3.9)$$

where \mathbf{g}_1 and \mathbf{g}_2 are one-dimensional gaussian kernels; $\mathbf{g}_1 = (0, 0, 0, 1, 3, 8, 19, 37, 59, 79, 86, 79, 59, 37, 19, 8, 3, 1, 0, 0, 0)$; $\mathbf{g}_2 = (2, 4, 6, 10, 16, 22, 30, 38, 45, 50, 52, 50, 45, 38, 30, 22, 16, 10, 6, 4, 2)$. Using separable gaussian kernel box filters allows to speed up computation times, since the 2D convolution can be replaced by two 1D convolutions.

3.6.2 Comparison of sharpness operators

Sharpness operators are employed on each image in the image stack in order to establish the range map. Therefore, the time required for this processing is critical when real-time applications are considered. Figure 3.7 presents a comparison of sharpness processing times, for different image sizes, measured on a Core2Duo 2.66GHz computer. The image sizes considered are: the native resolution of the macroscopic system: 997×1016 ; the native resolution of the micro camera used in miniature prototype: 768×576 ; and two synthetic image sizes obtained by selecting a region of interest (ROI) on existing image stacks: 512×512 and 256×256 . Note that the 3×3 difference filter is slow since it wasn't optimized. All other methods use optimized functions defined in [55]. As expected, processing times do not depend on the window size w for variance operators. The largest image size is incompatible with real-time operation, as the processing time for a single image can exceed 70 ms. In contrast, restricting the analysis to a 256×256 ROI ensures that the processing time remains below 10ms, which is an adequate value for real-time depth from focus imaging.

All sharpness operator are sensitive to noise, and this behavior can cause measurement errors. One approach to limit the influence of noise is to increase the support w : integrating variations over a larger area should give smoother results. However, in

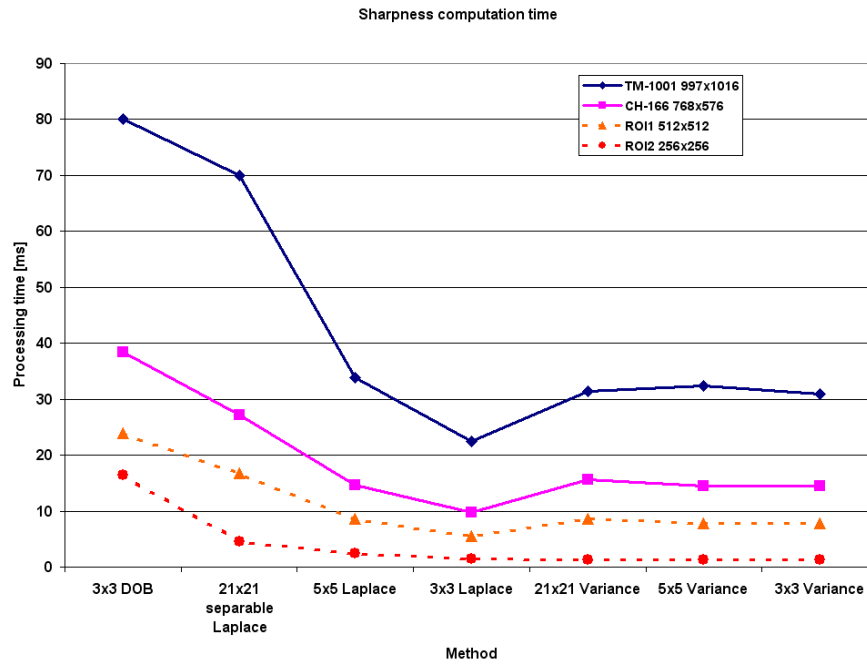


Figure 3.7: Comparison of sharpness processing time, for different image sizes, measured on a Core2Duo 2.66GHz computer. The 3x3 DOB method is slowest since it uses unoptimized convolution functions. Note that processing times do not depend on the window size w for variance operators in this implementation.

many experimental situation, the better approach is to use a small support sharpness operator, and to perform noise filtering in post processing steps. This allows to retrieve the finest resolution features of the scene. This is illustrated by the example of table 3.8. In this example, the depth from focus processing was applied to one image stack of resolution 768×576 , and 5 different sharpness operators were used. The middle column in table 3.8 shows the produced range maps. Computation intensive operators such as 21×21 variance or Laplace do not outperform the small support operators such as 3×3 or 5×5 Laplace.

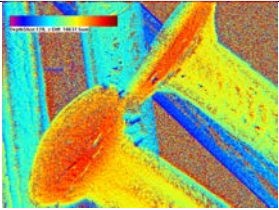

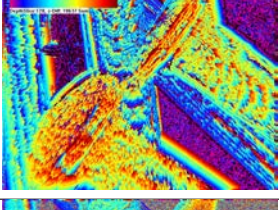
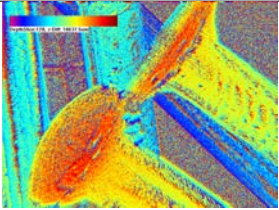
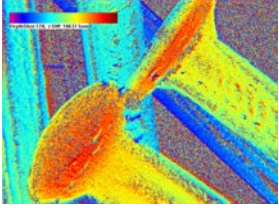
It can be observed in table 3.8 that regions where no object was imaged in the scene have random depth values. In order to provide a realistic depth map, these outliers must be filtered out. An efficient, multi-resolution noise filtering scheme for depth from focus measurement is presented in Zamofing & Hügli [126]. This noise filtering step was also used in this work, and allowed to produce the depth maps illustrated in figure 3.10.

3.6.3 Multiresolution filtering

Zamofing & Hügli [125] proposed a multi-resolution analysis scheme to lower the noise in DFF measurements. Multi-resolution analysis uses the sharpness information as a confidence indication (fig. 3.8, top row) for the depth value measured (fig. 3.8, middle row). Combining depth maps at different resolutions (6 levels in figure 3.8) allows to obtain a filtered depth image (fig. 3.8, bottom left). Parameters for this

3.6. Depth from focus software processing

Table 3.8: Comparison of sharpness operators. Using sharpness operators with a larger support region w does not necessarily produce better range images.

Operator	Range map	Processing time
DOB (Difference of boxes)		50 ms/image
Variance 21x21		15ms/image
Laplacian 21x21 (sep)		27 ms/image
Laplacian 5x5		14 ms/image
Laplacian 3x3		10 ms/image

Chapter 3. Depth from focus vision system for micro assembly

combination (number of resolution levels, weight of each level) must be adapted for each type of sample observed, in order to have a good balance between preservation of high frequency features and noise in the output image. Noise is still present in the final image; but the readability of the image was significantly improved when compared to initial high-resolution depth.

3.7 Miniature system - Experimental results

The implemented miniature imager uses a Kappa CH-166 micro-camera, that contains an 1/6" CCD sensor with PAL resolution (768×576). The pixel pitch is $\epsilon_x, \epsilon_y = 3.0 \mu m \times 3.0 \mu m$. The key elements of this system are summarized in table 3.2. Depending on the field to cover, different optical magnifications M must be employed. The camera objective has focal length $f = 15 mm$, and spacer elements between objective and sensor allow to span magnifications ranging from $M = 0.25$ to $M \approx 2.0$. The micro-camera is shown on figure 3.4. The mass of the imager device (including objective) does not exceed $20 g$. When compared to the expectations in table 3.1, this indicates that a mass budget of $80 g$ can be spent on the z-motor in the development of the embedded system. Note that currently, the prototype setup uses much bulkier motors ($mass > 2 kg$); this first step allows to evaluate the miniature imager only.

The software component, realized as a C++ application on PC, is identical to the reference system. This application controls camera displacement, image acquisition, sharpness maximization, noise filtering and display of range maps. Image acquisition is performed through a Matrox Meteor II frame-grabber. The sharpness evaluation algorithms are implemented either with MIL [80] or OpenCV [123], and sharpness maximization is performed in parallel with image stack acquisition. Nevertheless, sharpness determination remains a time consuming operation, especially when a large kernel is used.

3.7.1 Sample images

Depth maps acquired with the prototype miniature system illustrate the adequacy of the micro camera component when considering the task proposed in section 3.1. The first sample is a detail of a screw viewed from top (figure 3.9), that was acquired at high magnification ($M = 1.85$). The dimensions of the scene $L_x \times L_y \times L_z$ are approximately $1.3mm \times 1.0mm \times 2.5mm$. This example shows the potential of the miniature system to provide accurate depth data at high resolution, for high aspect ratio scenes. The 3D rendering allows to see that the system is able to measure the slope in the screw helix. The second sample (figure 3.10) was acquired with low magnification ($M = 0.28$). The image shows a random arrangement of nails, which serves as an example of bulk part feeding situation[50]. With range information, it is easy to distinguish the top nails from the bottom ones, so that an assembly robot can be programmed to automatically pick one of these parts for assembly.

3.7.2 Depth resolution

Resolution evaluation is difficult for passive 3D measurement systems, since range image quality depends on the image contents. To estimate depth resolution, simple scenes (described by a simple geometric model) are measured. The scene used in our experiments, shown on figure 3.11 (a), features two identical disks, with diameter

3.7. Miniature system - Experimental results

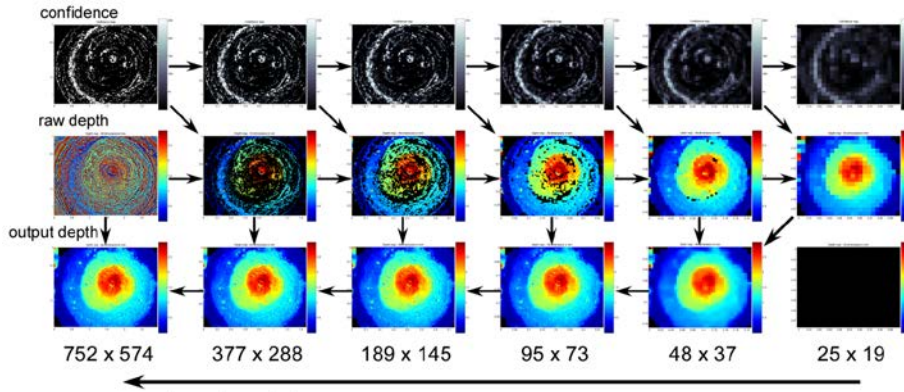


Figure 3.8: Overview of multiresolution filtering for DFF imaging. In this example, 6 resolution level are used.

Table 3.9: Main components in miniature depth from focus prototype [83]. Compare with reference system components in table 3.2. The miniature prototype can be used to check the performance of micro-camera and associated optics.

Element	Description
Image sensor	Kappa CH-166 interline transfer CCD camera 768 × 576, color, 1/6 inch sensor.
Image formation system	15mm objective, spacers for different magnifications, 0.36×, 0.72× 1.10× 1.85× .
Vertical motion actuation	Märzhäuser focus motor, MC-2000 controller with serial interface to PC.
Software	Custom software for Windows OS, controls microscope z and sample xy motions, five sharpness function are implemented, includes multi-resolution filtering for noise reduction.

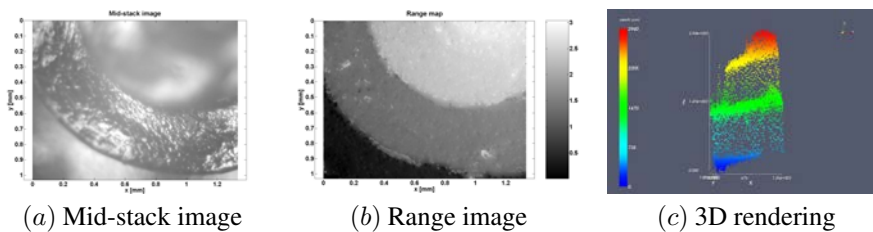


Figure 3.9: Sample image for miniature system : screw tip. The helix shape of the screw tip is measured accurately.

Chapter 3. Depth from focus vision system for micro assembly

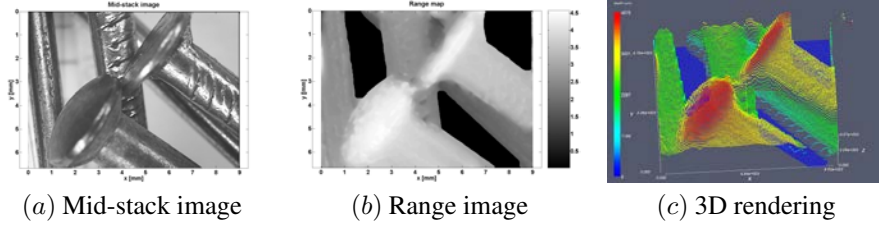


Figure 3.10: Sample image for miniature system : nails. The range image allows to determine the relative positions of the nails.

19 mm. The height difference between the parts is 1.5 ± 0.1 mm. This scene was imaged with the miniature prototype in different optical configurations (labels 1 to 4 in image 3.11 and table 3.10), resulting in increasingly narrow fields of view (see figure 3.11 (a)). For each test scene, the number of images in the stack was 128, the scanned depth range was $\Delta Z \approx 2$ mm. The geometric model for each disk is a perfectly flat and horizontal surface. In each range map, two regions of interest (covering approximately one quarter of the image field) are selected: R_b (on the bottom disk) and R_t (on the top disk), and the deviations σ_b and σ_t from the perfect planar model are measured for both R_b and R_t . The system resolution is then estimated as the range standard deviation σ , averaged for those two regions of interest :

$$\sigma = \frac{\sigma_l + \sigma_h}{2} \quad (3.10)$$

Measurement results are summarized in table 3.10, which also recalls accuracy values obtained with high-performance macroscopic systems [125], as reference values. Accuracy results for scenes 1 to 4 were also reported on figure 3.5, where they can

Table 3.10: Measurements with miniature imager (1 to 4), compared to accuracy obtained with microscope setups [125] (5, 6)

Scene	1	2	3	4	5	6
Imager	Miniature	Miniature	Miniature	Miniature	MZ12	DMLA
Field of view [mm]	6.7×5.1	3.4×2.6	2.2×1.7	1.3×1.0	0.9×0.9	0.3×0.3
Magnification	0.36	0.72	1.10	1.85	10.0	31.0
σ [μm]	160	78	28	20	5	2

be compared to the depth of field values. As expected, the highest depth accuracy is obtained for narrowest field of view. The best accuracy obtained with our miniature imager is $20\mu\text{m}$. This result does not comply with the expectations of table 3.1 (a factor of 2 is missing). A bulky microscope is 10 times more accurate (DMLA [125]). Figure 3.12 shows a comparison of test scene acquired with miniature imager or microscope. However, such a microscope is typically heavier than 3 kg. The observed 10-fold reduction in accuracy follows a 150-fold reduction in mass. Such a tradeoff is necessary if an embedded system is to be developed.

3.7. Miniature system - Experimental results

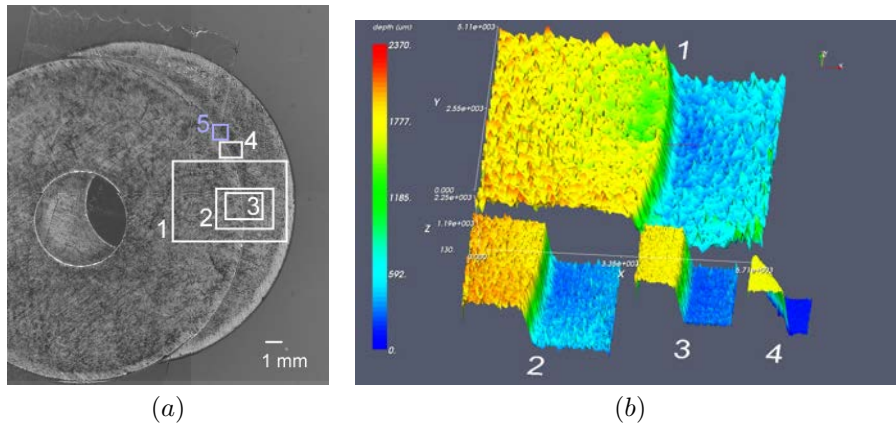


Figure 3.11: Resolution measurement (Stacked metal disks) - (a) top view of the scene - (b) 3D rendering of measured range data for different fields of view (see table 3.10)

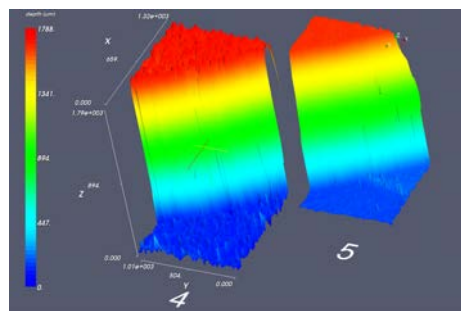


Figure 3.12: Range image acquired with miniature system (left) - Range image acquired with microscope system (right)

3.8 Miniature system - Perspectives

As mentioned above, the design of an embedded 3D vision sensor is far from completion. Many challenges remain if all expectations defined in table 3.1 are to be met. Table 3.11 summarizes the performance of the developed system and lists some perspectives for improvement in each area.

3.8.1 Low-mass depth from focus motor

We have seen that depth from focus with a miniature imager can reach accuracy specifications in the order of $20\mu\text{m}$, with a mass budget of 20g for the imager. The following step in development of an embedded depth from focus sensor is the selection of an appropriate z-motor, capable of moving this 20g imager mass over a stroke of 5mm or more, while the motor mass stays under 80g. This step is required for the completion of a first embedded 3D sensor prototype. Apart from mass, criteria to consider for motor selection are : linear accuracy, length of stroke, and speed of operation. For a first embedded prototype, mass and linear accuracy are considered critical, while length of stroke and speed of operation are secondary.

3.8.2 High-frame rate imaging

In order to meet the frame rate specification of table 3.1, additional steps are required. First, the z-motor must move the imager package with 10Hz period. Second, the image sensor must acquire images at a rate of 200fps (under the assumption that one range image requires a stack of 20 2D images). Finally, the software component must compute sharpness images at the same rate, i.e. in less than 5ms. To reach this goal, sharpness determination could be performed on a small region of the image only (using a 256×256 region of interest allows to reduce the computation time by a factor higher than 10). Alternatively, smart imagers with on-chip processing could be used to speed up the computation of sharpness values.

Table 3.11: Present performance of miniature depth from focus prototype, with perspectives for improvement.

Property	Expectation	Current implementation	Improvement by
Mass	$m \leq 100\text{ g}$	$m > 2000\text{ g}$ (imager : 20 g , motor : 2000 g)	Low-mass z-motor
Spatial resolution	$r_x, r_y, r_z \geq 0.1\ \mu\text{m}^{-1}$	$r_x, r_y, r_z = 0.6 \times 0.6 \times 0.05\ \mu\text{m}^{-3}$	(Higher optical aperture)
Volume imaged	$L_x, L_y, L_z \geq 1\text{ mm}$	$L_x, L_y, L_z = 1.3 \times 1.0 \times 3.0\text{ mm}^3$	
Frame rate	$R \geq 10\text{ fps}$	$R < 0.25\text{ fps}$	Fast z-motor, fast camera, ROI processing

3.9 Conclusion

This chapter considered the development of a 3D vision system suited for a micro-assembly robot. The depth from focus approach was selected since it turns the small depth of field observed at high magnifications as an advantage rather than an inconvenience. During the first step of development presented in this work, a depth from focus system using miniature image sensor and optical imaging system, but bulky z-motor, was realized, in order to evaluate the performance of a miniature imager.

Experiment results show that the depth resolution for the system with miniature imager (mass < 20g) is close to 20 μ m, which represents a resolution degradation by a factor of 10 when compared with a classical system based on a bulky microscope. The next step identified in the design of an embedded 3D sensor based on the depth from focus approach is the integration of a low-mass z-motor. Finally, a fully functional embedded 3D sensor supposes real-time processing. Experiments have shown that the target computation time of 5ms can be reached for various sharpness operators if the field of view is limited to 256 \times 256 region of interest.

4

Time-of-flight camera system

This chapter begins by recalling the basic physical principle used in time-of-flight measurements and discusses the main characteristics of continuous wave (CW) time-of-flight imaging. The analysis then focuses on the Swissranger [82] device, which was used throughout this work for acquisition of experimental data. This study aims to improve the quality of the depth information returned by the camera by identifying and trying to reduce the causes of erroneous measurements. Noise, causing stochastic errors, is measured (sec. 4.3) and compared to deterministic errors caused by multipath and scattering (sec. 4.5).

4.1 Characteristics of CW TOF cameras

The principle of operation of TOF cameras was presented in section 2.2.2. Here, we recall briefly the characteristics of CW TOF cameras. This study allows to better understand the different error sources to account for with such devices.

4.1.1 Continuous wave TOF signal demodulation

The demodulation of the return signal is realized by sampling it at defined intervals inside its period. The Nyquist theorem states that at least 2 samples per period are required for signal demodulation. Practical considerations of pixel design [19] push towards employing an even number of samples per period. Figure 4.1 shows an example of a *4-tap* demodulation scheme: the photogenerated charges $c(t)$ are sampled 4 times during the period. Then the amplitude and phase difference of the signal are then simply computed as :

$$A = \frac{\sqrt{(c(\tau_3) - c(\tau_1))^2 + (c(\tau_0) - c(\tau_2))^2}}{2} \quad (4.1)$$

$$\Delta\varphi = \arctan\left(\frac{c(\tau_3) - c(\tau_1)}{c(\tau_0) - c(\tau_2)}\right) \quad (4.2)$$

4.1. Characteristics of CW TOF cameras

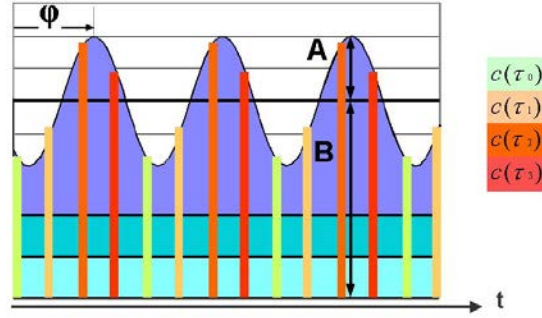


Figure 4.1: 4-taps demodulation of TOF return signal [81]

Finally, the signal amplitude offset B , which corresponds to background light and plays an important role in noise analysis, is computed as the average of the charge measurements:

$$B = \frac{c(\tau_0) + c(\tau_1) + c(\tau_2) + c(\tau_3)}{4} \quad (4.3)$$

Intuitively, it appears that the ratio A/B is a modulation contrast: if this ratio is high, the signal to noise ration (SNR) will be good, whereas the SNR will be bad if this ratio is too small.

4.1.2 Range accuracy limits for TOF cameras

Lange [71] and Büttgen [19] reviewed the factors limiting range accuracy in TOF camera base on solid state sensors. It appears that the main limiting factor is noise. Following Büttgen [19], we use equation 2.4 to define the standard deviation σ_r of the range measurement:

$$\sigma_r(\mathbf{i}) = \frac{c}{4\pi f} \cdot \sigma_\varphi(\mathbf{i}) \quad (4.4)$$

where σ_φ is the standard deviation of the phase difference measurement. The major noise sources listed by Lange [71] are electronic and optical shot noise, thermal noise, reset noise, $1/f$ noise and quantization noise. Lange [71] identified shot noise as being the most fundamental factor limiting range accuracy in CW TOF cameras, since it cannot be circumvented. Moreover, shot noise varies with the incident light intensity [114]. As Büttgen [19] points out, other noise sources contribute to a noise floor level independent of the incident light intensity. While this noise floor level could theoretically be lowered arbitrarily, this is often impractical since it would involve inconveniently long integration times or complex cooling devices to limit temperature effects. The behavior of different noise sources for a typical CCD device [19, 114] is illustrated in figure 4.2. All these factors can be integrated in a noise model [19], allowing to identify possible options to increase the range accuracy (see sec. 4.1.2.1).

4.1.2.1 Noise model for CW TOF camera

Based on an extension of the Lange's [71] derivation for shot noise in TOF devices, Büttgen [19] proposes the following noise model for a TOF camera:

$$\sigma_r = \frac{c}{4\pi f \sqrt{2}} \cdot \frac{\sqrt{B}}{A} \quad (4.5)$$

Chapter 4. Time-of-flight camera system

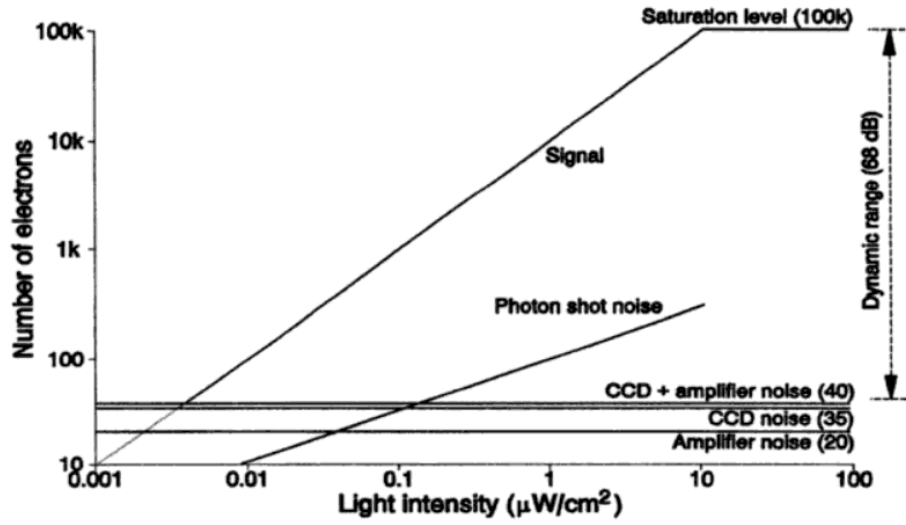


Figure 4.2: Illustration of the different noise sources of a CCD imager in terms of number of electrons and as a function of light intensity - Theuwissen [114]

where (see fig 4.1):

- A is the demodulation amplitude : the number of photoelectrons per pixel and per sampling point generated by the modulated light source,
- B is the offset : the number of photoelectrons per pixel and per sampling point caused by all light sources, including the average of the modulated light.

But the offset B also accounts for the different noise sources. We have:

$$B = A + B_G + B_D + N_{ps} \quad (4.6)$$

where:

- B_G is the background light, dependent on illuminations sources different from the camera's active illumination;
- B_D is the pixel dark current; B_D depends on sensor technology and on temperature;
- N_{ps} corresponds to the number of pseudo electrons generated by all other added noise sources.

This noise model allows to identify three options to improve range accuracy : increasing the modulation contrast $c_m A$; increasing the modulation frequency f ; or decreasing the offset component B . These options are discussed in the following section.

4.1.2.2 Noise reduction in CW TOF cameras

The first option for noise reduction is increased illumination contrast. The appropriate illumination will depend on the targeted application. The SR-3000 camera[82], used indoors, has a fixed LED illumination. But customized illumination allows the

4.1. Characteristics of CW TOF cameras

same technology to be used outdoors in automotive applications [94]. PMD TOF cameras [113] are designed to support different illuminations modules, chosen again depending on the application. However, the active illumination contrast A is always limited by the camera power budget and by eye-safety requirements, so that it can not be increased at will for noise reduction.

Büttgen [19] notes that σ_r varies inversely with the modulation frequency in eq. 4.4. Therefore, increasing the modulation frequency allows to increase accuracy, at the cost of a lowered unambiguous operation range. Technology currently puts an upper limit on the modulation frequency used, since demodulation speed is limited by charge transport in the solid state sensor. But recent advances in sensors allow for example the SR-3000 and SR-4000 cameras to be operated at 30 MHz for increased accuracy.

Dark current and thermal noise are lowered when the device is operated at low temperatures, but this is often not possible in practice. Note that [61] proposed and implemented an optical feedback path to greatly reduce the influence of dark current variations with temperature. The reduction of all other noise sources can only be achieved through improvements in pixel design and demodulation strategies, discussed in detail in [19].

4.1.3 From TOF range maps to cartesian coordinates

TOF cameras produce range measurements: the range data values $\mathbf{r}(i, j)$ from the 2D imager are distances to the center of the camera. For subsequent processing, range maps often need to be converted to a point cloud \mathcal{P} expressed in a cartesian coordinates system: $\mathcal{P} = \{(x, y, z)\}$. This conversion requires an appropriate camera model: the image formation system determines the projection from real-world positions (x, y, z) to pixel coordinates (i, j) . In the following discussion, lens distortion is neglected, and a simple pinhole camera perspective projection model [118] is used.

Let f_o be the camera focal length, d_x and d_y the pixel pitch in the x (resp. y) direction. With (i_c, j_c) as the coordinates of the imaging system's *optical center* on the sensor array, normalized sensor coordinates (X_c, Y_c) are defined as:

$$X_c = (i - i_c) \cdot d_x \quad Y_c = (j - j_c) \cdot d_y \quad (4.7)$$

The transformation between the range map $\mathbf{r}(i, j)$ and 3D coordinates relative to the camera position is then given by :

$$z = r \cdot \frac{f_o}{\sqrt{f_o^2 + ((i - i_c) \cdot d_x)^2 + ((j - j_c) \cdot d_y)^2}} \quad (4.8)$$

$$x = z \cdot \frac{X_c}{f_o} = z \cdot \frac{(i - i_c) \cdot d_x}{f_o} \quad (4.9)$$

$$y = z \cdot \frac{Y_c}{f_o} = z \cdot \frac{(j - j_c) \cdot d_y}{f_o} \quad (4.10)$$

Note that although the point cloud is expressed in 3D, the depth data is measured only for the frontmost surface. This is common to all monocular range imaging systems. This is why such systems are sometimes referred as 2.5D.

4.2 SR-3000 TOF camera characteristics

The SwissRanger[82] SR-3000 and SR-3100 cameras used in our experiment are USB devices, capable of delivering up to 30 frames per second (*fps*). These cameras use

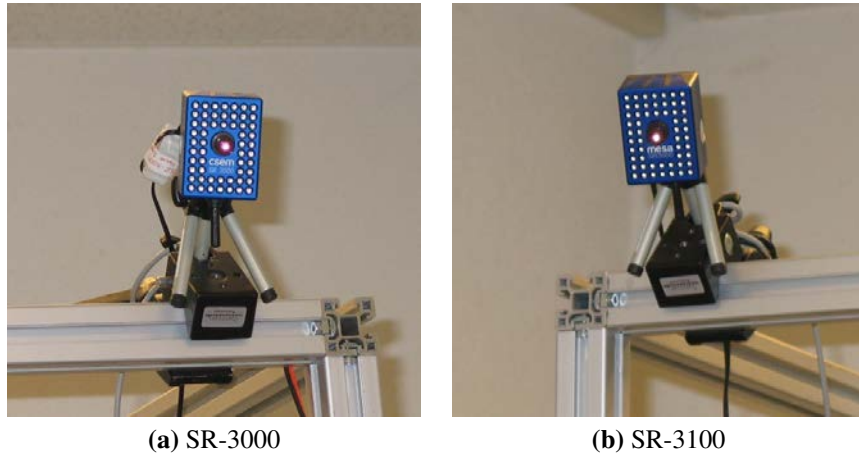


Figure 4.3: SR-3000 and SR-3100 time-of-flight cameras. The SR-3100 camera has improved anti-reflective coatings to reduce scattering effects (those effects will be discussed in section 4.4.2).

continuous wave modulation to provide time-of-flight data. Illumination is provided by 55 infrared LEDs (see fig 4.3). The viewing angle for the camera is roughly 45° , and the focal length is fixed ($f = 8 \text{ mm}$). Unfortunately, the sensor size is currently relatively small (176×144), but next generations of time-of-flight cameras should have improved resolution. Nevertheless, this type of range imager is particularly well suited to applications requiring to detect the position or motion of humans indoors, i.e. safety or human interface systems.

Figure 4.4 shows the range and amplitude image delivered by the SR-3000 camera for an example scene: one person in an office room. Typically, such images are acquired at 20fps. The SR-3000 driver software includes functions performing the transformation from the range map \mathbf{r} to the associated point cloud $\mathcal{P} = \{(x, y, z)\}$ (see sec. 4.1.3), which can be rendered in a 3D visualization software. In figure 4.5, amplitude data is represented by the grayscale intensity. A red-to-blue colorbar is used to colorize the depth data in fig. 4.6. Note that those figures also illustrate data occlusion: objects close to the camera occlude objects farther away ; in fig. 4.6, the person casts a *shadow* over the wall.

The SR-3000 camera is typically operated at a frequency $f = 20\text{MHz}$. Therefore, the unambiguous measurement range R_0 is $[0, 7.5]\text{m}$. The optical power emitted by the camera LEDs is approximately 1W. This is generally sufficient to cover the unambiguous measurement range for indoor operation. Unfortunately, direct sunlight can cause saturation of the sensor pixels, so that this device is not well suited to outdoors usage. Note that the camera user can choose the operation frequency from 4 discrete settings: 19MHz, 20MHz, 20MHz, 30MHz. This option is advantageous in multi-camera configurations, but also to reduce noise levels.

4.3. Noise in SR-3000 cameras

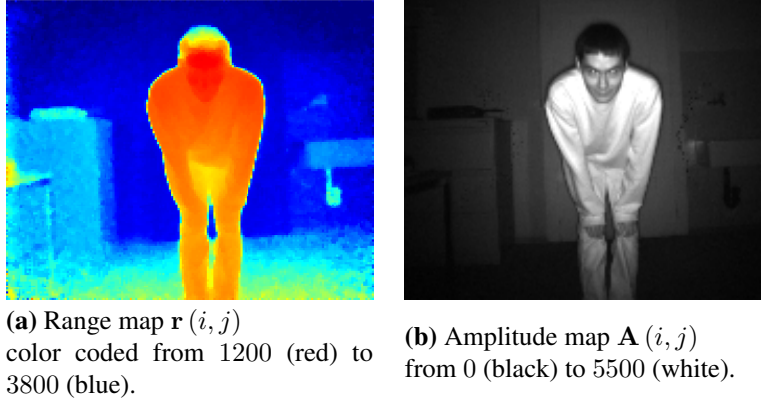


Figure 4.4: Range and amplitude maps returned by the SR-3000 camera for scene containing one person in an office room.

4.3 Noise in SR-3000 cameras

In this section, the limitations in accuracy caused by noise in SR cameras are discussed. Examples of noisy data are provided, and the correlation between noise and the signal amplitude are illustrated. Temporal averaging is proposed to reduce noise in static scenes. For scenes with moving objects, the amplitude measurement can be used to estimate the range error.

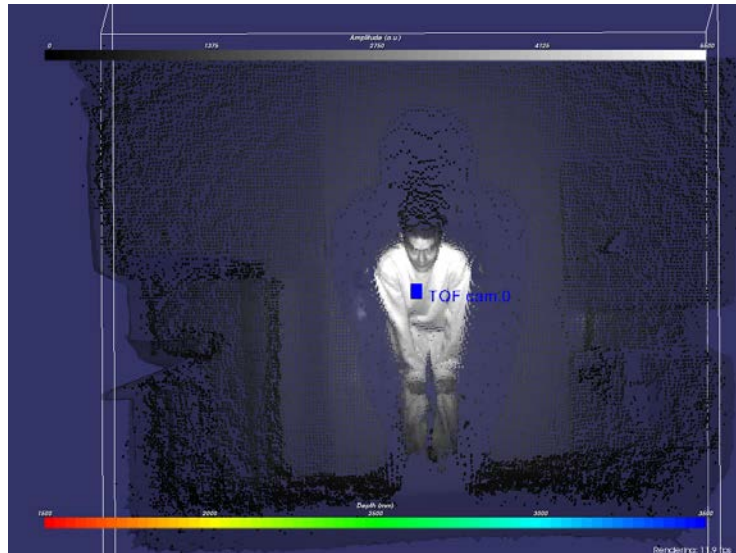
4.3.1 Average noise level

The left column of figure 4.7 provides an illustration of the typical noise level to account for when working with Swissranger cameras. The scene imaged is a part of an empty office, with the door, the side of a desk, and a bookshelf in the background. The camera used is a SR-3100, operated at $f = 20\text{MHz}$. To evaluate quantitatively the noise level in the range map, an individual range image \mathbf{r}_t and an average range image \mathbf{r}_{avg} (from 50 measurements) are compared; see figures 4.8 and 4.9. In this experiment, the two TOF cameras were imaging the same scene, but from different viewpoints. Note that the distribution of range differences is centered around zero, as expected for fluctuations caused by noise. The RMS noise value σ_{avg} for an image of size $N_{rows} \times N_{cols}$ is reported in table 4.1.

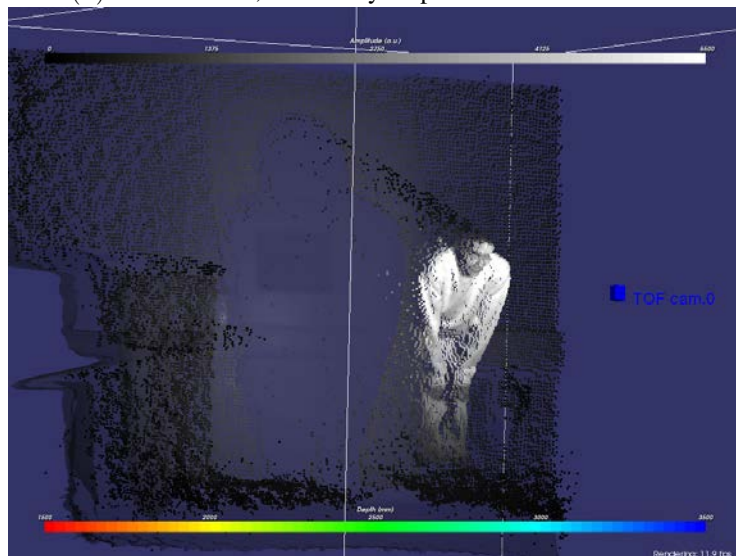
$$\sigma_{avg}(t) = \sqrt{\frac{1}{N_{rows} \cdot N_{cols}} \sum_{i=1}^{N_{rows}} \sum_{j=1}^{N_{cols}} \|r_t(i, j) - r_{avg}(i, j)\|^2} \quad (4.11)$$

Table 4.1: Noise evaluation - RMS range differences

	σ_{avg} [mm]
SR-3100 scene (fig. 4.8)	138
SR-3000 scene (fig. 4.9)	156



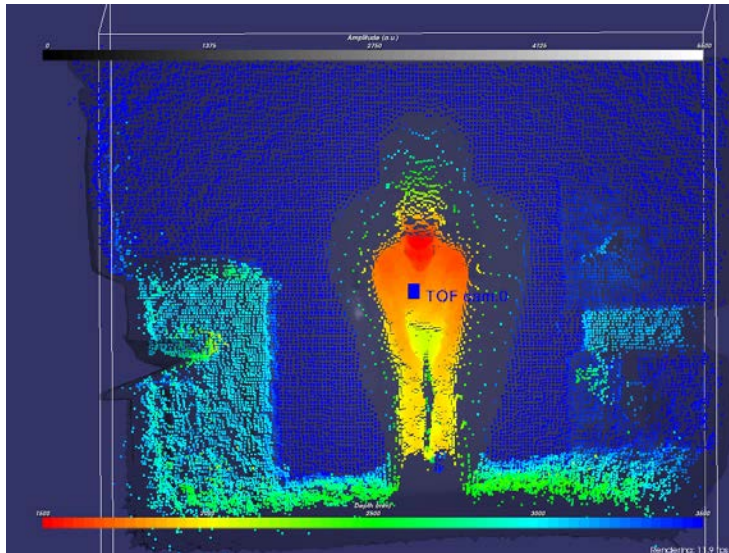
(a) Point cloud \mathcal{P} , colored by amplitude data. Front view.



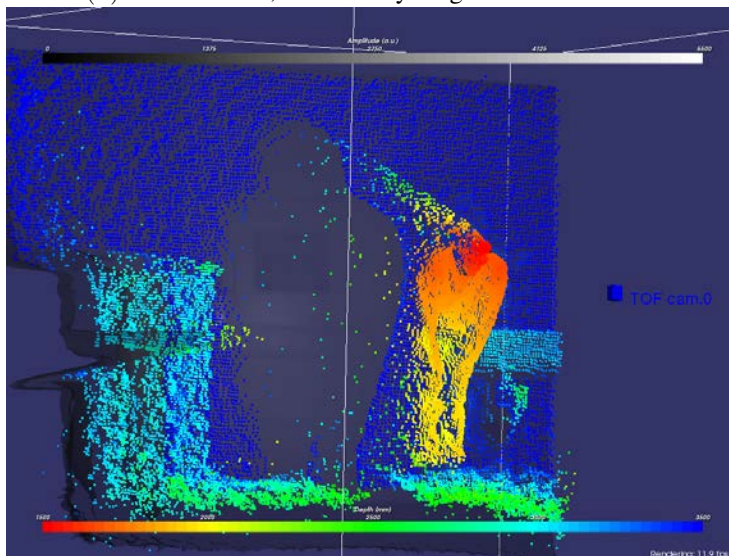
(a) Point cloud \mathcal{P} , colored by amplitude data. Side view.

Figure 4.5: 3D renderings of sample SR-3000 data: the scene shows a single person in an office room. Grayscale intensity maps the point cloud amplitude data, going from 0 (black) to 5500 (white).

4.3. Noise in SR-3000 cameras



(a) Point cloud \mathcal{P} , colored by range data. Front view.



(a) Point cloud \mathcal{P} , colored by range data. Side view.

Figure 4.6: 3D renderings of sample SR-3000 data: the scene shows a single person in an office room. The point cloud is colored with range data, going from 1500mm (red) to 3500mm (blue).

Chapter 4. Time-of-flight camera system

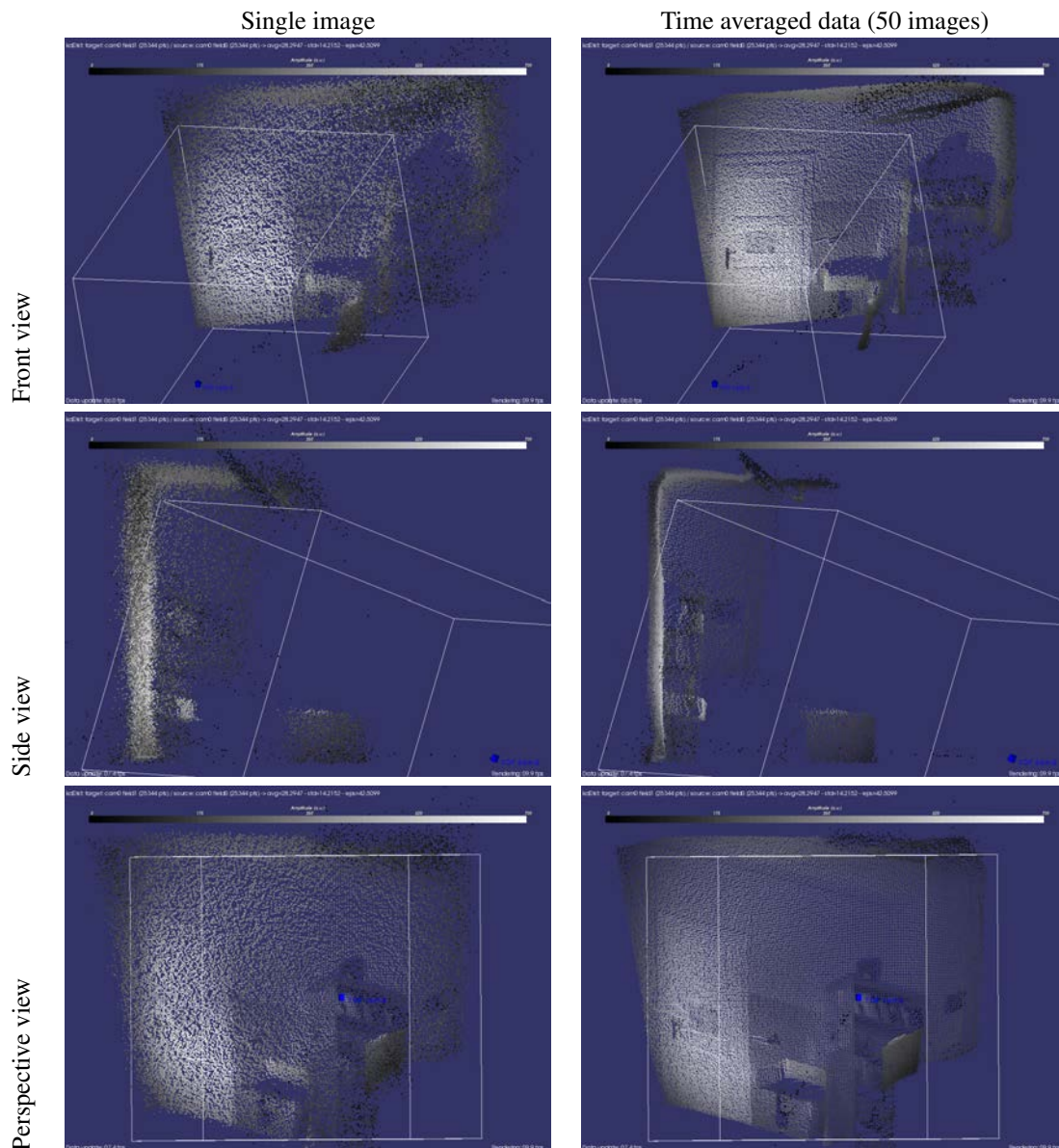


Figure 4.7: Comparison of raw data and averaged data for an empty scene, acquired with a SR-3100 camera. The closest distance from camera to background is approximately 3.7m. For any single image, noise limits the range image accuracy (left column). Time averaging allows to reduce noise effects (right column).

4.3. Noise in SR-3000 cameras

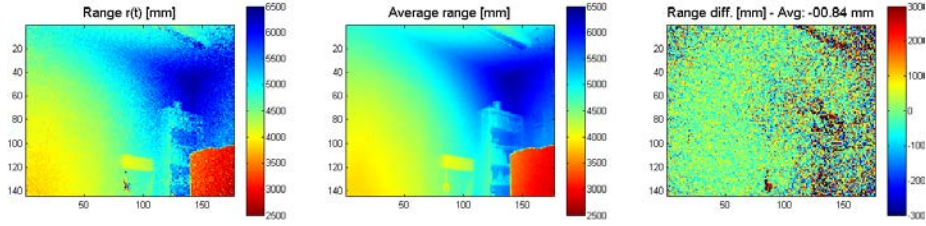


Figure 4.8: SR-3100, Typical indoor scene - Comparison between current range image r_t and average range image r_{avg} .

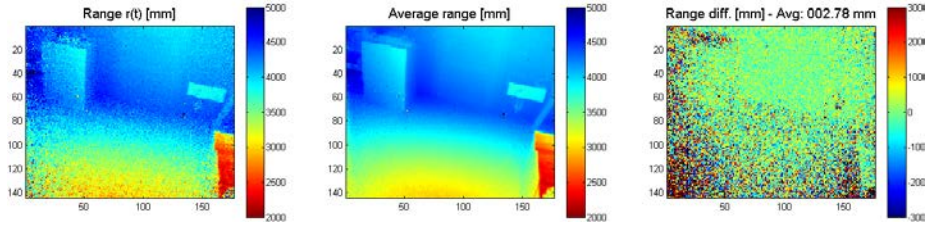


Figure 4.9: SR-3000, Typical indoor scene - Comparison between current range image r_t and average range image r_{avg} .

In these examples, the RMS error is around 15 centimeters. Those temporal variations are not negligible, even for visualization: the point cloud obtained from a single range image (fig. 4.7 left) is much more difficult to interpret than the point cloud obtained from time averaged data.

4.3.2 Amplitude dependent noise

Although the distance measurement in TOF cameras is based on the phase of the returning signal, amplitude plays a critical role in distance measurement accuracy; see eq. 4.5. Figure 4.10 provides another illustration of this behavior. The standard deviation $\sigma_r(i, j)$ of the N_{avg} images used to define the temporal average $r_{avg}(i, j)$ is compared to the amplitude data $A_{avg}(i, j)$.

$$\sigma_r(i, j) = \sqrt{\frac{1}{N_{avg}} \sum_{t=1}^{N_{avg}} \|r_t(i, j) - r_{avg}(i, j)\|^2} \quad (4.12)$$

σ_r is lower than 50 mm for objects with high amplitude, but can stay higher than 300 mm for low amplitude regions, i.e. objects with low reflectivity, or regions in the corner of the field of view. Corner regions suffer from vignetting and from lower illumination amplitude by the LEDs integrated in the SR camera.

Typical noise versus amplitude distributions are analyzed from eight scenes acquired with SR-3000 and SR-3100 cameras. Note that the data distribution is similar for the both cameras. The sets clearly show that σ_r is proportional to the inverse of the amplitude A . As expected σ_r goes to infinity when the amplitude goes to zero. Figures 4.12 and 4.13 show enlargements of the same data. The data distributions are used to fit a functional relation between the range standard deviation σ_r and the amplitude A

Chapter 4. Time-of-flight camera system

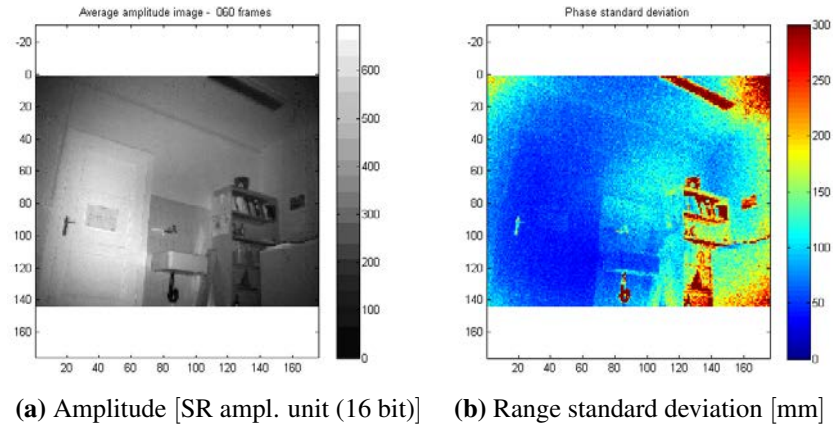


Figure 4.10: Comparison of average amplitude A_{avg} , and standard deviation of range measurement σ_r , for SR-3100 camera. Notice that σ_r varies inversely with A_{avg} .

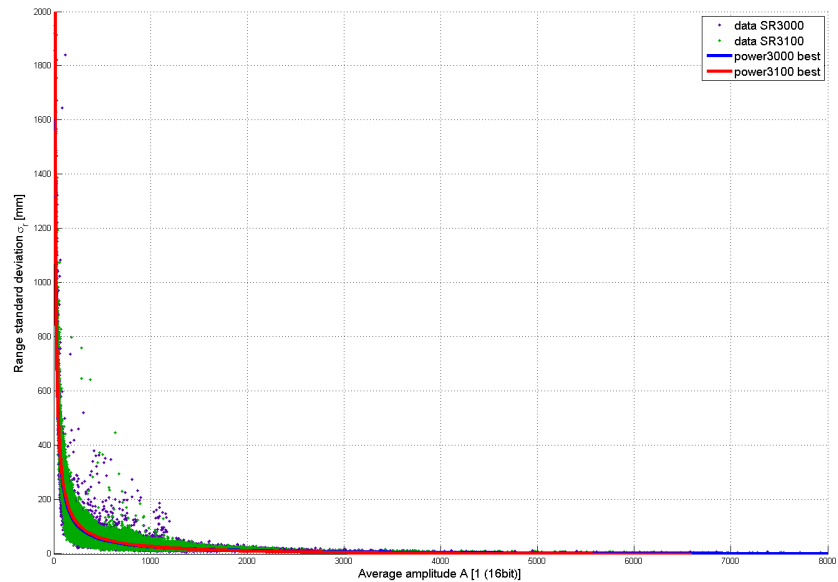


Figure 4.11: Standard deviation of range measurement as a function of average amplitude for SR-3000 (purple) and SR-3100 (green) cameras.

4.3. Noise in SR-3000 cameras

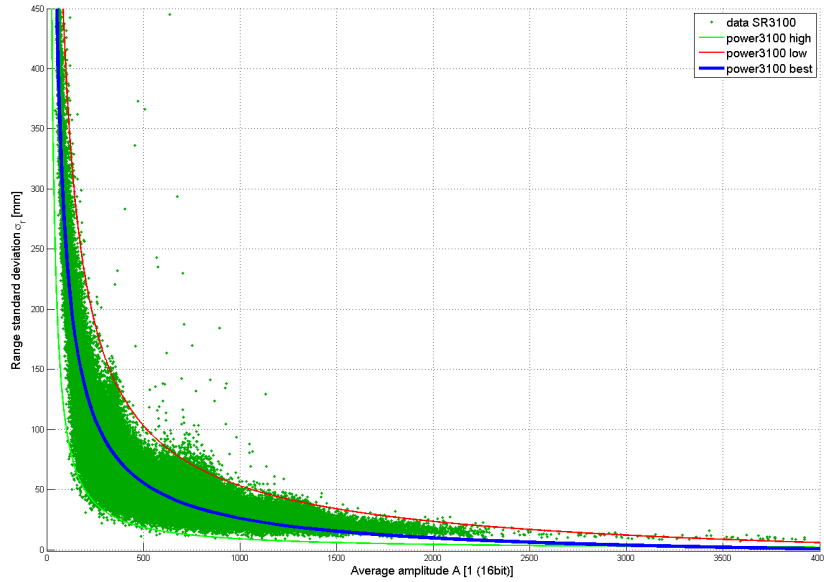


Figure 4.12: Standard deviation of range measurement as a function of average amplitude for and SR-3100 camera. Detail of the $0 < \sigma_r < 350$ region.

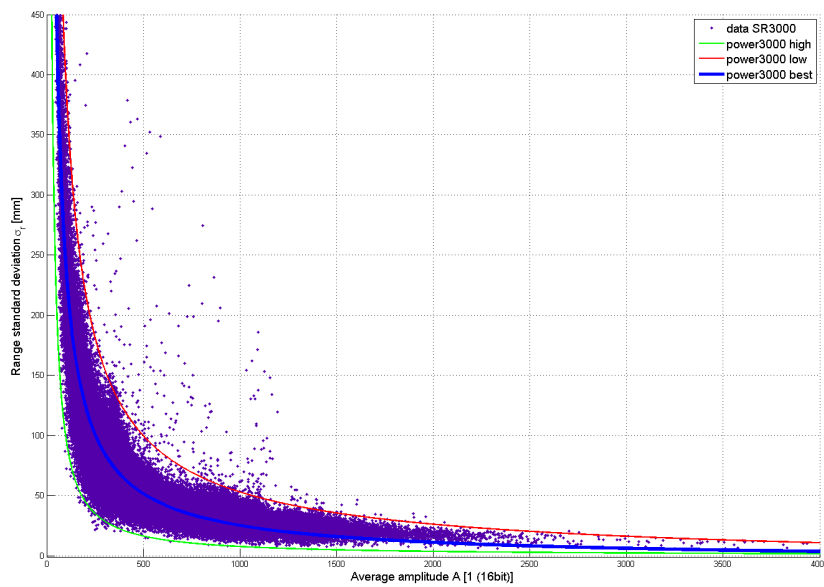


Figure 4.13: Standard deviation of range measurement as a function of average amplitude for SR-3000 camera. Detail of the $0 < \sigma_r < 350$ region.

Chapter 4. Time-of-flight camera system

reported by the SR device. The relationship was modeled as a power law:

$$\sigma_r = \alpha \cdot A^\beta + \gamma \quad (4.13)$$

where α , β and γ are model parameters; This model is relevant when compared to the noise model for CW TOF cameras (eq. 4.5): Values fitted with Matlab's curve fitting toolbox [79] are reported in table 4.2 for the SR-3100, respectively SR-3000 device used for measurements. The fits aren't very precise for large values of A : the value of parameter γ should be 0, according to eq. 4.5. However, γ is indeed small, especially when compared to α . Moreover, the model is mostly useful for low values of A , i.e. when the measurement errors are high. Note that the power β isn't exactly -1 , as could be expected from eq. 4.5. This is related to the fact that the camera's active illumination contributes to the background light noise level B (see eq. 4.6). The experimental data indicates that the camera active illumination isn't the largest contribution to the background noise level, since the factor $\beta \approx -0.9$ is actually close to -1 .

4.3.3 Noise reduction

From a camera user perspective, the options for reducing noise are limited; see 4.5. The modulation frequency f is constrained by the depth of the scene. Moreover, only a small set of discrete integration frequencies are proposed by camera manufacturers. Therefore, increasing the operation frequency is not a convenient solution for noise reduction.

The other options imply to enhance the contrast term $\frac{A}{\sqrt{B}}$. The reflectivity of the objects in the scene can generally not be enhanced; this would require a reflective coating, such as white powder, and is not practical in many situations.

However, noise can be reduced by integrating its temporal fluctuations. Two options for noise reduction are discussed in the next paragraphs.

4.3.3.1 Longer integration time

For all cameras, increasing the integration time allows to decrease the noise, as more photons are collected.

The typical integration time for a SR device is 9.80ms, it can be chosen from 0.20ms to 51.20ms in 0.20ms increments. However, the choice is constrained by frame rate requirements and by the dynamic range of the scene imaged. With an integration time of 51.20ms, the highest sustainable frame rate is approximately 5 fps, while 20 fps is easily attained for the standard integration time.

Figure 4.14 illustrates the pixel saturation effect that can occur when the integration time is not appropriate for the scene dynamic range. In this case, the integration time was set to 51.20ms. But an object close to the camera (here, a hand), causes pixels to

Table 4.2: Range standard deviation as function of amplitude. Power law parameters for best fit.

Model	α	β	γ
SR-3100	12770	-0.8468	-10.75
SR-3000	15270	-0.9008	-5.159

4.4. Deterministic error sources

saturate and report wrong amplitude and depth values. The integration time should be reduced in order to avoid saturation.

4.3.3.2 Frame averaging

Frame averaging can be considered for static scenes. Taking the time average of the TOF signal provides two advantages :

- noise is reduced,
- deviations from the averaged signal provide information on the noise characteristics,
- Saturation problems encountered with long integration times can be avoided by using shorter integration

One key issue when taking averages of TOF data is to consider the amplitude and phase data as part of a **complex signal**. Averaging separately amplitude and phase gives lower quality results for phase, as illustrated in figure 4.15. Since the TOF error σ_r varies with the inverse of amplitude A , taking the complex average naturally ensures that measurements with higher errors contribute less to the averaged phase measurement.

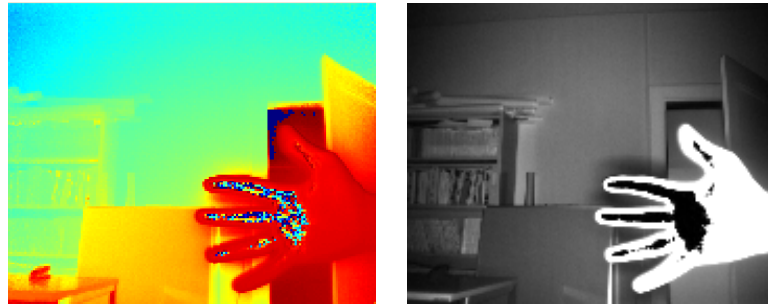
Noise reduction by frame averaging is illustrated in the second column of figure 4.7. Note that frame averaging is only valid for static scenes. In scenes with motion, all time averaging methods cause motion blur.

4.3.3.3 Spatial low-pass filtering

As it is done with conventional 2D cameras, low-pass filters could be used to reduce the noise level. The Swissranger driver implementation [24] includes a 3×3 median filter. However, low-pass filtering techniques cause blur. Since the image resolution is already small, low-pass filtering isn't among the preferred methods for noise reduction in TOF cameras. Moreover, non-linear low-pass filters such as median filters introduce artifacts that cannot be modeled with linear filters. Such filters should therefore be deactivated during data acquisition, in order to avoid extraneous artifacts when linear filters are used in subsequent parts of the image processing pipeline. See chapter 5 for an example of such a linear filtering process, aimed to reduce scattering effects (which are described in the next section).

4.4 Deterministic error sources

The errors sources considered in the previous section were caused by noise, i.e. stochastic processes. In this section, deterministic error sources for SR-300 cameras are discussed; these errors are reproducible. The discussion will focus on two error sources accounting for most of the perturbations observed in SR data: multipath and scattering. Both perturbations are caused by undesired reflections of the TOF light signal. In the case of multipath, the reflections occur in the scene, while for scattering, the reflections occur inside the camera device.



(a) Range map $r(i, j)$ color coded from 0m (red) to 7.5m (blue).

(b) Amplitude map $A(i, j)$ from 0 (black) to 6000 (white).

Figure 4.14: Saturation in SR-3100 data. This dataset was acquired with 51.20 ms integration time. Some pixels in the hand region are saturated and report erroneous amplitude and range values: black pixels in the amplitude image.

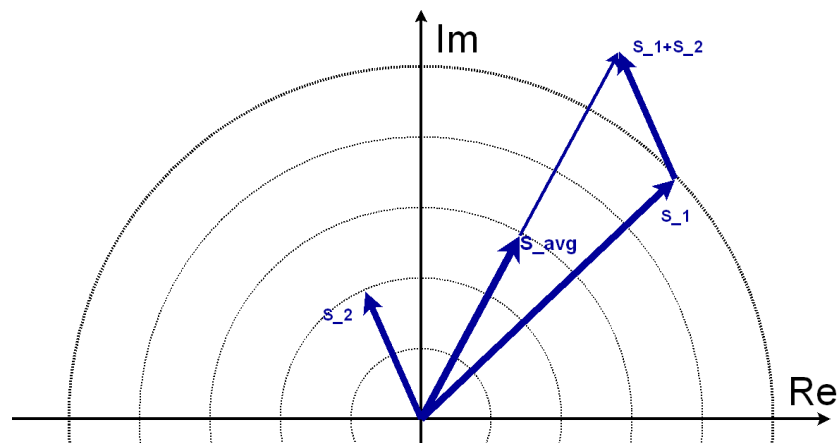


Figure 4.15: Average of two TOF signals S_1 and S_2 : the average phase is different from the average of individual phases.

4.4.1 Multipath

Multipath perturbations are caused by undesired reflections occurring in the scene. Typically, multipath effects are stronger (and therefore easier to observe) near corners. Figure 4.16 provides a simple illustration of multipath for a standard camera.

If the active TOF light signal can take different paths and still be imaged on the same sensor area, the measured range value will be biased towards longer distances, since indirect paths imply longer propagation times. A discussion of multipath in TOF imaging is presented in Guomundsson *et al.* [43]. In their experiments, the authors use a scene with two walls, one of which is fixed while the other one can be removed (see figure 4.17). The results clearly illustrate that multipath affects 3D measurement with the SR-3000 range camera. By fitting plane primitives to the data points, Guomundsson *et al.* [43] found a value of 122° for the dihedral angle between the two walls, while the ground truth value was 90° .

We performed a similar experiment. The scene used was the corner of a room, shown in fig. 4.18, containing 3 planes corresponding to a brick wall, a side wall, and the floor. The scene was imaged simultaneously by a SR-3100 and a SR-3000 TOF camera at different positions. In each dataset, the 3 planes were fitted using a RANSAC method (described in sec. 6.6.2) similar to the method used by Guomundsson *et al.* [43]. The results for the dihedral angle between the different planes are presented in table 4.3. It appears that the fitted planes are all far from orthogonal. But results are very different for both cameras. This data suggests that, while multipath perturbations are reproducible, they are critically dependent on the surfaces present in the scene and their orientations, relative to each other and to the camera. Multipath is therefore very difficult to model.

One general remark to be made is that multipath is strongest for the type of scenes illustrated above, i.e. planes intersections. In that case, a high fraction of the TOF illumination reflected by one plane is intercepted by the second plane. Fortunately, multipath effects are lower far from corners, since the light going through multiple reflections is then often scattered away from the camera. In many practical applications, such as surveillance for example, imprecise measurement of corners are not considered critical.

4.4.2 Scattering

Scattering perturbations are caused by undesired reflections occurring inside the camera device. The effects of scattering are observed in scenes with a large dynamic range. Such perturbations are generally called lens flare in photography; see 4.19.

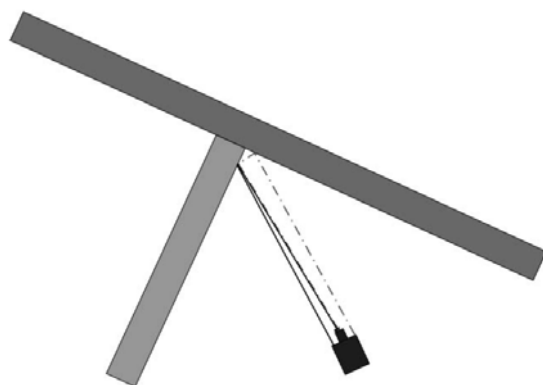
Since the illumination in TOF systems comes from the camera, large dynamic range corresponds to scenes where one object is close to the camera while other objects are

Table 4.3: Measured dihedral angle between fitted planes in scene with large multipath perturbations.

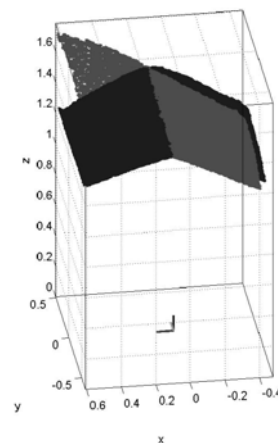
	Ground truth	SR-3100 data	SR-3000 data
Brick wall - Floor	90°	83°	121°
Brick wall - Plain wall	90°	81°	80°
Floor - Plain wall	90°	79°	140°



Figure 4.16: Multipath in standard photography. A flashlight is positioned near a corner. Although the flashlight illuminates only the right wall, the left wall is also visible. Most of the light incident on the right wall is diffused. Some of this light illuminates the left wall, and a fraction of this light is diffused towards the camera.



(a) The multiple reflection experiment setup. One image is taken with the lighter gray wall present and one without. The multiple reflection problem is also illustrated; the correct path is denoted by the black line, an erroneous double reflection by the dashed gray line.



(b) 3D point clouds. Dark points are the results in the presence of the both walls, gray are the one wall measurements.

Figure 4.17: Multiple reflection experiment Guomundsson *et al.* [43]. The reported range data, and the corresponding point cloud are largely affected when the second wall is added in the scene, showing that multipath effects can be important.

4.4. Deterministic error sources

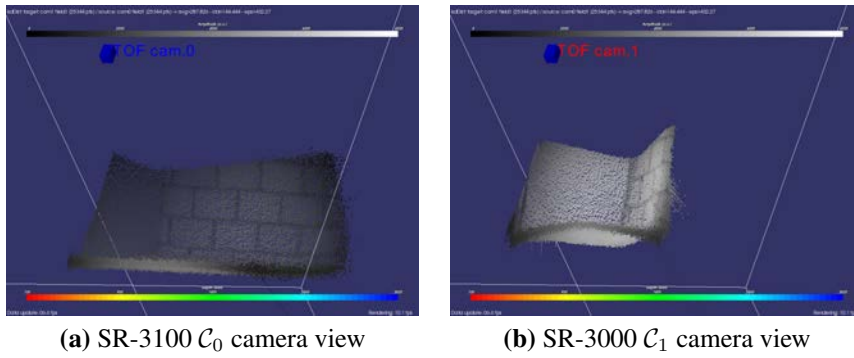


Figure 4.18: Scene for multipath experiment: corner of a room, with one brick wall, side wall and floor, imaged by SR-3100 and SR-3000 cameras (at different positions).

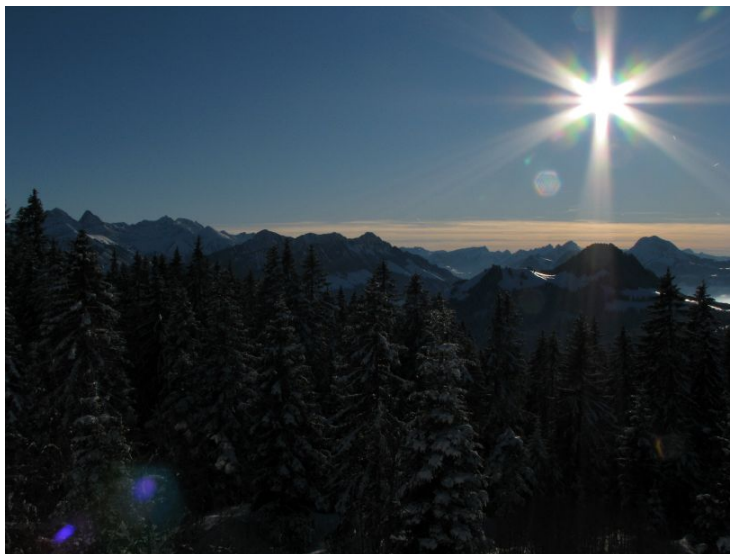


Figure 4.19: Scattering is one of the effects referred to as lens flare in standard photography. The bright light source (sun) causes secondary reflection inside the camera. Notice in particular the hexagonal shape (left and down from the sun). This is an image of an optical aperture inside the camera, confirming that this perturbation is caused by internal reflections.

Chapter 4. Time-of-flight camera system

at a much larger distance. This is often the case when a new object is added in a background scene. Section 4.4.2.1 presents a typical occurrence of scattering in TOF measurements. A qualitative explanation of the range image degradation by scattering is the topic of section 4.4.2.2, while section 4.4.2.3 introduces a metric that can be used to quantify scattering effects.

4.4.2.1 Occurrence of scattering in TOF measurements

The effect of scattering is first illustrated in fig. 4.20, which shows that the measured distance to a fixed wall is affected by the presence of a person. In this experiment, a

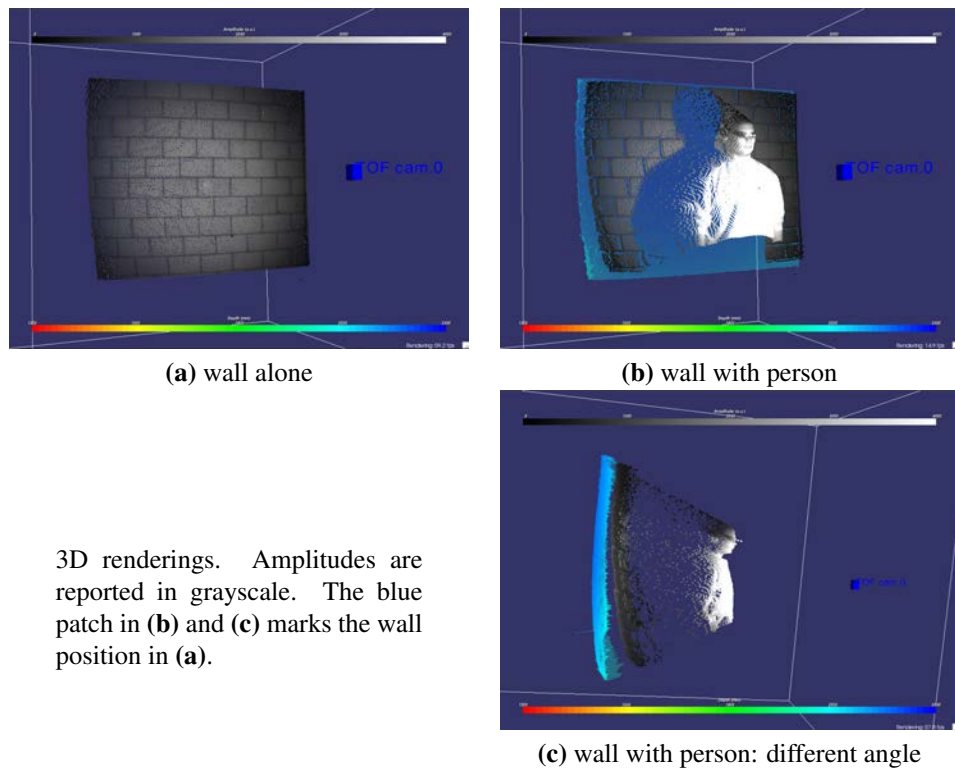


Figure 4.20: Scattering example: The measured distance to the wall is affected by the presence of a person. (b) and (c) compare the wall original position (blue) with its current position.

room is first imaged empty, so that the only object in the range image is the room's wall; fig. 4.20(a). For the second range image measurement, a person is added in the field of view; fig. 4.20(b). The flat wall in fig. 4.20(a) is far from the camera, and thus reflects less light than the foreground person in fig. 4.20(b). It appears clearly that range measurements for the wall region are different when the person is present. This difference is even more striking when the same scene is rendered from a different angle fig. 4.20(c). This type of artifacts will be described throughout this thesis as scattering artifacts, since the present understanding of TOF cameras indicates that those artifacts are produced by light scattering inside the TOF device.

4.4. Deterministic error sources

4.4.2.2 Range measurement degraded by scattering

When an ideal signal $S(\mathbf{i})$ is affected by a parasitic additive contribution $\Delta S(\mathbf{i})$ of different phase, produced for example by light scattering, the phase of the new complex signal $S + \Delta S$ differs from the original phase in a proportion that increases with the ratio $\Delta A/A$ of the respective signal amplitudes. The amplitude A influences therefore the reliability of the depth measurement. The major parameters affecting this quantity are the range r to the imaged object, the angle θ of light incidence on the surface and the object albedo ρ (see figure 2.12):

$$A \propto \frac{\rho \cdot \cos(\theta)}{r^2} \quad (4.14)$$

In a practical situation, the spread of possible values for r , θ and ρ results in a very high amplitude dynamic range that the camera must handle. In particular, when the spread of depths is wide, the $1/r^2$ behavior is critical, since it results in very strong amplitude level differences. Considering the amplitude A_1 of a first object, and amplitude A_2 for a dark object ($0 < \rho_2 \ll 1$) far from the camera, the dynamic range D is given by

$$D = \frac{A_1}{A_2} = \frac{\rho_1 \cdot \cos(\theta_1) r_2^2}{\rho_2 \cdot \cos(\theta_2) r_1^2} \quad (4.15)$$

A high dynamic range situation is illustrated in figure 4.21. The scene imaged is an office room (4.21a). A reference range image $r_{bg}(i, j)$ is taken when the room is empty (4.21b). A second range image $r_t(i, j)$ is acquired in the same conditions, except for the presence of a close person in the field of view (4.21c). Comparison of both range images shows how the depth of the unchanged background is affected by light from the close person. The range difference between the two acquisitions was reported in fig. (4.21d). In this example, the measured range for the unchanged background decreases in presence of the person by values ranging from 100 to 400 mm. Notice that the background regions most affected by scattering are the black tubing (low reflectivity ρ) and the floor region (low angle of incidence θ).

4.4.2.3 Quantitative evaluation of scattering

As shown in the examples of figures 4.20 and 4.21, scattering is best observed when a new foreground object is added to a scene. For qualitative assessment of the scattering defects, a human observer can compare the range measurement in such a case and estimate displacement of background pixels. In order to compare scattering compensation methods or models, it is necessary to use an objective metric assessing scattering severity.

Since the final range accuracy is most important, we propose to use as a metric the root mean square distance to the ideal (i.e. scattering free) range image. In order to measure this distance quantitatively, the image is manually segmented into background and foreground regions (see fig. 4.22). For background regions, the range data r_{avg} measured when no foreground object is present is taken as ground truth, so that the RMS scattering error σ_{scat} is :

$$\sigma_{scat} = \sqrt{\frac{1}{N_{bg}} \sum_{p=1}^{N_{bg}} \|r_t(coord(p)) - r_{avg}(coord(p))\|^2} \quad (4.16)$$

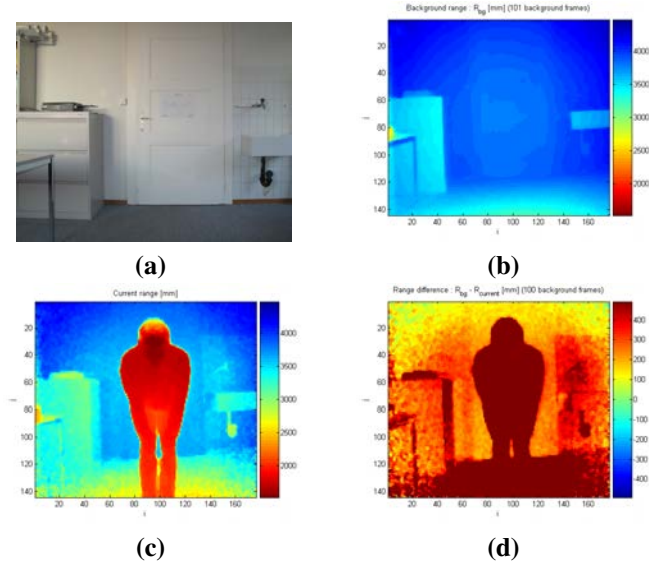


Figure 4.21: Illustration of depth artifacts - (a) Color image. - (b) Background range image $r_{bg}(i, j)$. - (c) Range image with foreground $r_t(i, j)$. - (d) Range image difference.

where $coord(p)$ is the list of pixels (i, j) segmented as belonging to the background, and N_{bg} is the number of background pixels. For foreground regions, the raw data is taken as ground truth. This is equivalent to assuming that scattering is negligible for closer objects. Then, it becomes possible to quantify displacement of foreground pixels introduced by scattering compensation algorithms, which are discussed in the next chapter.

As an illustration of the proposed metric, the RMS background displacement was computed for three scenes (figs 4.22 to 4.24). In each scene, the background was 2000 mm away from the camera. In the first measurement (fig. 4.22), a SR-3000 device was used. A different SR-3000 device was used in the second measurement (fig. 4.23), while a SR-3100 device was used in the third measurement (fig. 4.24). The results for RMS background displacement are reported in table 4.4. The closest object (1.05m) caused the largest background displacement: 305mm RMS. The second experiment with a SR-3000 device shows that scattering is reduced (139mm RMS) when the perturbing object is moved away from the camera (at 1.47m). Finally, the newer SR-3100 device, which includes special coatings reducing parasitic reflections, performs best: the background displacement is only 91mm RMS, for a perturbing object at 1.22m.

4.5 Comparison of TOF error sources

Error caused by unwanted reflections, i.e. multipath and scattering, can be compared to the noise level caused by stochastic error sources. Multipaths effects depend strongly on the scene, and are often larger than the noise level for corner configurations, but almost negligible in most imaging situations. In contrast, even in the situation with the lowest scattering in table 4.4: 91mm RMS background displacement, scattering effects are not negligible, as they are three times larger than the noise level measured in

4.5. Comparison of TOF error sources

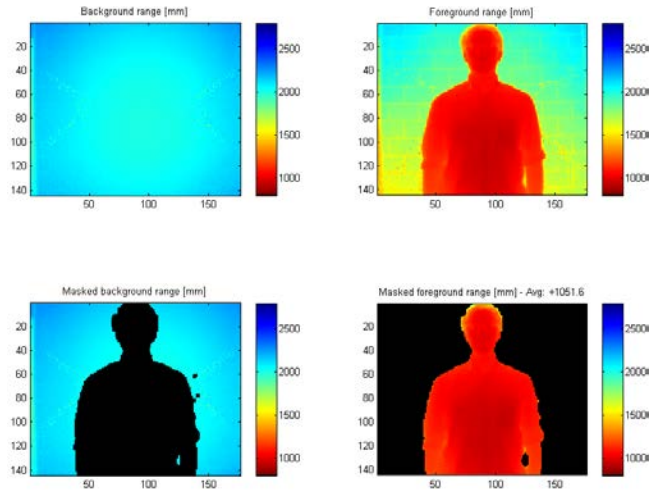


Figure 4.22: Example of image segmentation for scattering measurement. A background image (top left) and a foreground image (top right) are measured. Masks are manually defined in the foreground image to separate background regions (bottom left) and foreground regions (bottom right). Masked data displayed in black. - RMS background displacement: 305 mm

Table 4.4: Average background displacement caused by scattering for 3 simple scenes.

Scene	Camera	Object avg. dist. [m]	Noise level $\sigma_{\text{avg}}[\text{mm}]$	scattering error $\sigma_{\text{scat}} [\text{mm}]$
1 (see fig. 4.22)	SR-3000 dev. 1	1.05	29.2	305
2 (see fig. 4.23)	SR-3000 dev. 2	1.47	n.a.	139
3 (see fig. 4.24)	SR-3100 dev. 0	1.22	29.9	91

Chapter 4. Time-of-flight camera system

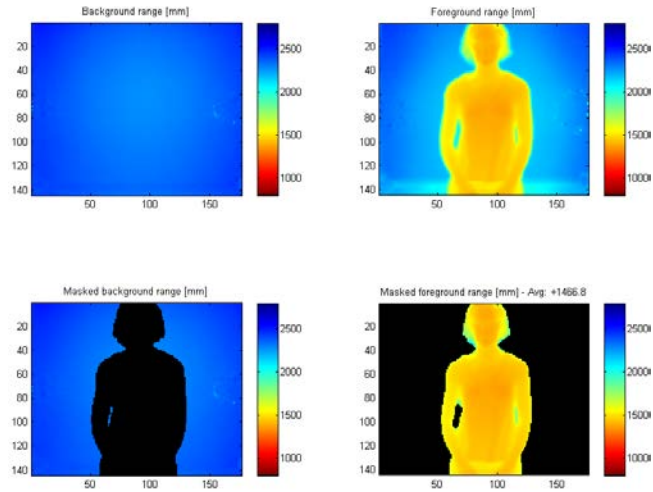


Figure 4.23: Scattering measurement - Camera: SR-3000 dev.2 - Foreground object average distance: 1467 mm - RMS background displacement: 139 mm

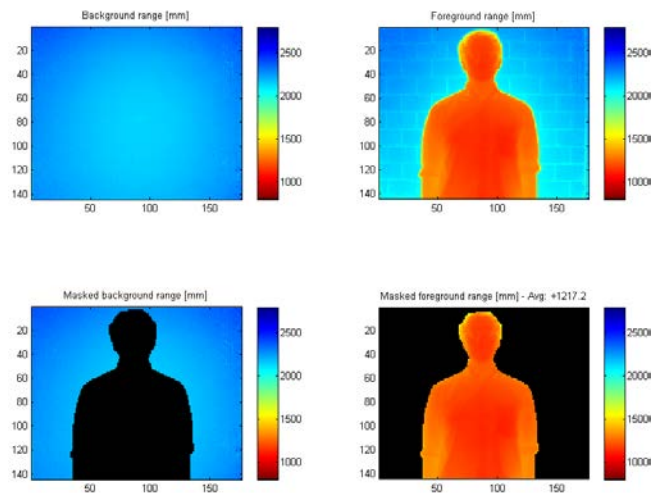


Figure 4.24: Scattering measurement - Camera: SR-3100 dev.0 - Foreground object average distance: 1217 mm - RMS background displacement: 91 mm

this scene (29.9mm). Moreover, while noise effects can be reduced by time averaging, scattering errors are stable over time. In a worst case scenario, scattering errors can be 10 times higher than camera noise. This indicates that scattering is a major source of error in TOF measurements, and should therefore be reduced to enhance the quality of the range images obtained.

4.6 Conclusion

This chapter presented TOF imaging. After a quick comparison of the TOF technologies available today, the discussion focused on TOF cameras using continuous wave emission. The noise properties of such cameras were discussed. The SR-3000 and SR-3100 cameras used in our experimental work were then introduced. A typical example allowed to evaluate the level of stochastic noise observed with these cameras. Experiments on various scenes allowed to observe that the range accuracy varies with the inverse of the measured amplitude. Empirical bounds for the range accuracy based were defined, based on the measured amplitude.

But we have also seen that deterministic error sources can have dramatic effects on the output range image. In particular, it was observed that scattering can cause range measurement errors much larger than the noise variance for a SR-3000 camera. This source of error is a major limitation in current TOF camera systems. This is why chapter 5 is devoted to the problem of scattering compensation.

5

Scattering compensation

This chapter describes an original method for scattering compensation in TOF cameras. First, an overview of the desired properties of scattering compensation algorithm is presented. Then, a rigorous mathematical formulation is introduced to describe the scattering phenomenon. Additional hypotheses are then added to lower the numerical complexity of the task. The proposed algorithm has been implemented in an image acquisition application, and its performance is analyzed.

5.1 Principle of compensation procedure

The reflection mechanisms involved in the production of scattering artifacts are simple to understand intuitively, but are challenging to describe accurately. Since scattering is repeatedly observed in TOF data, we model it as mathematical function h applied on the ideal TOF image data \mathbf{S} , which produces the image \mathbf{S}_{meas} returned by the TOF camera :

$$\mathbf{S}_{meas} = h(\mathbf{S}) \quad (5.1)$$

Based on this model, a scattering compensation algorithm must allow to define a function g which inverts the relationship in equation 5.1.

$$\mathbf{S} = I(\mathbf{S}_{meas}) \quad (5.2)$$

If the function h is invertible the solution will be $I = h^{-1}$. Unfortunately, h is generally unknown, so that there is no guaranty that it will be invertible. Nevertheless, the scattering compensation algorithm should allow to build a good model I of this inverse function.

To justify its use in real world systems, the algorithm must:

- cause minimal perturbation in the range image for low scattering situations,

$$\|I(\mathbf{S}_{meas}) - \mathbf{S}\| \rightarrow 0 \quad \text{if } \|h(\mathbf{S}) - \mathbf{S}\| \rightarrow 0 \quad (5.3)$$

5.1. Principle of compensation procedure

- significantly reduce scattering artifacts for high scattering situations,

$$\|I(\mathbf{S}_{meas}) - \mathbf{S}\| < \|h(\mathbf{S}) - \mathbf{S}\| \quad \text{if } \|h(\mathbf{S}) - \mathbf{S}\| \neq 0 \quad (5.4)$$

- be compatible with camera real-time operation in terms of computational complexity,
- be consistent with the hypothesis of artifacts production through multiple reflections inside the camera device.

5.1.1 Simplified scattering model

Scattering artifacts are caused by multiple internal reflections occurring inside the camera device [81], as illustrated in figure 5.1 by a simplified example with three pixels. This example shows how parasitic reflections from a strong signal source $S(1)$ can come in competition with the direct signal from far objects $S(2)$ and $S(3)$. Figure 5.2 shows the related complex signals $S(1)$, $S(2)$, and $S(3)$ in absence of scattering (5.2a), and then the measured signals $S_{meas}(1)$, $S_{meas}(2)$, and $S_{meas}(3)$ when scattering is present (5.2b). In the worst case scenario, scattering is assumed to create an optical coupling between all pixels. Under the assumption of a linear process, we can describe this coupling through coefficients arranged in a 3x3 matrix h , and the measured signals S_{meas} are then given by the expression :

$$S_{meas}(i) = \sum_{m=1}^3 h(i, m) \cdot S(m) \quad i = 1, 2, 3 \quad (5.5)$$

where the superposition of different signals is computed as an addition in the complex plane.

Since, in TOF imaging, far objects have a low amplitude (eq. 4.14), they are most affected by the scattering phenomenon. This is verified for signals $S(2)$ and $S(3)$ in our example. Moreover, this model explains why depth artifacts are often not necessarily associated with a significant change in the amplitude measured for the affected pixels. In figure 5.2, the phase difference between $S(1)$ and $S(3)$ is such that the perturbation by $S(1)$ on $S(3)$ creates a large range artifact: $\varphi_{meas}(3) \neq \varphi(3)$, but a negligible amplitude change: $A_{meas}(3) \approx A(3)$.

5.1.2 Scattering point spread function

Multiple reflection inside the camera are highly inefficient processes : only a small fraction η of the incident light is reflected by optical surfaces in the imaging system (typ. $\eta < 5\%$). This explains why the amplitude data output from TOF cameras is not critically affected by scattering artifacts. Therefore, we choose to describe scattering as an additive perturbation added to the ideal TOF signal. If $S(i, j)$ is the ideal TOF signal at a sensor pixel (i, j) , then the measured signal $S_{meas}(i, j)$ is the sum of the ideal signal and a small scattering signal $S_{scat}(i, j)$:

$$S_{meas}(i, j) = S(i, j) + S_{scat}(i, j) \quad (5.6)$$

Similarly to the formalism used in optics [103], we can describe the transformation from \mathbf{S} to \mathbf{S}_{meas} as the effect of a non-ideal *point spread function* (PSF) h , as was done in eq. 5.7:

$$S_{meas}(i, j) = \sum_{i'} \sum_{j'} h(i, j, i', j') \cdot S(i', j') \quad (5.7)$$

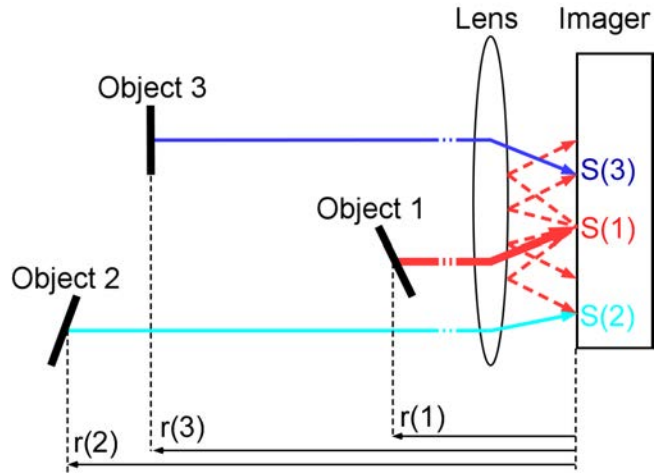


Figure 5.1: Light scattering in TOF camera (after [81]).

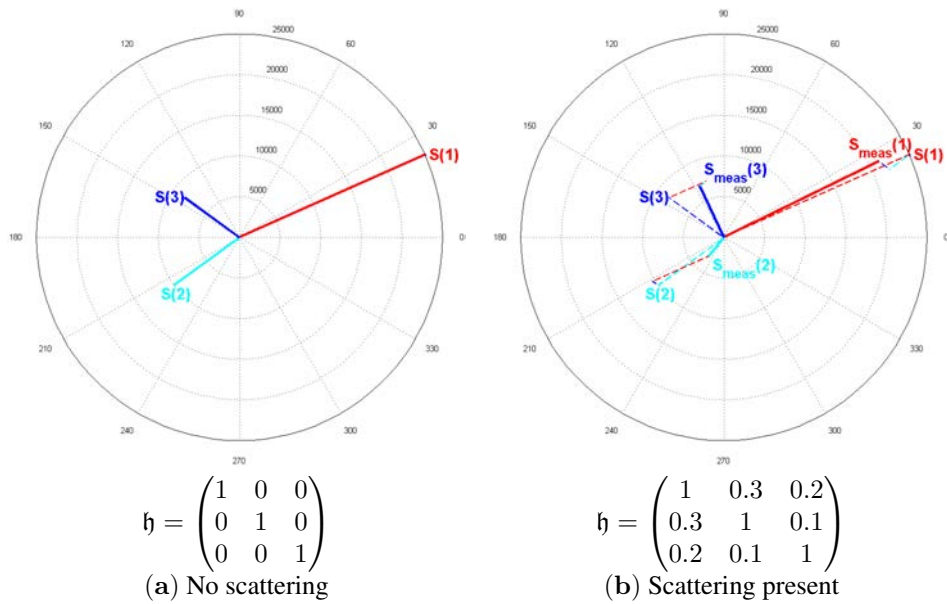


Figure 5.2: Example of linear coupling between three measurement points

5.2. Scattering models

The camera PSF can be decomposed in a standard component \mathbf{h}_0 and a scattering component $\Delta\mathbf{h}$:

$$\mathbf{h} = \mathbf{h}_0 + \Delta\mathbf{h} \quad (5.8)$$

The standard component describes the ideal imaging process ; \mathbf{h}_0 contains only one non-zero coefficient. With δ as the Kronecker symbol, we have:

$$h_0(i, j, i', j') = \delta_{i-i', j-j'} \quad (5.9)$$

In presence of scattering, various coefficients of $\Delta\mathbf{h}$ have a small, but non-zero value, and produce a parasitic scattering image \mathbf{S}_{scat} which adds to the original image. Rewriting equation 5.7 in order to let clearly appear the 'scattering-only' component, allows to see that it is equivalent to eq 5.6.

$$\mathbf{S}_{\text{meas}} = \sum_{i'} \sum_{j'} h(i, j, i', j') \cdot S(i', j') \quad (5.10)$$

$$= \sum_{i'} \sum_{j'} (h_0(i, j, i', j') \cdot S(i', j') + \Delta h(i, j, i', j') \cdot S(i', j')) \quad (5.11)$$

$$= S(i, j) + \sum_{i'} \sum_{j'} \Delta h(i, j, i', j') \cdot S(i', j') \quad (5.12)$$

$$= S(i, j) + S_{\text{scat}}(i, j) \quad (5.13)$$

To describe scattering, we need an estimate $\hat{\Delta\mathbf{h}}$ of the scattering PSF. For our discussion, we consider only the cases where the camera PSF is real (i.e. $h(i, j, i', j') \in \mathbb{R}$). This can be interpreted as making the assumption that the scattering phenomenon does not have an intrinsic phase shift. The low efficiency of scattering processes is modeled by enforcing:

$$\Delta h(i, j, i', j') \ll 1 \quad \forall \quad i, j, i', j' \quad (5.14)$$

5.2 Scattering models

The preceding section introduced a formulation of scattering effects in TOF cameras in terms of a scattering PSF $\Delta\mathbf{h}$. In order to perform scattering compensation, this scattering PSF must be accurately measured or modeled. Although the reflection mechanisms involved in the production of scattering artifacts are simple to understand intuitively (see fig. 5.1), they are not easy to describe quantitatively. Multiple reflections inside the TOF imager could theoretically be modeled using ray-tracing algorithms. However such an approach is not practical since:

- the exact optical design of the TOF camera is known only to the camera manufacturer,
- it is unlikely that all relevant parameters (especially reflectivity of lenses and chips) are precisely measured prior to camera assembly,
- any imperfection in the camera assembly process, such as centering errors, would cause critical changes in a ray-tracing model,
- other mechanisms possibly involved in the production of scattering artifacts would be neglected.

Chapter 5. Scattering compensation

The absence of a detailed quantitative model for scattering production inside the imaging device, does not prevent to attempt scattering compensation. Since scattering is repeatably observed in TOF data, a model based on range image data returned by the TOF device can be constructed.

The following section presents a comparison of different hypotheses used when defining scattering models. The discussion goes from the most general, space variant scattering model to a particular subset of space invariant model particularly well suited for scattering compensation in real-world systems.

5.2.1 Space variant models

In theory, the scattering PSF $\Delta h(i, j, i', j')$ could have arbitrary coefficients for all (i, j, i', j') tuples. However, a practical scattering model can not involve so many independent coefficients : for a 176×144 camera (such as the SR-3000), the number of coefficients would be $176^2 \times 144^2 = 642318336$. This quantity of data is too large to handle with present computer systems.

It is nevertheless possible to define space variant models with less free parameters. In 2008, Kavli *et al.* [64] introduced a scattering compensation method involving PSFs measured at 35 different sensor locations. Such an approach requires a significant calibration effort, since an accurate PSF measurement must be performed at each location. Kavli *et al.* [64] used an highly reflective target, which was placed at 35 discrete positions in the sensor field of view. For each target position, the amplitude variation of the TOF signal was measured; see fig. 5.3. This method requires to use a small integration

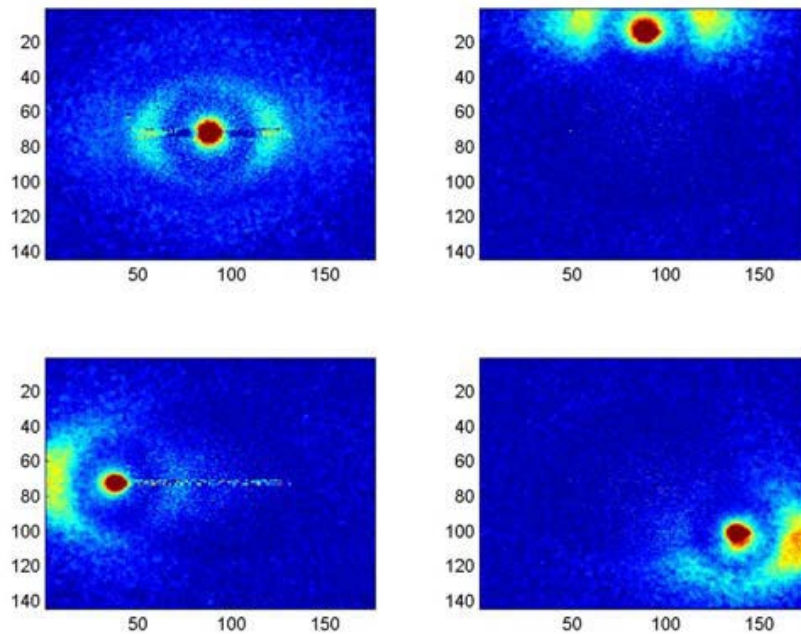


Figure 5.3: Kavli *et al.* [64]: Distribution of the scattered light for different placements of the reflector in the image. - Amplitude is reported from low (blue) to high (red). The bright red dot corresponds to the reflector position. Without reflector, the scene has a uniform low amplitude (blue).

time, in order to avoid saturation for pixels corresponding to the reflector. In [64], this time was set to 512 μs . This short integration time increases the sensitivity to noise in the measured amplitude, so that PSF obtained by this method are usually noisy. Such PSF are therefore difficult to use for scattering compensation, since the noise pattern mixed with the scattering PSF contributes to high frequency noise in the compensation output.

5.2.2 Space invariant models

Space invariant scattering models have many practical advantages over their space variant counterparts when it comes to practical implementation. For space invariance to be satisfied, the scattering PSF must satisfy to

$$\Delta h(i, j, i', j') = \Delta h(i - i', j - j') \quad \forall \quad i, j, i', j' \quad (5.15)$$

The scattering PSF degenerates to a simple 2D matrix $\Delta h(i, j)$. A space invariant model is therefore simpler to describe. Moreover, measuring/learning a single model per camera is easier than measuring space variant models.

Space invariance allows to describe scattering with a convolution operation on the 2D image data $\mathbf{S} = S(i, j)$:

$$\mathbf{S}_{meas} = \mathbf{S} * \mathbf{h} = \mathbf{S} + \underbrace{\mathbf{S} * \Delta \mathbf{h}}_{\mathbf{S}_{scat}} = \mathbf{S} + \mathbf{S}_{scat} \quad (5.16)$$

This scattering model is illustrated in figure 5.4.

5.2.2.1 Handling of image edges in the space invariant model

When a space invariant convolution model is used, applying the model around image edges poses a practical question. The extent of the space invariant model $\Delta h(i, j)$ is such that the convolution sum involves pixel outside the image boundaries. In that case, an assumption must be made on the values of those pixels. Typical assumption include:

- assuming all values are 0 outside the image boundaries,
- replicating edge values over the boundaries,
- mirroring image values over the edge,
- replicating values from the opposite edge (periodical wrapping of the image).

In the case of scattering compensation, the two first options are the most interesting. Setting undefined pixels at value 0 corresponds to assuming that light is perfectly absorbed in the areas around the image sensor. Replication of the edge values implies that areas around the image sensor have the same reflectivity. The edge value in this case is taken as a best estimation of the contribution from those areas to scattering. In this work, the replication option was selected. This choice was motivated by the necessity to discontinuities at image edges when FFT processing is performed.

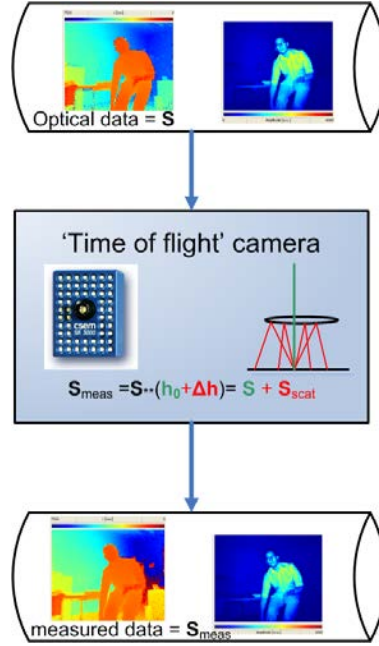


Figure 5.4: Convolution scattering model

5.2.2.2 Sum of gaussians model

In practice, the scattering PSF is very difficult to determine, since the coupling between adjacent pixels is very small. A useful assumption is that this PSF is continuous and smooth over the imager. Since scattering effects have influence on far away pixels, we know that the PSF is wide, probably as wide or wider than the image dimensions. From a computation point of view, this makes a straightforward extensive search prohibitive. A simple model parametrization can be used to facilitate the search.

In this work, we chose to parametrize the scattering impulse response as a weighted sum of gaussians:

$$\hat{\Delta h}(i, j) = \sum_{k=1}^G w(k) \cdot h_h(i, k) \cdot h_v(j, k) \quad (5.17)$$

where :

- h_h is a 1D horizontal gaussian kernel ($\in \mathbb{R}$): $h_h(i, k) = \frac{1}{\sqrt{2\pi}\sigma_h(k)} e^{-\frac{i^2}{2\sigma_h^2(k)}}$
- h_v is a 1D vertical gaussian kernel ($\in \mathbb{R}$): $h_v(j, k) = \frac{1}{\sqrt{2\pi}\sigma_v(k)} e^{-\frac{j^2}{2\sigma_v^2(k)}}$
- $w(k)$ is a scalar ($\in \mathbb{R}$) weight.

The weighted sum of gaussians ensures the smoothness of the impulse response, and also ensures a very sparse representation, since the parameters are : the number of gaussians in the sum G , and, for each gaussian, a standard deviation in the horizontal and vertical direction σ_h and σ_v , along with a weight w .

Section 5.6 introduces a method allowing to find a sum-of-gaussians scattering model that produces best results for a given set of training data.

5.3 Convolution based compensation

Section 5.2.2 introduced a representation of scattering processes through a convolution applied on the ideal image. According to eq. 5.16, subtracting the scattering image S_{scat} to the measured image S_{meas} allows to compensate for scattering. However, neither S nor Δh are available. Therefore, the best scattering image estimate S'_{scat} which can be employed for compensation is given by the expression:

$$S'_{scat} = S_{meas} * \hat{\Delta h} \quad (5.18)$$

where $\hat{\Delta h}$ is a *best-performing* estimation of the scattering PSF Δh . An overview of this scattering compensation approach is given in figure 5.5. Although this approach

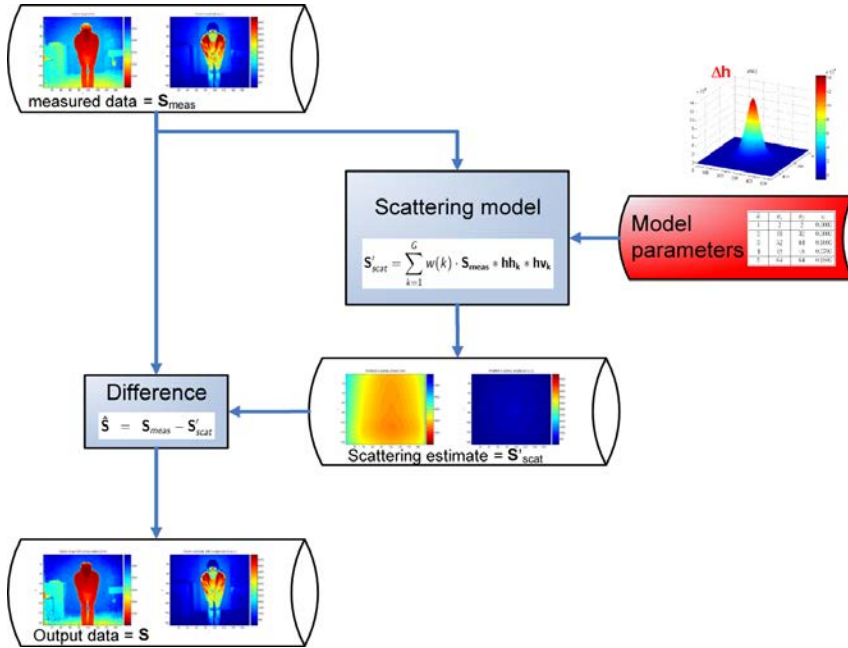


Figure 5.5: Schematic model of scattering compensation through convolution

is straightforward to implement, and can provide satisfying range image outputs, it presents some drawbacks. First, since the scattering estimation uses a raw TOF sensor image as input, no complete cancellation can occur, even if a perfect scattering PSF Δh were known, as second order scattering terms will always be present. However, the inefficiency of scattering processes ensures that this error stays small :

$$\begin{aligned} \hat{S} &= S_{meas} - S'_{scat} \\ &= (S + S * \Delta h) - (S + S * \Delta h) * \hat{\Delta h} \\ &= S + \underbrace{S * \Delta h - S * \hat{\Delta h}}_{\text{small if } \Delta h \approx \hat{\Delta h}} - \underbrace{S * \Delta h * \hat{\Delta h}}_{\text{small if } (\sum_i \sum_j \Delta h(i,j))^2 \ll 1} \end{aligned}$$

The second and most problematic drawback of scattering compensation through convolution is the long processing time required to actually compute the convolution results. This issue is extremely significant since scattering PSFs for TOF devices are as large,

Chapter 5. Scattering compensation

or sometimes larger than the TOF image. This consideration motivated the investigation of scattering compensation in Fourier space, since the Fourier transform allows to replace convolutions with much faster multiplications of the Fourier signals.

5.4 Compensation by Fourier division

In section 5.2.2, scattering was expressed as a convolution of the ideal signal \mathbf{S} with the scattering PSF \mathbf{h} . Practical experiments show that the scattering PSF can have a fairly large extent. For large extent PSFs, computing convolutions is a lengthy process. This process can be speeded up when remembering that a convolution in real space is equivalent to a multiplication in the Fourier space [103].

Equation 5.16 can be transposed in the Fourier domain ($\tilde{\mathbf{S}}(u, v) = \mathcal{F}\{\mathbf{S}(i, j)\}$, $\tilde{\mathbf{H}}(u, v) = \mathcal{F}\{\mathbf{h}(i, j)\}, \dots$)

$$\tilde{\mathbf{S}}_{meas} = \tilde{\mathbf{S}} \cdot \tilde{\mathbf{H}} \quad (5.19)$$

If a space invariant model for \mathbf{h} is known, the desired signal \mathbf{S} can be retrieved through a simple division in the Fourier domain :

$$\tilde{\mathbf{S}} = \tilde{\mathbf{S}}_{meas} \cdot \frac{1}{\tilde{\mathbf{H}}} \quad \mathbf{S} = \mathcal{F}^{-1} \left\{ \tilde{\mathbf{S}}_{meas} \cdot \frac{1}{\tilde{\mathbf{H}}} \right\} \quad (5.20)$$

This process is illustrated in figure 5.6. Since fast implementation of 2D FFT are

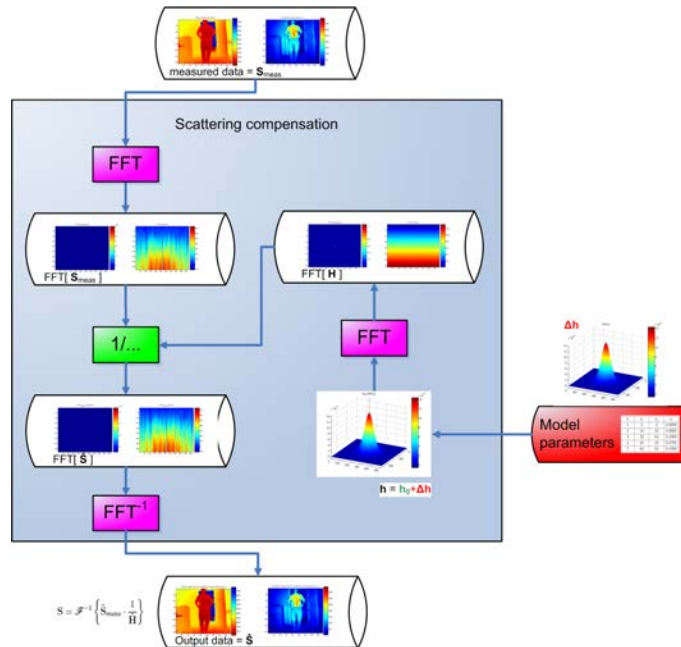


Figure 5.6: Schematic model of scattering compensation in Fourier domain

available [37], the computation time for scattering compensation is greatly reduced [84].

5.5. Complexity comparison of scattering compensation techniques

5.4.1 Windowing function for FFT processing

FFT processing can only be applied to a periodic signal. Therefore, each TOF image must be *windowed*, in order to be compatible with periodicity, prior to FFT processing. Without windowing, the filtering results would be distorted due to discontinuities at the image edges. In the specific case of scattering compensation, windowing isn't performed directly on the TOF image. The edge pixels of the $M \times N$ TOF image are replicated in order to obtain an image buffer of size $2M \times 2N$, whose center part is the original TOF image. The size of this extended buffer matches the size of the space invariant scattering PSF used. The window function is then applied on the extended image buffer: the window function is unity for the image center pixels, and falls with gaussian tails for pixels whose distance to an edge is small ($< N/2$). Using a flat window over the original TOF image region allows to avoid using inverse windowing filters when reading the results of the inverse FFT.

5.5 Complexity comparison of scattering compensation techniques

In order to be used in a real-time acquisition system, scattering compensation should not require long computation times. It is therefore useful to compare the computational load for scattering compensation. The following comparison involves 5 different approaches which can be used when considering space invariant scattering models:

- Image domain filtering with scattering PSF expressed as discrete function $\Delta h(i, j)$
- Image domain filtering with scattering PSF expressed as a sum of separable gaussians ($\Delta \mathbf{h} = \sum \mathbf{g}$)
- Image domain filtering with inverse filter expressed as discrete function $\Delta I(i, j)$
- Image domain filtering with inverse filter expressed as a sum of separable gaussians ($\Delta \mathbf{I} = \sum \mathbf{g}$)
- Fourier domain filtering with scattering PSF $\Delta \mathbf{h}$

In general, if the TOF image is $M \times N$ (i.e. M columns and N rows), the scattering PSF extent can be as large as $2M \times 2N$, in order to specify all possible pixel couplings.

Table 5.1 shows a comparison of the computational complexity involved for scattering compensation. The image space filtering approach is practical only if the inverse filter is expressed as a sum of separable gaussians. Nevertheless, the complexity is higher than in Fourier transform processing. Finally, the Fourier domain filtering is preferred since it involves the lowest complexity of computation, independently of the scattering descriptor used. Among the descriptors, the PSF expressed as sum of gaussians ($\Delta \mathbf{h} = \sum \mathbf{g}$) is preferred since it directly describes the phenomenological scattering model of fig. 5.1, and since it is specified by a small set of parameters Θ that may be easily optimized.

Table 5.2 compares the processing time for 4 different scattering compensation implementations, running on the same computer, and producing identical results.

- Image domain filtering with discrete function $\Delta I(i, j)$
- Image domain filtering with inverse filter expressed as a sum of separable gaussians ($\Delta \mathbf{I} = \sum \mathbf{g}$)

Chapter 5. Scattering compensation

Table 5.1: Comparison of deconvolution complexity for different scattering models

		Image domain filtering	Fourier domain filtering
Scattering descriptor	<ul style="list-style-type: none"> • $\Delta h(i, j)$ • $\Delta \mathbf{h} = \sum \mathbf{g}$ • $\Delta I(i, j)$ 	Complexity $\mathcal{O}(M^2 N^2)$	Complexity $\mathcal{O}(MN(\log(MN)))$
	<ul style="list-style-type: none"> • $\Delta \mathbf{I} = \sum \mathbf{g}$ 	Complexity $\mathcal{O}(MN(M + N))$	

- Image domain filtering with inverse filter $\Delta \mathbf{I}$ expressed as a sum of separable gaussians, realized with optimized image processing library (IPL 2.5)[55]
- Fourier domain filtering with PSF $\Delta \mathbf{h}$, realized with FFTW library Frigo & Johnson [38].

Table 5.2: Average processing time for different scattering descriptors.

	Image domain filtering			Fourier domain filtering
	Discrete function $\Delta I(i, j)$	Sum of gaussians $\Delta \mathbf{I} = \sum \mathbf{g}$	Optimized sum of gaussian $\Delta \mathbf{I} = \sum \mathbf{g}$	$\Delta \mathbf{h} = \sum \mathbf{g}$
Complexity	$\mathcal{O}(M^2 N^2)$	$\mathcal{O}(MN(M + N))$		$\mathcal{O}(MN \log(MN))$
Average proc. time per frame	46.0 s	0.460 s	0.085 s	0.033 s

The comparison clearly verifies that Fourier domain processing allows for faster scattering compensation: only 33ms are required for scattering compensation using the FFT method, while an optimized convolution method requires 85ms for separable gaussian kernels. For non-separable gaussians kernels, the computation time with the convolution approach reaches 46s, which would be totally incompatible with real time operation. Therefore, the FFT approach is preferred for all scattering compensation operations.

5.6 Optimization of scattering model parameters

When a segmentation of scenes in which scattering occurs is available (see sec. 4.4.2.3), it becomes possible to evaluate the efficiency of a given scattering model $\Delta \mathbf{h}$ by measuring the RMS background displacement when this model is used for compensation. Trying to obtain the best scattering PSF by extensive search over all possible PSFs is

5.7. Compensation results

not feasible. A simple model parametrization is needed to facilitate the search.

5.6.1 Family of models tested

In this work, we chose to parametrize the scattering PSF $\Delta \mathbf{h}$ as a weighted sum of gaussians (see section 5.2.2.2). The weighted sum of gaussians ensures the smoothness of the impulse response, and also ensures a very sparse representation, since the parameters are : the number of gaussians in the sum G , and, for each gaussian, a standard deviation in the horizontal and vertical direction σ_h and σ_v , along with a weight w . Although this has not been attempted yet, this representation could enable using genetic algorithms to find an optimum scattering impulse response.

A cruder but very effective approach is to keep the number of gaussians as well as their standard deviations constant, and to allow the weights to be modified. Figure 5.7 illustrates that using 3 gaussians already allows to define a wide variety of impulse responses. The empirically defined standard deviations are reported in table 5.3.

5.6.2 Optimization experiments

In optimization experiments, the error metrics used are the RMS background and foreground displacements as introduced in section 4.4.2.3. Figure 5.8 illustrates how those quantities are computed for a particular set of parameters.

Using a family of PSFs generated by a sum of three gaussians, with 12 possible weights for each gaussian in the sum, creates 1728 PSFs for which the scattering compensation performance can be evaluated. Figure 5.9 presents the results of such an experiment, for the three scenes illustrated in figures 4.22 to 4.24. A zoom on the data (fig. 5.10) allows to verify that the set of PSFs studied spans the PSF space correctly: some PSF bring almost no improvement when compared to the uncompensated situation (bottom of the curves in fig. 5.10), many PSF further degrade the RMS background distance (tail of the curve on the right hand side of the minimum), while a few PSF do actually bring an improvement in RMS background distance, at the cost of an augmented RMS foreground distance (tail of the curve on the *left* hand side of the minimum).

5.7 Compensation results

The previous sections presented the steps required to obtain appropriate scattering models for compensation. Here, we illustrate the validity of the models obtained on real examples.

Table 5.3: Geometric parameters of gaussian kernels used in optimization experiment.

k	σ_h	σ_v
1	32	64
2	48	48
3	64	64

Chapter 5. Scattering compensation

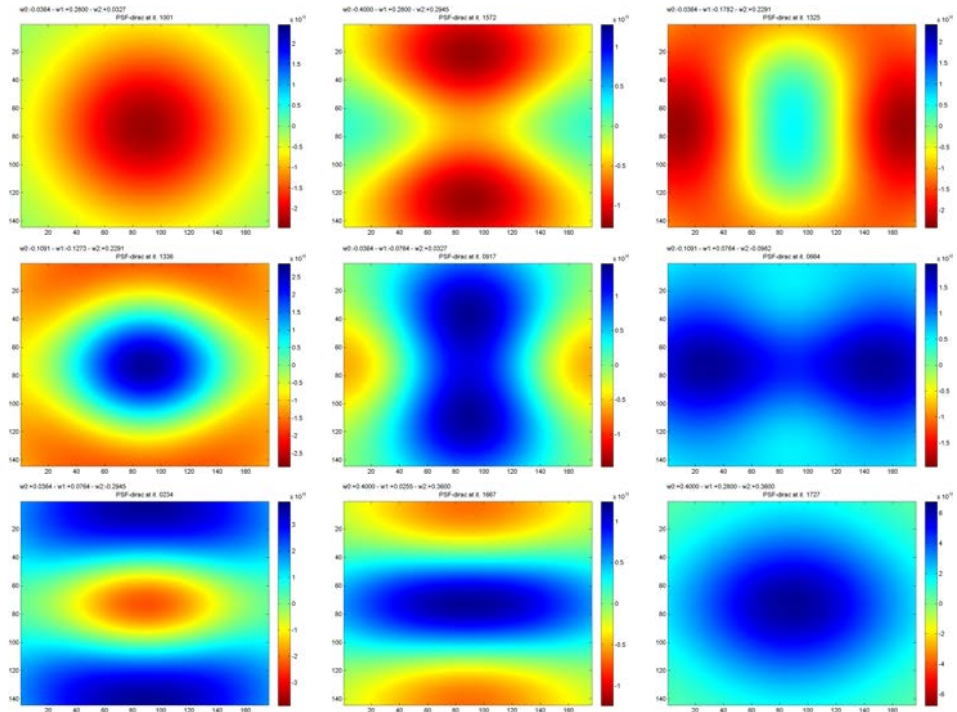


Figure 5.7: Examples of scattering compensation point spread functions generated by weighted sums of 3 gaussian kernels. The dimensions of the gaussian kernels are fixed, only weights are varied.

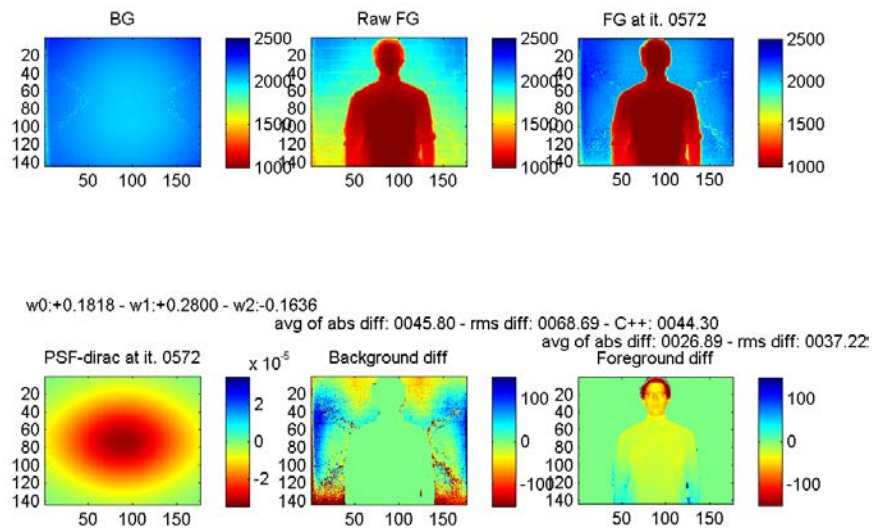


Figure 5.8: Errors against background and foreground when using scattering compensation on scene from example in fig 4.20

5.7. Compensation results

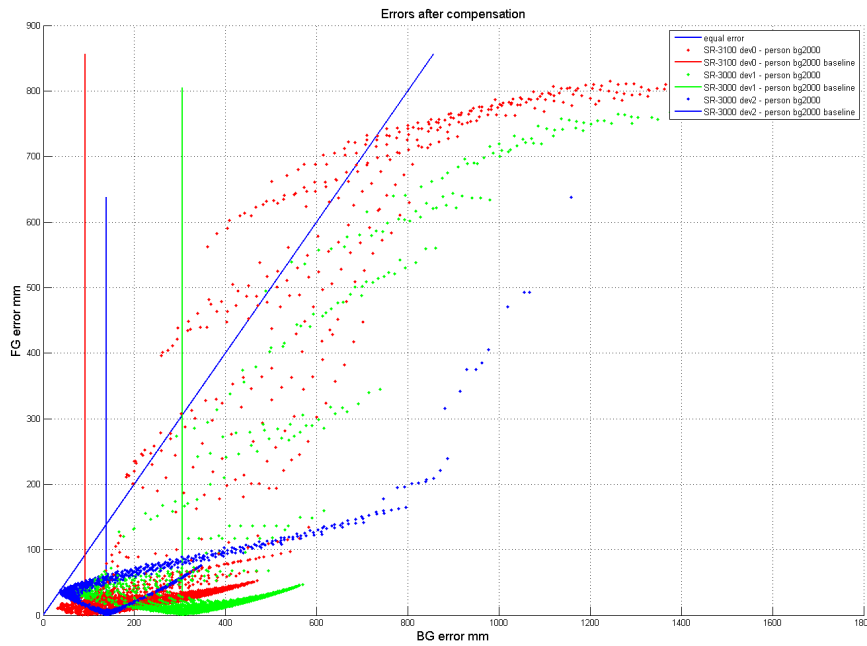


Figure 5.9: Comparison of RMS background and foreground displacements - 1728 PSFs were evaluated, for the 3 scenes illustrated in fig. 4.22 to 4.24. Plain lines indicate the baseline for background error (no compensation).

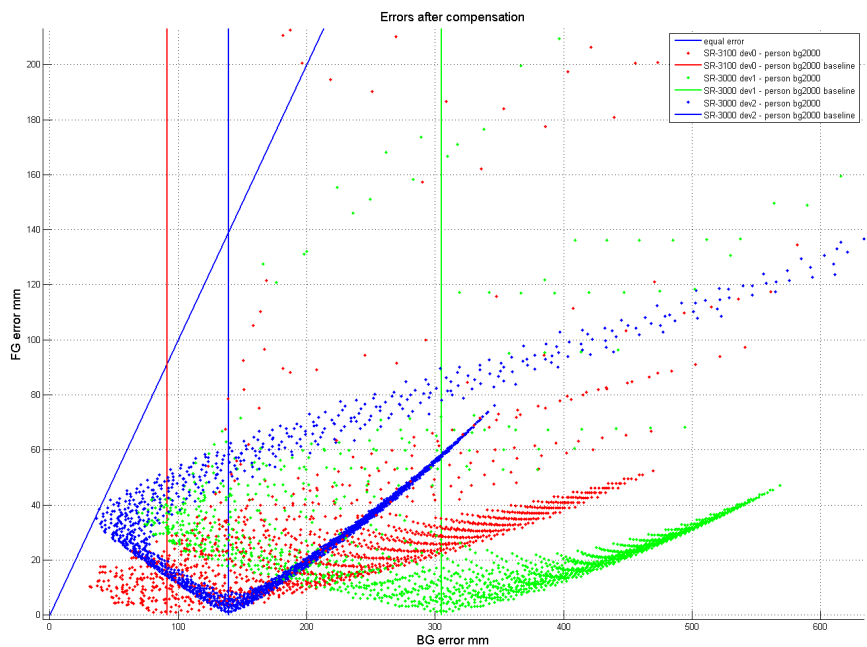


Figure 5.10: Comparison of RMS background and foreground displacements - Detail - Plain lines indicate the baseline for background error (no compensation).

Chapter 5. Scattering compensation

Figure 5.11 presents scattering compensation results in form of points clouds compared without and with compensation, for the example test scene of figure 4.22. Results for the other example scenes of section 4.4.2.3 are summarized in table 5.4. In all cases, the RMS background displacement caused by scattering is greatly reduced. It then becomes similar to noise errors observed in SR-TOF data. Note that results are best for the SR-3100 camera, which is the camera with lowest scattering.

A disappointing note is the difficulty to find a universal scattering PSF for a particular camera. The PSF used are often valid for a specific scene configuration (distance from camera to background, distance from camera to scattering object, etc), and for small variations close to this configuration. But often a different model is required if the scene is totally different.

5.8 Limitations of scattering PSF model

The PSF model used for scattering has limitations which are important to consider when analyzing its performance. The first limitation is the assumption of linearity of scattering coupling; see eq. 5.7. Linearity can be interpreted as taking only the first term in a Taylor series expansion of the measured signal :

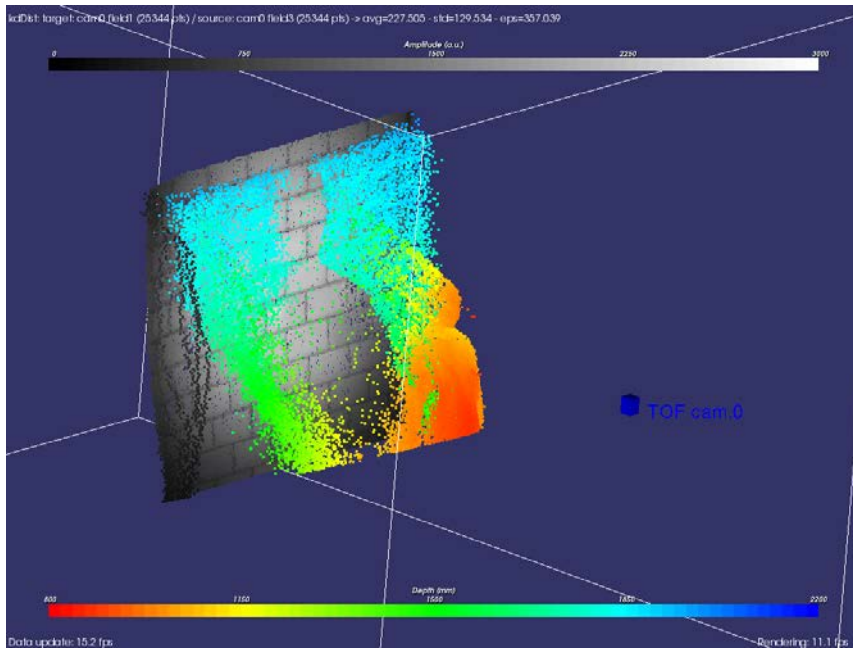
$$\begin{aligned}
 S_{meas}(i, j) = & S(i, j) + \sum_{i'} \sum_{j'} \Delta h(i, j, i', j') \cdot S(i', j') \dots \\
 & + \sum_{i'} \sum_{j'} \Delta h_{(2)}(i, j, i', j') \cdot S(i', j') \cdot \|S(i', j')\| \\
 & + \sum_{i'} \sum_{j'} \Delta h_{(3)}(i, j, i', j') \cdot S(i', j') \cdot \|S(i', j')\|^2 \quad (5.21) \\
 & + \dots \\
 & + \sum_{i'} \sum_{j'} \Delta h_{(n)}(i, j, i', j') \cdot S(i', j') \cdot \|S(i', j')\|^{n-1}
 \end{aligned}$$

The accuracy of scattering compensation algorithms could be increased by taking into account higher order terms. In particular, the quadratic term (associated with PSF $\Delta h_{(2)}(i, j, i', j')$) may be better fitted to the physical scattering phenomenon, if undesired reflection occurring inside the camera are quadratic with light intensity. This approach wasn't tested in this work. But the observation that linear scattering PSF models must be adapted when the scene changes seems to indicate that higher order terms should be considered.

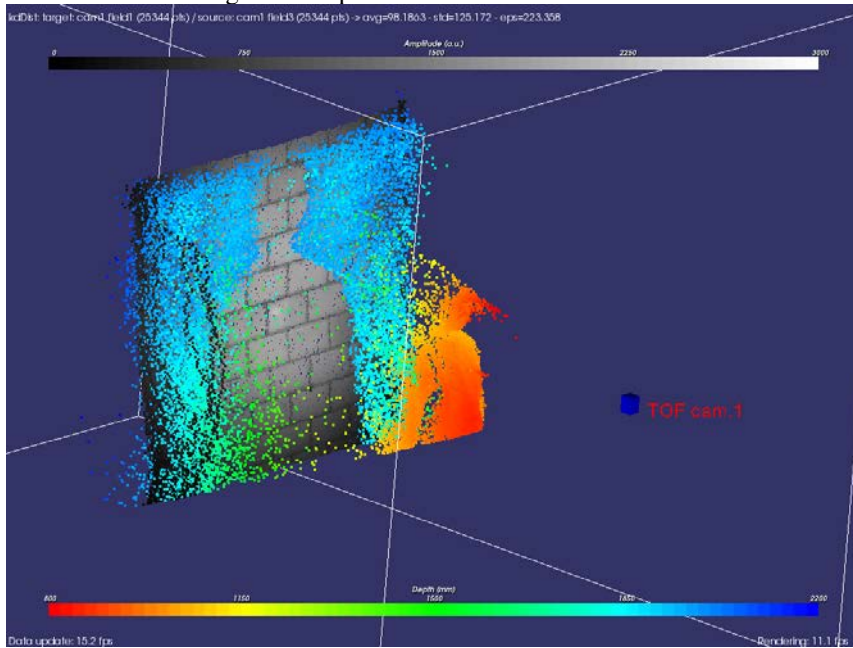
Table 5.4: Scattering compensation results for 3 simple scenes.

Scene	Camera	Noise level σ_{avg} [mm]	σ_{scat} (raw data) [mm]	σ_{scat} after compensation [mm]
1 (see fig. 4.22)	SR-3000 dev. 1	29.9	305	68.7
2 (see fig. 4.23)	SR-3000 dev. 2	n.a.	139	36.3
3 (see fig. 4.24)	SR-3100 dev. 0	29.2	91	31.5

5.8. Limitations of scattering PSF model



(a) Background (grayscale) and current (color) point clouds.
RMS background displacement : 305mm.



(b) Background (grayscale) and current (color) point clouds
with scattering compensation activated. RMS background
displacement : 68.7mm.

Figure 5.11: Scattering compensation results for SR-3000 example scene: single person in front of a wall. The displacement of background pixels is mostly compensated.

Chapter 5. Scattering compensation

But the most fundamental limitation of the scattering compensation methods presented here lies in the assumption that scattering is invertible. In practice, many reflected light rays fall outside the area of the camera sensor, so that this information is lost. Scattering is therefore not invertible, and a perfect compensation is impossible.

5.9 Conclusion

This chapter addressed the compensation of scattering for TOF cameras. First, basic requirements for the compensation algorithm were expressed. Then, a mathematical formulation of the scattering phenomenon inside TOF device was presented, based on a scattering point spread function (PSF). Although the scattering PSF can generally be space variant, we have assumed space invariance for this function to greatly simplify the complexity of the scattering compensation task.

Under the space invariance hypothesis, scattering compensation can be performed either by convolution or by division in the Fourier space. It was shown that Fourier space processing allows faster computation. A testing method for candidate scattering PSFs constructed as sum of separable gaussians was developed. This testing allows to select a best performing PSF when the segmentation between foreground and background data is known. In an example case, the RMS background displacement could be reduced from 305mm to 68.7mm. Moreover, the good performance of scattering PSFs based on sum of gaussians PSFs was illustrated on various scenes containing scattering.

A limitation of the current implementation is that the validity domain of a given scattering PSF model seems limited to scenes similar to the one on which the model was trained. A possible area for improvement would be to include quadratic coupling terms in the scattering model.

6

Registration of noisy range images

Most 3D vision devices are inherently affected by occlusion. In order to acquire complete data sets in presence of occlusion, multiple views must be combined. Moreover, the field of view of a given device is limited. When large scenes must be observed, using multiple views allows to synthetically extend the field of view.

Multi-view range imaging systems can bring improvements in many applications, but such systems require to:

- register the different views : each range image is expressed in its own reference frame; all images must be expressed in a common reference frame.
- integrate the different views : the range image data must be merged.

In this chapter, we will discuss the registration of point cloud data obtained from noisy range images. Considering point clouds allows to use a simple integration approach : registered point clouds can be summed into a larger point cloud. Advanced integration schemes, allowing for example to reduce noise by exploiting the redundancy in the registered data [108] are not considered in this thesis.

Our contribution will rather be focused on registration of point clouds in data sets with limited overlap. However, the large noise observed in range images will affect the selection of the registration method. Since individual points are often not reliable enough, robust alignment features must be built by using as many points as possible. Therefore, this chapter discusses five registration methods for two point clouds, and compares them in terms of complexity, but also on fitness to noisy TOF datasets. The extension to systems with more than two views can be done by sequential pairwise registration. Error balancing techniques for multi-view systems [7] are outside the scope of this thesis.

6.1 3D point clouds registration

Registration of two point clouds \mathcal{P}_0 and \mathcal{P}_1 acquired by two devices $\mathcal{C}_0, \mathcal{C}_1$ at different positions, separated by a rigid-body transformation T , requires to eliminate the 7

Chapter 6. Registration of noisy range images

degrees of freedom (DoF) between the two views. These seven DoF correspond to :

- an arbitrary scaling (1 DoF),
- an arbitrary 3D rotation (3 DoF),
- an arbitrary 3D translation (3 DoF).

For the following discussion, the scaling degree of freedom will not be considered. This is reasonable since the range imagers studied in this thesis provide *absolute valued* point clouds, i.e. the point positions are already scaled against the reference length unit, the meter. As a convention, we use \mathcal{C}_0 's reference frame as the common reference frame. The registration problem is then to determine the rigid-body transformation $T_{\mathcal{C}_0, \mathcal{C}_1}$ which allows to express \mathcal{P}_1 in \mathcal{C}_0 's reference frame. We assume that no *a-priori* knowledge of the cameras positions is available. The registration procedure shall be based only on the data returned by the range cameras. Note that this data is usually noisy, so robustness to noise in the range image inputs is a desired property for the registration algorithms studied.

6.2 Motivation

This chapter specifically addresses the problem of registration of a pair of point clouds in situations with limited overlap. Two main objectives are defined: field of view extension, and occlusion removal.

The discussion of field of view extension is motivated by considering the possible application of TOF cameras in surveillance systems: often a single camera \mathcal{C}_0 does not fully cover the region to be watched. In this case, adding more cameras $\mathcal{C}_1 \dots \mathcal{C}_{N_{cam}}$ is beneficial, as it allows field of view extension, as illustrated in figure 6.1. Since the cost of a TOF camera is high, it is generally desired to cover the largest possible field of view with a small number of devices. This in turn requires to have a small overlap between individual devices' fields of view. In the example of figure 6.1, a person walks in front of two TOF cameras. The reference camera \mathcal{C}_0 is aimed at the upper body, while the second camera \mathcal{C}_1 points toward the legs. The resulting point cloud $\mathcal{P}_0 + \mathcal{P}_1$ clearly shows the full person walking.

Occlusion removal can also be addressed by multi-camera systems, as illustrated in fig. 6.2. Using a single TOF camera results in large unmonitored areas on the wall behind a person close to the camera (fig. 6.2b). The occlusion is removed when a second TOF camera is added (fig. 6.2d), allowing to confirm that a single human is present in the cameras field of view. A typical application of occlusion resolution is human monitoring applications, where the number of humans present in a defined area must be reliably determined.

6.3 Registration based on intensity images : bundle adjustment

Range maps produced by TOF devices are always produced with associated amplitude images. A first attempt at TOF data registration could use standard techniques developed for conventional intensity cameras.

A classical approach based on a planar target was proposed by Tsai [118] and later

6.3. Registration based on intensity images : bundle adjustment

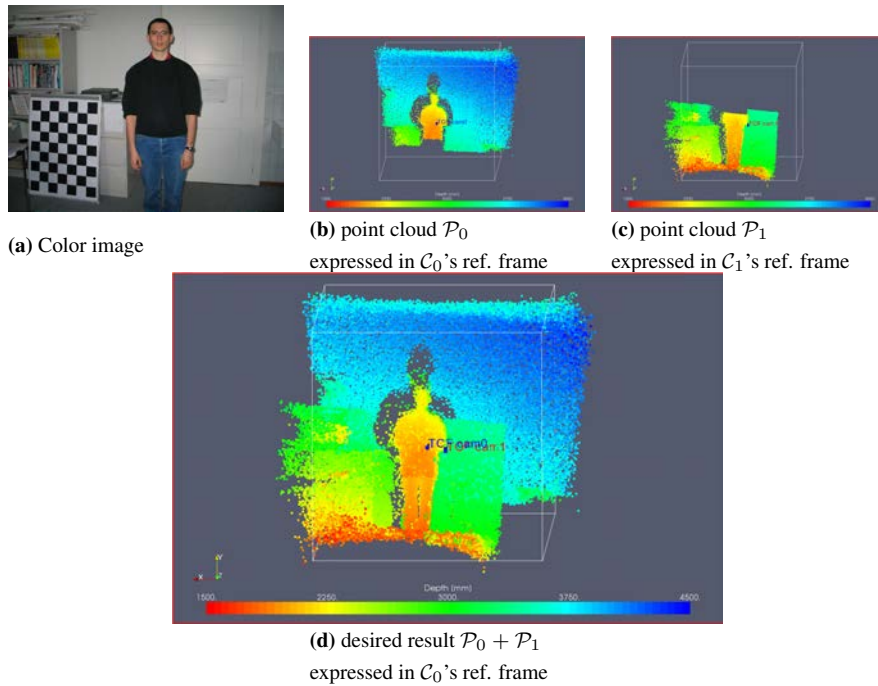


Figure 6.1: Point clouds registration for field of view extension.

refined by Zhang [130] and Bouguet [17]. Multiple view registration for intensity and color cameras is so commonly encountered that the method proposed by Bouguet [17] was integrated into the OpenCV library [123]. In this approach, the planar checkerboard target is imaged at different positions in the cameras' fields of view. Reference feature points (usually corners) are extracted in the images (an example is shown in fig. 7.10). The geometry of the target object is precisely known, i.e. the number of squares in the checkerboard is known and the square dimension is well defined. Then, the perspective projection model for the camera can be inferred from the acquired images by bundle adjustment for the feature points. Note that this technique involves a significant amount of human operation in the calibration procedure, since the calibration pattern must be moved and matched in many positions inside the cameras fields of view prior to bundle adjustment.

This photogrammetric calibration procedure provides camera *intrinsic* parameters such as focal length, position of the optical center and distortion, but also *extrinsic* parameters, namely the position of the camera with respect to the target pattern origin. If extrinsic parameters are matched for images of the *same* target pattern acquired by two devices \mathcal{C}_0 , \mathcal{C}_1 at different positions, the relative position of the two devices can be computed. It is important to note here that this alignment method is based only on intensity images produced by the TOF cameras : the TOF range map is not exploited in this alignment procedure. Unfortunately, Kahlmann & Ingensand [59] notes that the sensor lateral resolution is too low to use standard calibration targets for bundle adjustment for the SR-3000 camera. Nevertheless, Lindner & Kolb [75] successfully applied the OpenCV calibration procedure to register the range image of a PMD TOF camera

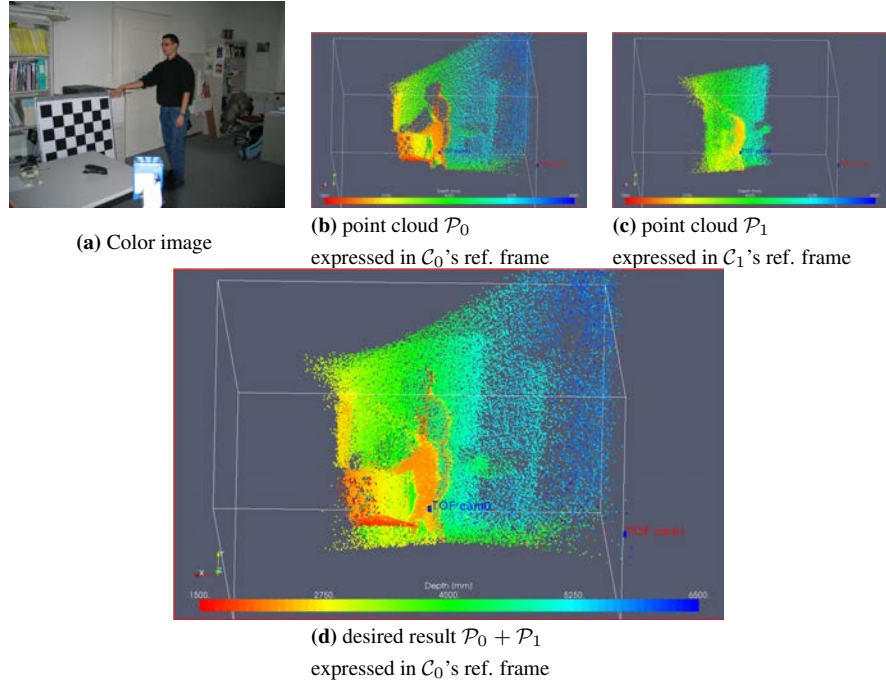


Figure 6.2: Point clouds registration for occlusions removal.

with a color image produced by a standard camera (in this case, the lateral resolution of the color imager is higher than the TOF sensor resolution). Experimental results for this method are presented in section 7.4.1.

6.4 Matched set of reference points

Given two point clouds \mathcal{P}_0 and \mathcal{P}_1 , expressed in the coordinates systems of range imaging devices \mathcal{C}_0 , respectively \mathcal{C}_1 , we want to determine the rigid body transformation $T_{\mathcal{C}_0, \mathcal{C}_1}$ allowing to express the point cloud \mathcal{P}_1 in the coordinates system of \mathcal{C}_0 . With at least three points matched across clouds \mathcal{P}_0 and \mathcal{P}_1 , it is possible to determine a transformation $T_{\mathcal{C}_0, \mathcal{C}_1}$, i.e. translation \mathbf{t} and rotation \mathbf{R} , which minimizes the distances between between the two point sets in a least square sense.

6.4.1 Least squares transform for two matched point sets

Arun *et al.* [3] propose to determine \mathbf{R} and \mathbf{t} by singular value decomposition (SVD) of a 3×3 matrix. Let $\{p_i^{(0)}\}$ and $\{p_i^{(1)}\}$ be the N matched points in \mathcal{P}_0 , respectively \mathcal{P}_1 ($i = 1, \dots, N$). The function to minimize is :

$$e = \sum_{i=1}^N \left\| p_i^{(0)} - \left(\mathbf{R} p_i^{(1)} + \mathbf{t} \right) \right\|^2 \quad (6.1)$$

6.4. Matched set of reference points

Subtracting the centroids $p^{(0)}$ and $p^{(1)}$ of the point sets allows to separate the translation and rotation problems. With the centroids defined as :

$$p^{(0)} = \frac{1}{N} \sum_{i=1}^N p_i^{(0)} \quad p^{(1)} = \frac{1}{N} \sum_{i=1}^N p_i^{(1)} \quad (6.2)$$

it is possible to define reduced coordinates $q^{(0)}$ and $q^{(1)}$:

$$q_i^{(0)} = p_i^{(0)} - p^{(0)} \quad q_i^{(1)} = p_i^{(1)} - p^{(1)} \quad (6.3)$$

Calling $\hat{\mathbf{t}}$ and $\hat{\mathbf{R}}$ the least squares solution minimizing eq. 6.1, we have :

$$\hat{\mathbf{t}} = p^{(0)} - \hat{\mathbf{R}} p^{(1)} \quad (6.4)$$

and the error function to minimize can be rewritten as :

$$e = \sum_{i=1}^N \left\| q_i^{(0)} - \mathbf{R} q_i^{(1)} \right\|^2 \quad (6.5)$$

which can be expanded :

$$e = \sum_{i=1}^N \left(q_i^{(0)} - \mathbf{R} q_i^{(1)} \right)^T \cdot \left(q_i^{(0)} - \mathbf{R} q_i^{(1)} \right) \quad (6.6)$$

$$= \sum_{i=1}^N \left(q_i^{(0)T} \cdot q_i^{(0)} + q_i^{(1)T} \mathbf{R}^T \mathbf{R} q_i^{(1)} - q_i^{(0)T} \mathbf{R} q_i^{(1)} - q_i^{(1)T} \mathbf{R}^T q_i^{(0)} \right) \quad (6.7)$$

$$= \sum_{i=1}^N \left(q_i^{(0)T} \cdot q_i^{(0)} + q_i^{(1)T} \cdot q_i^{(1)} - 2 q_i^{(0)T} \mathbf{R} q_i^{(1)} \right) \quad (6.8)$$

Minimizing e is therefore equivalent to maximizing :

$$f = \sum_{i=1}^N q_i^{(0)T} \mathbf{R} q_i^{(1)} = \text{Trace} \left(\sum_{i=1}^N \mathbf{R} q_i^{(1)} q_i^{(0)T} \right) \quad (6.9)$$

$$= \text{Trace} \left(\mathbf{R} \sum_{i=1}^N q_i^{(1)} q_i^{(0)T} \right) = \text{Trace} (\mathbf{R} \mathbf{H}) \quad (6.10)$$

where:

$$\mathbf{H} = \sum_{i=1}^N q_i^{(1)} q_i^{(0)T} \quad (6.11)$$

Arun *et al.* [3] use SVD decomposition of \mathbf{H} to solve this problem. The SVD decomposition of \mathbf{H} can be written as :

$$\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{V} \quad (6.12)$$

where \mathbf{U} and \mathbf{V} are 3×3 orthonormal matrices, and $\mathbf{\Lambda}$ is a 3×3 diagonal matrix with nonnegative elements [109]. Arun *et al.* [3] show that $\mathbf{X} = \mathbf{V} \mathbf{U}^T$ is the orthonormal matrix which maximizes f . Usually \mathbf{X} is a rotation, with $\det(\mathbf{X}) = 1$, and the least squares solution to the problem is $\hat{\mathbf{R}} = \mathbf{X}$. Arun *et al.* [3] also discuss the case where

\mathbf{X} is a reflection, with $\det(\mathbf{X}) = -1$; this can occur only if the points $\{p_i^{(0)}\}$ are coplanar. In this case, one of the eigenvalues in $\mathbf{\Lambda}$ is zero; $\hat{\mathbf{R}}$ will be given by $\mathbf{X}' = \mathbf{V}' \mathbf{U}^T$ where \mathbf{V}' is \mathbf{V} with a sign inversion in the column corresponding to the null eigenvalue. Finally, if all points $\{q_i\}$ are collinear, the solution \mathbf{X} is not unique : there exists an infinity of rotations minimizing e . Fortunately, this case can easily be avoided by checking the input point sets $\{p_i\}$ prior to attempting least squares minimization, by requiring the matrix \mathbf{H} to have rank 3.

6.4.2 5 spheres calibration object for SR camera

Registration based on reference point correspondences requires to image a known calibration object \mathcal{C}_0 's and \mathcal{C}_1 's view. Images of this calibration object should allow to define the required reference points in each view. In order to avoid complications arising when the set of points is coplanar (see sec. 6.4.1), the calibration object should be three-dimensional. At the same time, self-occlusion should be limited, in order to guaranty the visibility of reference points across the different views.

For TOF cameras registration experiments, a calibration object was built from 5 highly reflective balls rigidly connected together (fig. 6.3a). Some balls have different diameters, in order to allow identifying each ball, independently of the object pose. The object defines a large volume, but is mostly empty, in order to limit self-occlusion. Ball centers should be used as reference points, as these points should not be affected by the object's pose relative to the camera.

The object is imaged simultaneously by two TOF cameras \mathcal{C}_0 and \mathcal{C}_1 , and the range images obtained $\mathbf{r}_0, \mathbf{r}_1$ are used to compute two 3D point clouds \mathcal{P}_0 and \mathcal{P}_1 (fig. 6.3b,c). The center of each ball is estimated in both point clouds, allowing to produce the point sets $\{p_i^{(0)}\}$ and $\{p_i^{(1)}\}$. Unfortunately, the quality of range map obtained for spherical objects is poor. This is caused by:

- sparse sampling of the sphere surface : in some cases, only 20 pixels are imaging a given sphere;
- low amplitude of the TOF return signal, since most of the camera active illumination is scattered away from the sensor by the spherical surface.

In the current implementation, the extraction of the sphere centers is done manually. This process could be automated by using a least squares spherical shape detector, but the detector should be customized to account for the aberrations caused by light scattering on the sphere surface.

With 5 matched sphere centers, the overdetermined system is solved, allowing to compute all coefficients of $\hat{\mathbf{R}}$ and $\hat{\mathbf{t}}$. Experimental results obtained with this intuitive registration method are presented in section 7.1.

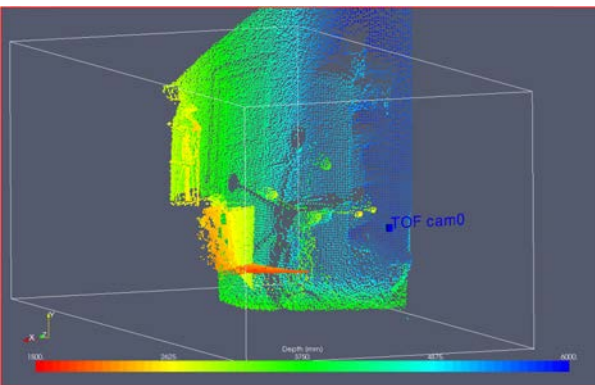
6.5 Iterative Closest Points (ICP) for registration

Since its introduction in the early 1990s (Besl & Mckay [12], Chen & Medioni [21]), the Iterative Closest Points (ICP) algorithm has been widely used for 3D data registration. The procedure provides a solution to the registration problem of two point clouds under the assumption of a rigid body transformation, and is based on minimization of an error function defined from distance from point in one cloud to their closest counterpart in the other cloud.

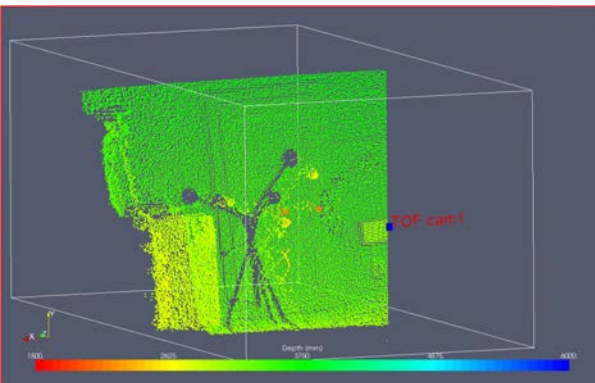
6.5. Iterative Closest Points (ICP) for registration



(a) Calibration object (illustrative color image). 5 highly reflective balls, assembled into a three-dimensional object. Ball centers are used as reference points.



(b) Point cloud \mathcal{P}_0 , measured from camera \mathcal{C}_0 . Each ball surface is sampled by 20 to 300 points in \mathcal{P}_0



(c) Point cloud \mathcal{P}_1 , measured from camera \mathcal{C}_1 . Each ball surface is sampled by 20 to 300 points in \mathcal{P}_1

Figure 6.3: Simple calibration object and associated point clouds.

6.5.1 Algorithm principle

The algorithm takes two points clouds \mathcal{P}_0 and \mathcal{P}_1 as input, along with an initial transform T_0 . The algorithm then produces an estimation T_{end} of the transformation that best registers \mathcal{P}_1 on \mathcal{P}_0 . The basic stages of the algorithm are:

1. **Matching (pairing)** of points between the sets \mathcal{P}_0 and \mathcal{P}_1 . Typically, each point $p_i^{(1)}$ of \mathcal{P}_1 , transformed by T_k , is matched to the closest point $p_{nn,i}^{(0)}$ in \mathcal{P}_0 :

$$p_{nn,i}^{(0)} = \operatorname{argmin}_{p_m^{(0)} \in \mathcal{P}_0} \left\| T_k \left(p_i^{(1)} \right) - p_m^{(0)} \right\|^2 \quad (6.13)$$

2. **Error metric** $e(T_k)$ assignment based on the point pairs. Typically, the total euclidean distance between paired points is computed:

$$e(T_k) = \sum_{p_i^{(1)} \in \mathcal{P}_1} \left\| T_k \left(p_i^{(1)} \right) - p_{nn,i}^{(0)} \right\|^2 \quad (6.14)$$

3. **Minimization** of error metric (see for example sec. 6.4.1), and update of transformation:

$$T_{(k+1)} = \operatorname{argmin}_{T \in E^{+(3)}} e(T) \quad (6.15)$$

4. **Termination check:** the process is iterated until the error metric gets below a threshold thr_{abs} , or until the variation from last iteration gets below a threshold thr_{rel} , or until the number of iteration i exceeds a fixed maximum N .

This workflow is illustrated in figure 6.4. A ready to use implementation of the ICP algorithm is available in the VTK library [67] [vtkIterativeClosestPointTransform](#)

6.5.2 Limitations of ICP methods

Unfortunately, since ICP methods rely on pairing of data in the two sets to register, they tend to fail when the overlap between the datasets is low. Chetverikov *et al.* [22] reports that advanced ICP algorithm still require around 50% overlap. Therefore, ICP methods do not seem appropriate for the case of FOV extension addressed in this work. Nevertheless, ICP can be used as a reference for specific datasets where the overlap is high, as was done in section 7.3.1.

6.6 Geometric primitives for registration

To determine the rigid body transformation between two point clouds, 6 degrees of freedom need to be eliminated. Section 6.4 showed that a matched set of 4 non-coplanar reference points could be used to this effect. Nevertheless, to increase the reliability of the transformation found, it is desirable to involve as many points as possible in the determination of the transformation. ICP techniques (sec. 6.5) are not convenient for the problem considered here (field of view extension) since they require a large overlap between clouds.

A promising approach to exploit as much information as possible from the point cloud is to extract geometric shapes describing accurately large parts of the point cloud.

6.6. Geometric primitives for registration

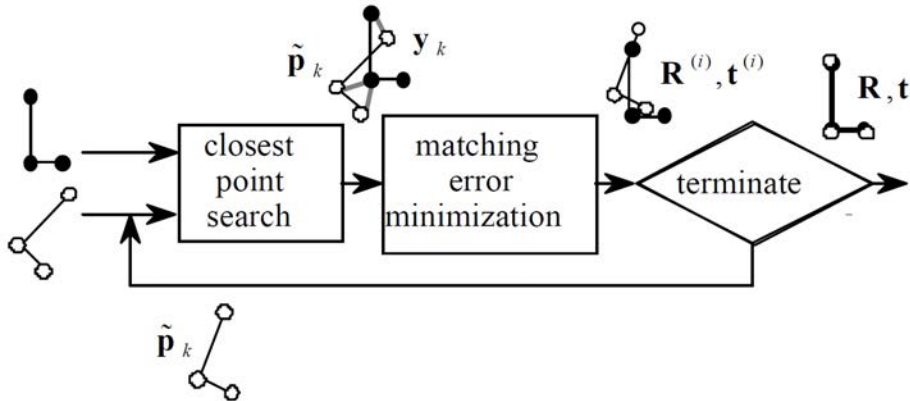


Figure 3.1 Data flow of ICP algorithm

Figure 6.4: ICP algorithm workflow principle - excerpt from [106]

Matched geometric primitives in different views allows elimination of some degrees of freedom. And, in contrast with ICP, using geometric primitives provides robustness even in conditions where the overlap area between the different views is small: the reliability of each primitive depends on the number of points on which it is constructed in its own cloud. In section 6.6.1, a non exhaustive list of usable geometric primitives is presented. Section 6.6.2 then introduces a method to extract plane primitives from a noisy point cloud.

6.6.1 Geometric primitives and degrees of freedom

Different geometric shapes will eliminate different DoF. Gelfand *et al.* [41] analyzed the geometric stability of ICP matching to compute the number of DoF left undetermined for the shapes presented in figure 6.5 . A plane primitive π involves 4 parameters:

$$\pi : \quad n_x \cdot x + n_y \cdot y + n_z \cdot z + d = 0 \quad (6.16)$$

where $\mathbf{n} = (n_x, n_y, n_z)^T$ is the plane normal vector and d is its distance to the origin. One condition imposed on the parameters is that \mathbf{n} has unit length: $\|\mathbf{n}\| = 1$. Matching a plane across two views leaves 3 DoF undetermined (fig.6.5, left): one rotation

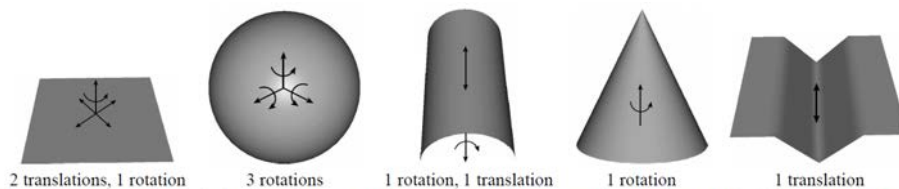


Figure 3: Some examples of simple shapes that are unstable. For each shape, the corresponding covariance matrix will have some number of small eigenvalues, and for those, the corresponding eigenvector specifies the direction of instability. Below each figure, the number and types of the instabilities are noted. A helix, which has one unstable screw motion, is missing, but helical shapes are not likely to arise in scanned data.

Figure 6.5: Geometric shapes and corresponding undetermined degrees of freedom (corresponding to ICP instabilities). Excerpt from [41]

Chapter 6. Registration of noisy range images

DoF corresponding to rotations in the plane and two translation DoF corresponding to translations in the plane.

A sphere primitive \mathcal{S} also involves 4 parameters, its center position $\mathbf{x}_0 = (x_0, y_0, z_0)^T$ and its radius r :

$$\mathcal{S} : (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 - r^2 = 0 \quad (6.17)$$

Matching a sphere across two views leaves the three rotations DoF undetermined (fig.6.5, second from left).

A cylinder primitive \mathcal{H} involves 7 parameters: the cylinder orientation \mathbf{v} , the position of its reference point \mathbf{x}_0 (axis point closest to the origin), and its radius r .

$$\mathcal{H} : \|\mathbf{x} - (\mathbf{x}_0 + ((\mathbf{x} - \mathbf{x}_0) \cdot \mathbf{v}) \mathbf{v})\|^2 - r^2 = 0 \quad (6.18)$$

Since the cylinder orientation vector \mathbf{v} has unit length ($\|\mathbf{v}\| = 1$) and is normal to \mathbf{x}_0 ($\mathbf{v} \cdot \mathbf{x}_0 = 0$), a cylinder has only five DoF. Matching a cylinder across two views leaves one rotation DoF and one translation DoF undetermined (fig.6.5, center).

Rabbani & van den Heuvel [99] used plane, cylinders, and spheres matching for registration of LIDAR data.

In this work, we focus our analysis on plane primitives. The motivation for this selection is that shapes with varying normals such as spheres, cones and cylinders are harder to acquire reliably with a TOF camera and its directional illumination. Combinations of plane primitives can be used to eliminate all DoFs. von Hansen *et al.* [121] used plane patches were used by to align LIDAR data. A wedge, i.e. two intersecting planes (fig.6.5, right), leaves only one DoF undetermined, corresponding to translations along the intersection axis of the planes. A cube corner, i.e. three intersecting planes, allows to eliminate all degrees of freedom, as is shown in section 6.7. An experimental limitation in our work was that only a small number of plane primitives could be reliably extracted from TOF point clouds. Section 6.7 discusses registration when the same cube corner is imaged by both cameras. Section 6.8 presents an registration method based on a single plane, but upon which reference points can be defined. But first, section 6.6.2 introduces a method allowing to obtain plane primitives fitted to point cloud data.

6.6.2 Random Sample Consensus (RANSAC) for plane primitive extraction

The Random Sample Consensus (RANSAC) is a technique which allows to fit a specific model to a dataset even when a high amount of outliers are present. The basic idea behind RANSAC is to fit the model only to a small set of points, but to check the model validity against the whole dataset.

The basic stages of the RANSAC algorithm for fitting a plane to a point cloud \mathcal{P} are:

1. **Initialize** the model π_{best} and the model compliance value c_{best} to 0.
2. **Randomly select a small** set $\{p_i\}$ of N points in \mathcal{P} . Typically, $3 < N < 10$.
3. **Fit** a plane model π_k to the set $\{p_i\}$.

6.7. Registration from cube corner planes

4. **Test compliance** c_k with the model π_k on the full point set \mathcal{P} .

$$c_k = 0; \quad \forall p_i \in \mathcal{P} : \quad \text{if } dist(p_i, \pi_k) < thr \quad : \quad c_k = c_k + 1 \quad (6.19)$$

5. **Memorize best model:** if $c_k > c_{best}$ then: $\pi_{best} := \pi_k$ and $c_{best} := c_k$.

6. **Iterate** the process.

A RANSAC algorithm for plane extraction is easily parametrized: the important parameters are the number of iterations $Niter$ and the distance threshold thr which is used to determine whether a point complies with the model or not.

6.7 Registration from cube corner planes

Section 6.6.1 stated that all DoF for registration can be eliminated by a matched set of three intersecting planes, that we will call abusively *cube corner planes*. This denomination is slightly abusive since it is not necessary for the planes to be orthogonal. The following discussion aims to prove this statement, and accurately describes the methods used for this registration. The proof will be based on the work by Faugeras & Hébert [33], who studied registration from a set of N matched plane.

Using the same notation as [33], we describe the geometry G of a scene through a set $(P(\mathbf{u}_i))_{i=1\dots N}$ of regions approximated by N plane surfaces, \mathbf{u}_i being the parameter vector of the i -th primitive. For plane primitives, the parameter vector combines the normal vector \mathbf{n}_i and the distance to the origin d_i : $\mathbf{u}_i = (\mathbf{n}_i, d_i)$. A matching M between two descriptions G and G' is a set of corresponding pairs $(P(\mathbf{u}_i), P'(\mathbf{u}'_i))$. Since the parameters \mathbf{u}_i and \mathbf{u}'_i are not expressed in the same coordinate systems, the rigid body transformation T linking those two systems must be determined. T should map each primitive P' of G' into the corresponding primitive P in G . The transformation is determined by minimizing the consistency measure :

$$g(M) = \min_T \sum \|\mathbf{u}_i - T(\mathbf{u}'_j)\|^2 \quad (6.20)$$

6.7.1 Decomposition of rigid-body transformation

The rigid body transformation T between two acquisitions of the same scene can be decomposed in a rotation R followed by a translation \mathbf{t} , so that :

$$T = \mathbf{t} * R \quad (6.21)$$

Under the assumption that the axis of rotation contains the origin of the reference coordinate system, this decomposition is unique. Applying this transformation to the plane $P(\mathbf{n}, d)$ yields the new plane $P_1(\mathbf{n}_T, d_T)$ where

$$\mathbf{n}_1 = R \mathbf{n} \quad \text{and} \quad d_1 = \mathbf{n}_1 \cdot \mathbf{t} + d \quad (6.22)$$

Given a matching M for N plane pairs: $M = (P(n_i, d_i), P'(n'_i, d'_i))_{i=1\dots N}$, the consistency measure to be minimized (eq. 6.20) can be rewritten as :

$$g(M) = \min_{T=\mathbf{t}*R} \sum_i \left(\|\mathbf{n}_i - R\mathbf{n}'_i\|^2 + W \cdot \|d'_i - d_i - \mathbf{n}_i \cdot \mathbf{t}\|^2 \right) \quad (6.23)$$

where W is a weighing factor. As proposed by Faugeras & Hébert [33], this sum is split into two terms, allowing to determine separately the best rotation R and translation \mathbf{t} .

6.7.2 Rotation estimation

Simple geometric consideration suffice to show that two non-parallel planes matches are sufficient for rotation determination. Using a single plane match does not constrain the degrees of freedom expressed by rotations and translations in this plane. Adding a second plane which intersects the first constrains all rotation degrees of freedom, leaving only one degree of freedom corresponding to translations along the intersection line of the two planes.

From equation 6.23, the function f which must be minimized for rotation determination is :

$$f = \sum_i \|\mathbf{n}_i - R\mathbf{n}'_i\|^2 \quad (6.24)$$

This minimization must account for the constraint that R is a rotation matrix, so that $R^T \cdot R = \mathbf{I}$. This constrained minimization problem is not straightforward to solve, especially since the function f is not linear.

Faugeras & Hébert [33] have introduced an elegant solution to this problem by using the quaternion representation of rotations. The rotation R can be expressed as quaternion products, so that for every vector \mathbf{v} in \mathcal{R}^3 , the relation

$$R \cdot \mathbf{v} = q \cdot (0; \mathbf{v}) \cdot \bar{q} \quad (6.25)$$

holds, where q and its conjugate \bar{q} are unit-norm quaternions representing the rotation. In the following, we will abusively use the notation \mathbf{v} to describe the corresponding quaternion $(0; \mathbf{v})$. Then, equation 6.24 can be rewritten as :

$$f = \sum_i \|\mathbf{n}_i - q \cdot \mathbf{n}'_i \cdot \bar{q}\|^2 \quad (6.26)$$

$$f = \sum_i \|\mathbf{n}_i \cdot q - q \cdot \mathbf{n}'_i\|^2 \quad (6.27)$$

The function $\mathbf{n}_i \cdot q - q \cdot \mathbf{n}'_i$ is linear in the coefficients of q , and can be represented by a 4×4 matrix A_i .

$$\mathbf{n}_i \cdot q - q \cdot \mathbf{n}'_i = q \cdot A_i \quad \text{where} \quad A_i = \begin{bmatrix} 0 & (\mathbf{n}_i - \mathbf{n}'_i)^T \\ (\mathbf{n}'_i - \mathbf{n}_i) & (\mathbf{n}'_i + \mathbf{n}_i)^o \end{bmatrix} \quad (6.28)$$

where the notation \mathbf{v}^o designs the 3×3 antisymmetric matrix obtained from the components of \mathbf{v} :

$$\mathbf{v}^o = \begin{bmatrix} 0 & v_3 & -v_2 \\ -v_3 & 0 & v_1 \\ v_2 & -v_1 & 0 \end{bmatrix} \quad (6.29)$$

Rewriting explicitly the norm in the function f allows to transform it into :

$$f = \sum_i q \cdot A_i \cdot A_i^T \cdot q^T = q \cdot B \cdot q^T \quad \text{where} \quad B = \sum_i A_i^T \cdot A_i \quad (6.30)$$

The minimization problem is now expressed as the minimization of $q \cdot B \cdot q^T$ under the constraint $\|q\| = 1$. This classical problem of linear algebra [109] is solved by finding the eigenvector q_{min} corresponding to the smallest eigenvalue of B . In practice, this can easily be achieved by means of singular value decomposition (SVD).

6.8. Master plane point cloud alignment

6.7.3 Translation estimation

Simple geometric considerations suggest that a minimum of three planes with a single intersection are required for translation determination Rabbani & van den Heuvel [99]. Following again Faugeras & Hébert [33], we define the sum S to be minimized as :

$$S = \sum_{i=1}^N \|d'_i - d_i - \mathbf{n}_i \cdot \mathbf{t}\|^2 \quad (6.31)$$

where N is the number of matched planes. The sum can be rewritten using $N \times 1$ difference vector D and the $N \times 3$ normals matrix C

$$S = \|D - C \cdot \mathbf{t}\|^2 \quad \text{where} \quad D = \begin{pmatrix} d'_1 - d_1 \\ \dots \\ d'_N - d_N \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} \mathbf{n}_1^T \\ \dots \\ \mathbf{n}_N^T \end{pmatrix} \quad (6.32)$$

This classical least squares problem can be solved by the *pseudoinverse* method [109]. The best translation is then given by :

$$t_{min} = (C^T \cdot C)^{-1} \cdot C^T \cdot D \quad (6.33)$$

If the plane list does not contain at least three independent planes, the rank of matrix C will be less than 3, so that the translation can not be determined. Experiments based on this registration methods are presented in sections 7.1.2 and 7.3.2.

6.8 Master plane point cloud alignment

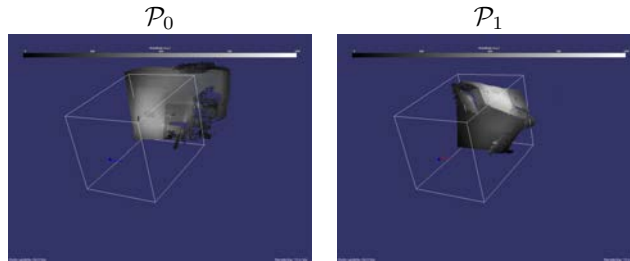
The basic idea of this registration method is to split the 3D registration problem in two problems easier to handle. We propose the following decomposition:

1. **Match one plane** across the datasets. This allows to eliminate 2 rotations and 1 translation DoF. Typically, RANSAC can be used to define the plane primitives $\pi^{(0)}$ and $\pi^{(1)}$ in both point clouds. For convenience, the point clouds can then be:
 - rotated so that this reference plane becomes aligned with the z-axis
 - translated to that the reference plane includes the origin.
2. **Affine alignment** in the reference plane. This alignment is based on feature points belonging to the matched plane: $\{p_i^{(0)}\} \in \pi^{(0)}$ and $\{p_i^{(1)}\} \in \pi^{(0)}$. Note that this operation is a 2D process. If the reference plane is transformed into the xy plane as suggested above, the z component can be entirely left out of the computations.

The splitting described above allows to potentially leverage many elaborate alignment methods developed for 2D images. The workflow of master plane alignment is illustrated in figure 6.6. In the following discussion, we will only consider 2D affine alignment from a set of matched points. In the experiments presented in this work (sec. 7.3.3 and 7.4.3), the matched points are selected manually by a human observer. But we note that more advanced image processing techniques, such as correlation or SIFT [77], could be used to provide the point matches automatically.

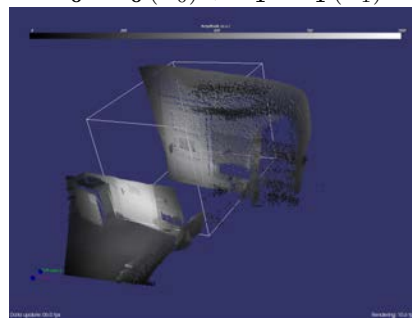
Chapter 6. Registration of noisy range images

Starting from unregistered point clouds:



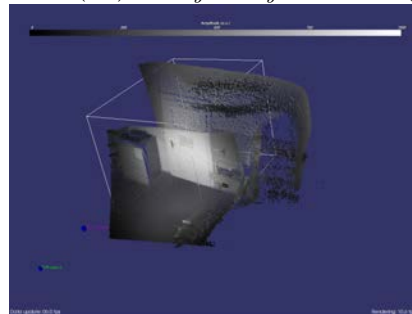
Align master plane with xy plane:

$$\mathbf{R}_0 \circ \mathbf{T}_0 (\mathcal{P}_0) + \mathbf{R}_1 \circ \mathbf{T}_1 (\mathcal{P}_1)$$



Align datasets in xy plane:

$$\mathbf{R}_0 \circ \mathbf{T}_0 (\mathcal{P}_0) + \mathbf{T}_{xy} \circ \mathbf{R}_{xy} \circ \mathbf{R}_1 \circ \mathbf{T}_1 (\mathcal{P}_1)$$



Transform back into \mathcal{C}_0 's reference frame:

$$\mathcal{P}_0 + (\mathbf{R}_0 \circ \mathbf{T}_0)^{-1} \circ \mathbf{T}_{xy} \circ \mathbf{R}_{xy} \circ \mathbf{R}_1 \circ \mathbf{T}_1 (\mathcal{P}_1)$$

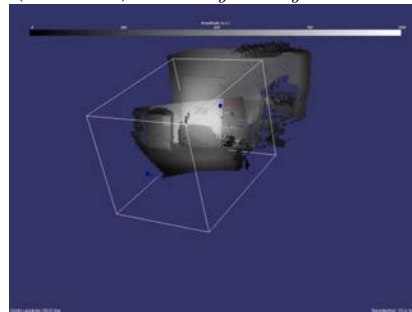


Figure 6.6: Workflow for master plane alignment procedure.

6.8.1 Usage considerations

Master plane alignment requires to have a plane object imaged by both cameras. However, the overlap region may be small. Since the plane primitive is isolated in each point cloud, the robustness of the plane parameters used can be increased by points belonging to the plane, but which are *not* in the overlap region. This increased robustness is useful for noisy datasets. Another key property of this registration method is its speed : the registration is obtained in only a few minutes. The idea is to benefit from plane regions already present in the scene, such as walls, to guide the registration. Once an overlapping plane patch is found, human selection of matched points in the plane for affine registration can be performed in a few minutes or seconds. In comparison bundle adjustment registration requires to move an ad-hoc target to various positions inside the cameras fields of view. This data acquisition step alone takes more time than master plane registration. The reduction in time required for registration is especially beneficial when a multi-camera system is initially set up : the cameras can be quickly repositioned to test a specific configuration.

6.9 Conclusion

The expected success of registration methods discussed above for registration of TOF data is summarized in table 6.1. The bundle adjustment or calibration object techniques allow for registration, even in situations with small overlap between the views. But the amount of data points actually involved in the registration is small : 5 (x, y, z) points for the calibration object, 48 pairs of pixel coordinates (i, j) for bundle adjustment. It remains to be seen whether the registration output from these techniques is sufficiently reliable when the input data is noisy, as is the case for systems based on TOF cameras. In contrast to these methods, ICP registration and cube corner planes registration involve a large number of data points in the registration, of the same order of magnitude as the total number of points in each cloud. This increases the registration stability when noisy input is considered. However, this improvement is only attained in configurations with large overlap between views (70% or more).

Finally, the master plane method represents a compromise between those situations. It involves as many data points as possible in the computation of the master plane orientation, allowing to reliably eliminate 3 DoF. But it does not require large overlap, since the 3 remaining DoF are eliminated by using a small set of reference points in the overlap region of the plane surface. Moreover, this registration approach is fast, as it is not tied to a specific calibration object being brought into the scene. Chapter 7 illus-

Table 6.1: Expected success of registration techniques for SR data

Technique	Large overlap	Small overlap
Bundle adjustment (on intensity images)	✓	✓
5 spheres calibration object	✓	✓
ICP registration	✓✓✓	✗
Cube corner planes	✓✓	✗
Master plane	✓✓	✓✓

Chapter 6. Registration of noisy range images

trates the application of those different registration methods to real datasets, allowing to select the most appropriate method for the combination of TOF views.

7

Registration experiments on noisy range images

This chapter presents registration experiments performed on TOF data. Evaluation metrics are introduced in section 7.1. Then, the different registration methods discussed in chapter 6 are tested. Two Swissranger TOF cameras were used for data acquisition. To avoid interference, the first TOF camera was operated at $f_0 = 20\text{MHz}$, while the second camera was operated at $f_1 = 21\text{MHz}$. Section 7.2 presents results obtained with a simple calibration object. The other methods are tested on two datasets; the first one has large overlap (section 7.3), while the overlap is small ($< 33\%$) for the second configuration (section 7.4).

7.1 Evaluation of registration error

For the evaluation of registration results, two methods are used:

- a qualitative evaluation, where the match between rendered point clouds is examined visually by a human observer.
- a quantitative evaluation by nearest neighbor distance.

Quantitative measures complement human observations, and can be used to evaluate alignment methods on datasets that may confuse a human observer.

7.1.1 Nearest neighbor distance metric

We have seen that the ICP algorithm works by minimizing the nearest neighbor distance across the point clouds (see eq. 6.14). The quantity minimized is the squared sum of squared distances $d(i)^2$ between each transformed point $p_{T,i}^{(1)}$ of \mathcal{P}_1 , and its

Chapter 7. Registration experiments on noisy range images

nearest neighbor $p_{nn,i}^{(0)}$ in \mathcal{P}_0 . The distance $d(i)$ is given by:

$$d(i) = \sqrt{\|p_{T,i}^{(1)} - p_{nn,i}^{(0)}\|^2} \quad (7.1)$$

The RMS registration error σ_{reg} is then given by

$$\sigma_{reg} = \frac{1}{N_1} \sum_{i=1}^{N_1} d(i)^2 \quad (7.2)$$

where N_1 is the number of points in cloud \mathcal{P}_1 .

One issue worth mentioning here is that this measure is appropriate only in situations when the overlap between views is high. Since the problem we try to solve is field of view extension, this condition will not be satisfied in many test datasets. In this case, this metric can nevertheless be used, but only on matching subsets $\mathcal{P}_{s,0}$ and $\mathcal{P}_{s,1}$ of the point clouds. Those subsets can be defined by introducing a threshold τ_m on the distance for matched points. If the nearest neighbor for a point $p_{T,i}^{(1)}$ of \mathcal{P}_1 is farther away than the threshold, this point isn't included in the subset $\mathcal{P}_{s,1}$. Formally, a weighing coefficient w_i is introduced [106]:

$$\sigma'_{reg} = \sqrt{\frac{1}{W-1} \sum_{i=1}^{N_1} w_i \cdot d(i)^2} \quad (7.3)$$

where :

$$w_i(d(i)) = \begin{cases} 1 & \text{if } d(i) < \tau_m \\ 0 & \text{otherwise} \end{cases} \quad W = \sum_{i=1}^{N_1} w_i(d(i)) \quad (7.4)$$

An alternative proposed by Zhang [129] is to use the sum of the average the nearest neighbor distance μ and its standard deviation ς as the error measure ϵ :

$$\mu = \frac{1}{N_1} \sum_{i=1}^{N_1} d(i) \quad \varsigma = \sqrt{\frac{1}{N_1-1} \sum_{i=1}^{N_1} (d(i) - \mu)^2} \quad (7.5)$$

$$\epsilon = \mu + \varsigma \quad (7.6)$$

In this case again, the terms can be weighted to limit the influence of outliers [106]:

$$\mu' = \frac{1}{W} \sum_{i=1}^{N_1} w_i \cdot d(i) \quad \varsigma' = \sqrt{\frac{1}{W-1} \sum_{i=1}^{N_1} w_i \cdot (d(i) - \mu')^2} \quad (7.7)$$

$$\epsilon' = \mu' + \varsigma' \quad (7.8)$$

The threshold τ_m must be chosen appropriately for each scene; a good practice is to set τ_m to approximately twice the noise level in the individual point clouds.

7.1.2 Illustration first order metric on depth-from-focus microscope datasets

To illustrate nearest neighbor distance accuracy metric ϵ , a real microscope 3D dataset obtained by the depth from focus method is used. 3D renderings of this dataset are

7.1. Evaluation of registration error

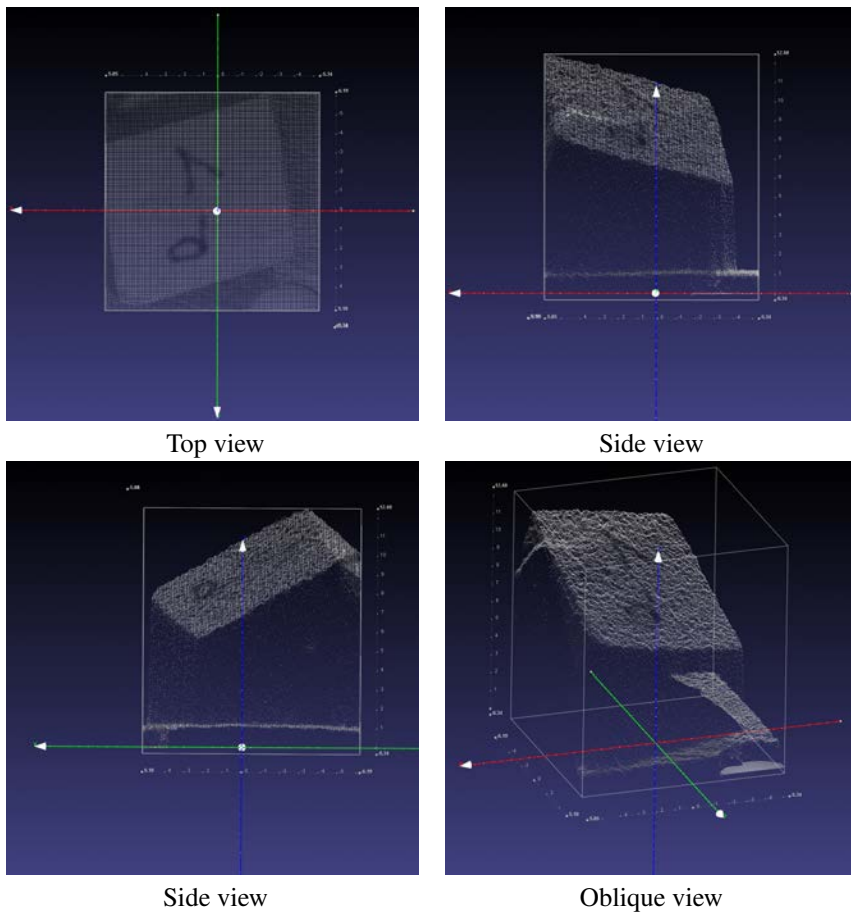


Figure 7.1: Microscope scene for synthetic registration experiment

Chapter 7. Registration experiments on noisy range images

presented in figure 7.1. This dataset contains large plane regions, and is therefore appropriate for a registration procedure using cube corner planes. For this experiment, a synthetic rigid-body transformation was applied to the dataset : it was first rotated around the axis $\mathbf{w} = (1, 3, 1)^T$ by an angle $\theta = 38^\circ$, and then translated by $\mathbf{T} = (-5, 3, 4)^T$. Table 7.1 show the parameters of the transformation in its first row. The second row contains the parameters obtained through registration from cube corner planes. The results are satisfying, but can be improved upon : figure 7.2(b) shows small discrepancies between the original dataset and the re-aligned data. This discrepancies can be further reduced by using the ICP method for alignment. The resulting parameters are presented in the third row of table 7.1. As can also be seen in figure 7.2(c), the final alignment is significantly improved. After ICP alignment, the nearest neighbor distance metric is $\epsilon = 0.0401$ mm.

This example showed that the nearest neighbor distance metric is useful to determine the quality of a registration between two point clouds. The results correlate well with the qualitative evaluation obtained by visual comparison of the point clouds. In section 7.4 and 7.3, this metric is used to evaluate registration quality on real TOF datasets.

7.2 TOF data registration based on 5 spheres calibration object

A first validation experiment was carried out by using camera configuration most favorable to point cloud registration : A single SR-3100 camera was used, and two range images were recorded *sequentially* with the same device and unchanged settings; the camera was translated upward by a small fixed amount between the two image acquisitions. This procedure maximizes the probability of successful registration since : no inter-device variability is introduced, the fixed pattern noise is the same for both range maps, and the geometrical matching is simplified (the optical axes are strictly parallel). Results of this experiment (fig. 7.3) indicate that a successful registration of two point clouds produced by a Swissranger camera is indeed possible : the two point clouds merge seamlessly.

Further tests were carried out while removing some of those favorable conditions. Test configurations and registration results are summarized in table 7.3. It appears that the calibration experiments failed in all cases where the rotation matrix \mathbf{R} was non trivial. This is most probably related to the fact that the sphere center coordinates are not determined with enough accuracy from the point cloud data. Such inaccuracies can lead to an erroneous rotation matrix \mathbf{R} . Small rotation errors cause large discrepancies between the merged point clouds, since the scene extends far away from the calibration object.

Table 7.1: Rigid body transformation parameters for alignment experiment on synthetically misaligned 3D microscope data.

	Rotation axis			θ	T		
Synthetic	0.30151	0.90453	0.30151	38.000	-5.0000	3.0000	4.0000
Cube corner planes	0.29920	0.90721	0.29571	38.312	-5.2816	2.8284	4.1720
ICP post-process.	0.29997	0.90561	0.29981	37.994	-4.9248	2.8844	4.2110

7.2. TOF data registration based on 5 spheres calibration object

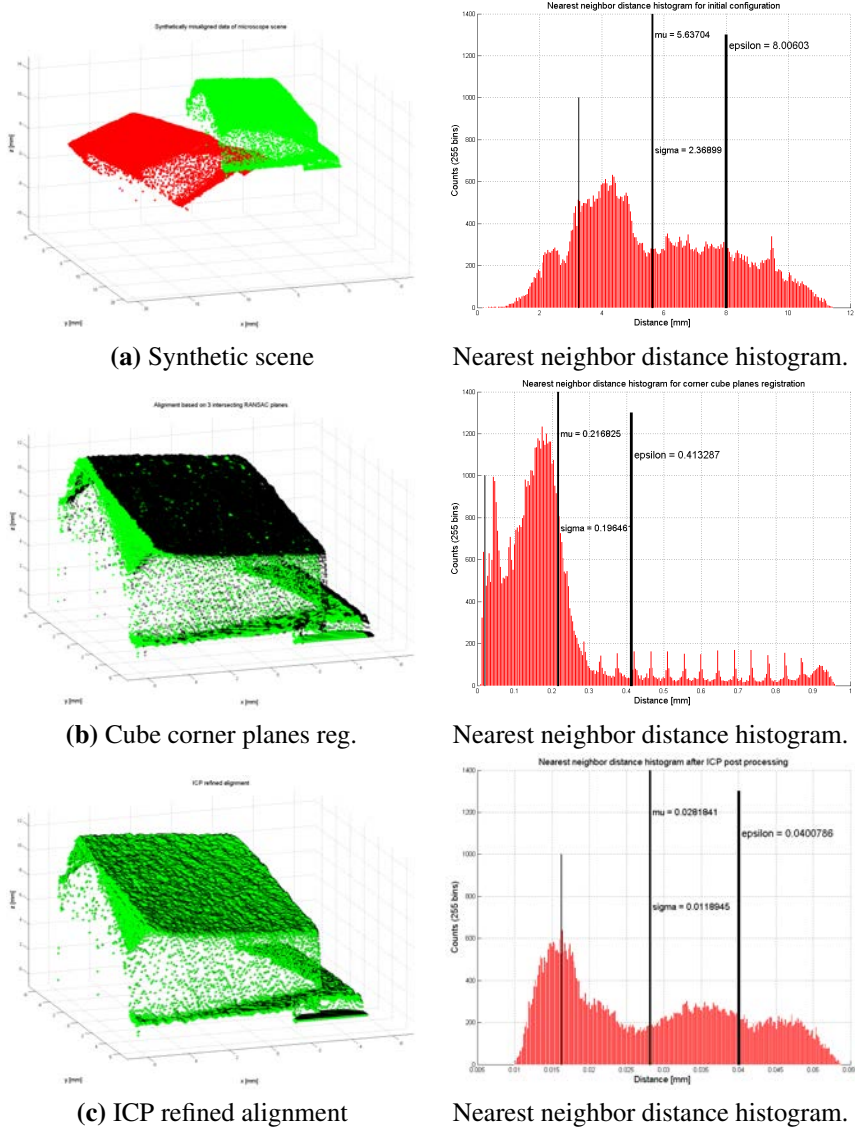


Figure 7.2: Synthetic registration experiment

Table 7.2: Illustration of nearest neighbor distance metrics - Microscope dataset

	Initial configuration	Cube corner planes registration	ICP post-processing
μ [mm]	5.64	0.217	0.0282
ς [mm]	2.37	0.197	0.0119
ϵ [mm]	8.01	0.413	0.0401
$\epsilon = \mu + \varsigma$			

Chapter 7. Registration experiments on noisy range images

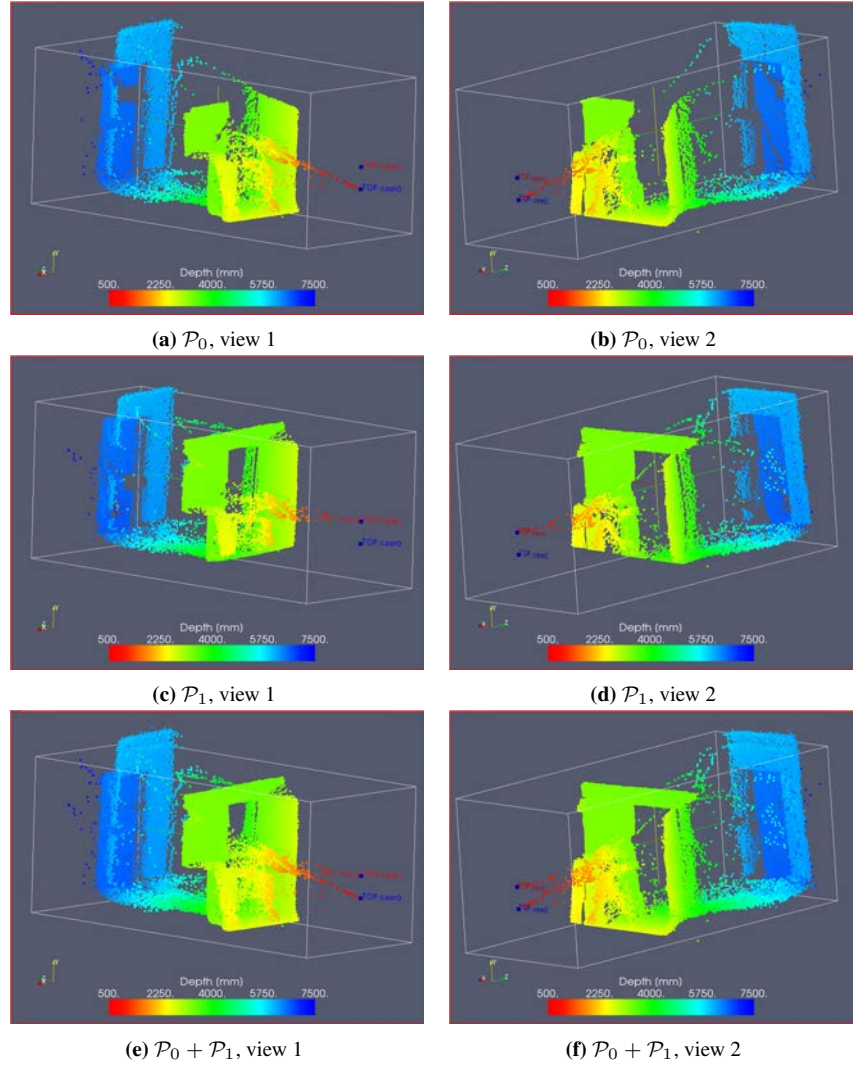


Figure 7.3: Registration based on calibration object : for most favorable acquisition configuration (test case A), the range image registration is qualitatively good.

Table 7.3: Point cloud registration results for specific multicamera configurations, based on matching of calibration object formed by 5 spheres.

	Test case →	A	B	C	D	E	F
Test config.	Same camera	×	×	×	×		
	Translation only	×		×			
	Same frequency	×	×			×	
Registration successful		×		×			

7.3. Large overlap between point sets

Nevertheless, we must note here that even in cases where the registration fails, the procedure allows to reduce the distance between the two point clouds, and provides a first approximation of the rotation matrix R , which would be very difficult to estimate manually. The translation parameters could possibly be interactively refined by a human supervisor to provide more accurate point cloud merging. However, this was not attempted since more advanced registration methods can be employed, as described in the next sections.

7.3 Large overlap between point sets

Figure 7.4 shows the corner of a room imaged by a SR-3100 device \mathcal{C}_0 and a SR-3000 device \mathcal{C}_1 , located slightly closer to the scene, so that the field of view of \mathcal{C}_1 is totally included in \mathcal{C}_0 's field of view. Note that one of the walls is made of bricks (rectangular shapes are visible). This structured plane surface allows to ensure that the planes can easily be matched. This dataset will be used to compare three registration methods : ICP, cube corner planes and master plane.

7.3.1 ICP registration

Using the ICP algorithm provided in VTK [67] on this dataset allows to determine a rigid body transformation T which projects the point cloud \mathcal{P}_1 into the coordinates system of camera \mathcal{C}_0 . In this experiment, 300 ICP iterations were allowed. T can be expressed by the 4matrix :

$$\mathbf{T} = \begin{pmatrix} 0.888671 & 0.189498 & 0.417558 & -262.392 \\ -0.255927 & 0.960559 & 0.108753 & 193.048 \\ -0.360481 & -0.20351 & 0.902119 & 683.903 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (7.9)$$

The registration results are good both visually: fig. 7.5 and numerically: table 7.4. The distance error ϵ between the point sets was reduced from 432 mm to 36.9 mm. Note that although alignment metrics are good, some inconsistencies are visible for the brick wall : the gaps between bricks seem slightly misaligned.

7.3.2 Registration based on cube corner planes

Figure 7.4 shows a dataset containing three easily matched intersecting plans, and is therefore a good candidate for alignment using plane primitives. The first step in the matching process is the estimation of plane primitives for each view. This is done using the RANSAC algorithm presented in section 6.6.2. The RANSAC plane search used 3000 iterations. The threshold for considering that a given points belong to the plane was set to $thr = 25\text{mm}$. RANSAC search was performed successively three times, while points already assigned to a plane were removed from the point cloud.

In both datasets, RANSAC first returned a plane corresponding to the brick wall, then a plane corresponding to the floor, and finally a plane corresponding to the remaining wall; see figure 7.6.

Since three (intersecting) planes are matched across views, all degrees of freedom of the transformation can be determined from the plane coefficients. The final rigid body transformation T which projects the point cloud of the second camera (SR-3000)

Chapter 7. Registration experiments on noisy range images

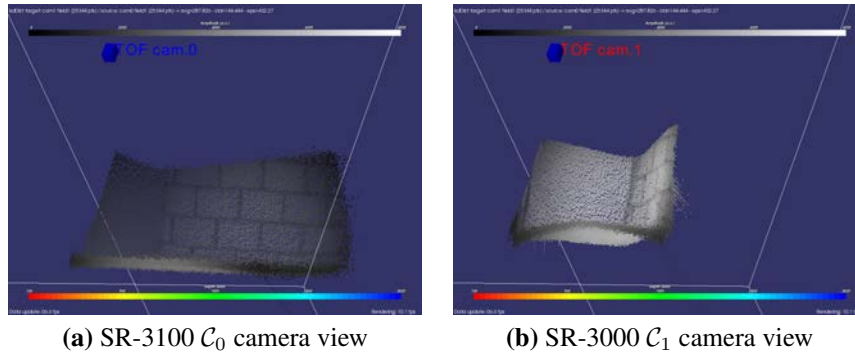


Figure 7.4: Scene with large overlap : corner of a room

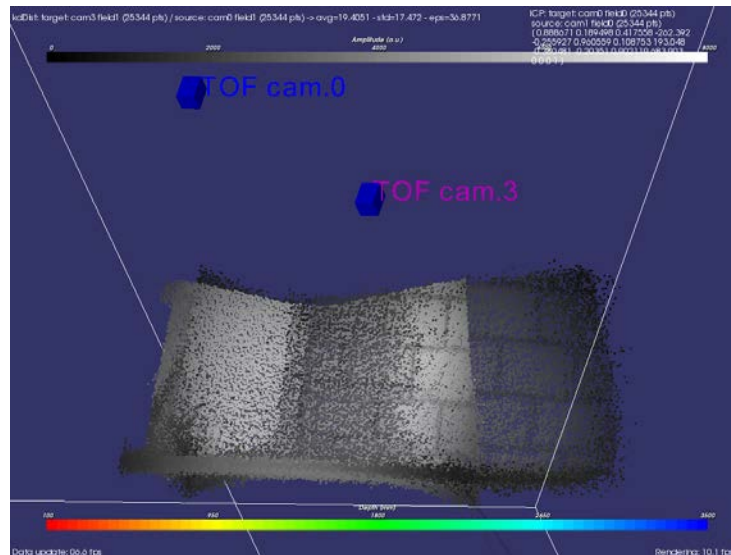


Figure 7.5: Room corner scene - point clouds successfully registered using ICP algorithm.

7.3. Large overlap between point sets

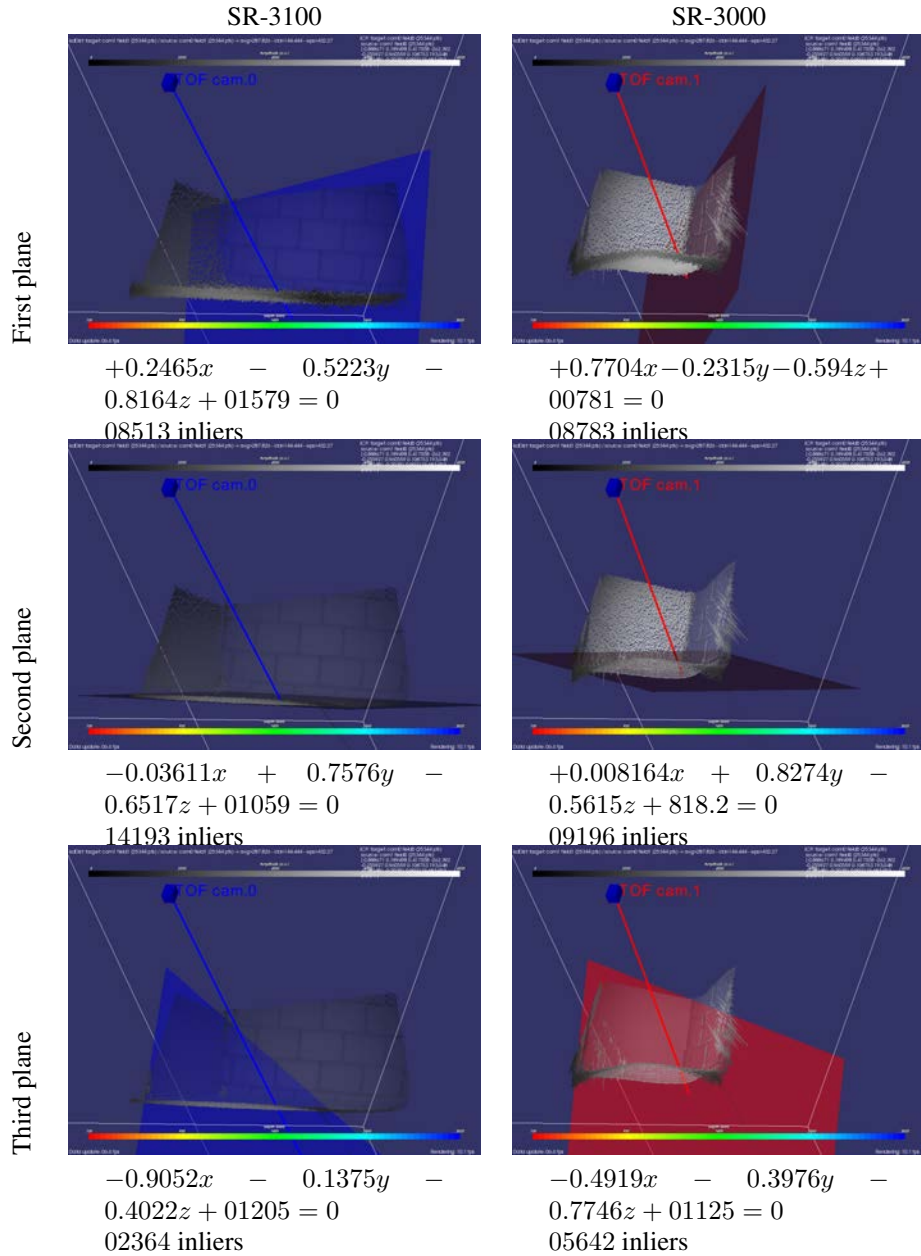


Figure 7.6: RANSAC extraction of plane primitives for room corner scene; 3000 RANSAC iterations were performed, the inlier threshold was set to $thr = 25\text{mm}$.

Chapter 7. Registration experiments on noisy range images

into the coordinates system of the first camera (SR-3100) can be expressed by the 4matrix :

$$\mathbf{T} = \begin{pmatrix} 0.801337 & 0.33839 & 0.493306 & -271.935 \\ -0.387585 & 0.92183 & -0.00273992 & 283.418 \\ -0.455671 & -0.189002 & 0.869852 & 714.036 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (7.10)$$

The registration results are good: fig. 7.7. As was the case for ICP registration on the same dataset, the ϵ error is below 50 mm. Table 7.4 shows that ICP registration performs slightly better, but plane based registration is a valid alternative registration approach for large overlap datasets.

7.3.3 Master plane registration

In this experiment, the brick wall was chosen as a master plane, since it is the first plane primitive returned by 3000 iterations of the RANSAC algorithm with a threshold set at $thr = 25\text{mm}$. Moreover, the intensity pattern of the bricks intersections allowed to easily define five reference points on the plane surface. The final rigid body transformation T which projects the point cloud of the second camera (SR-3000) into the coordinates system of the first camera (SR-3100) is expressed by the 4matrix :

$$\mathbf{T} = \begin{pmatrix} +0.838137 & +0.230656 & +0.494292 & -378.7 \\ -0.320188 & +0.941684 & +0.103494 & +199.3 \\ -0.441595 & -0.245009 & +0.863113 & +687.1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (7.11)$$

The registration results (see fig. 7.8) are good for the plane taken as reference, but show discrepancies between the two point sets for other regions. In particular, the position of the room side wall shows a clear gap between the two datasets.

7.3.4 Comparison of registration methods

Table 7.4 compares the error metric obtained on the large overlap dataset with ICP, cube corner planes and master plane registration. The lowest error is obtained with ICP, which marginally outperforms cube corner planes and master plane registration. The results can be compared to the noise level in the scene. Averaged point clouds $\mathcal{P}_{avg,0}$, resp. $\mathcal{P}_{avg,1}$ are computed from averaging 50 consecutive range images. Then error metrics are computed between single point clouds \mathcal{P}_0 , resp. \mathcal{P}_1 , and the averaged point clouds; see table 7.5. It can be seen that the registration errors are up to eight times large than the noise level for this scene. A partial explanation for this large

Table 7.4: Nearest neighbor distance metrics - Large overlap dataset - Comparison of ICP, cube corner planes and master plane alignment.

	Initial configuration	ICP registration	Cube corner planes	Master plane
σ_{reg} [mm]	322	24.8	29.7	47.9
μ [mm]	287	18.4	21.1	33.1
ς [mm]	144	16.7	20.9	34.6
ϵ [mm]	432	35.1	41.9	67.8

7.3. Large overlap between point sets

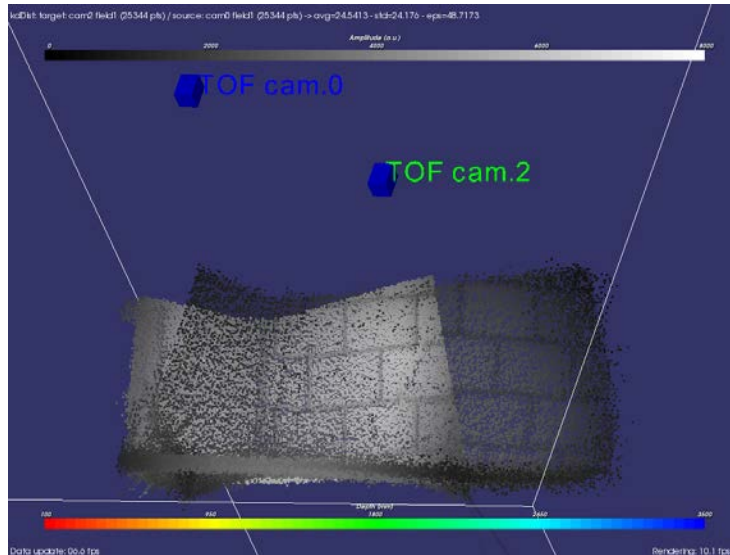


Figure 7.7: Room corner scene - point clouds have been registered using cube corner planes.

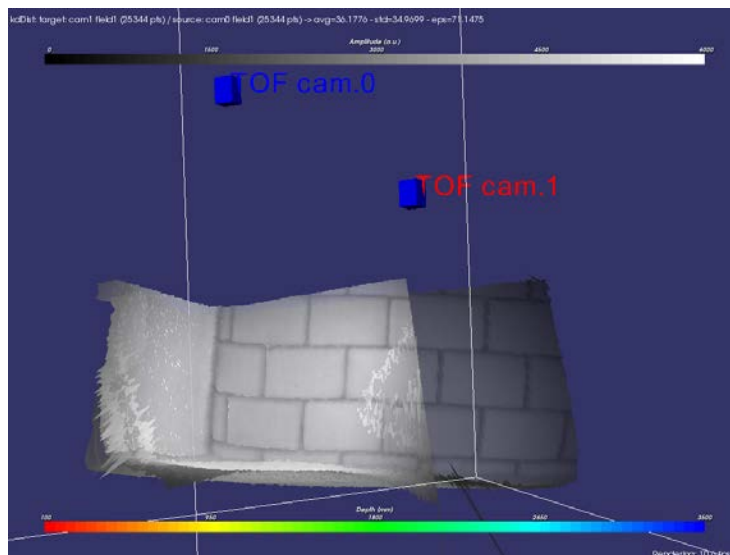


Figure 7.8: Room corner scene - point clouds have been registered by matching the brick wall plane as a master plane.

Chapter 7. Registration experiments on noisy range images

Table 7.5: Nearest neighbor distance metrics - Large overlap dataset - Noise levels for each point cloud produced by the TOF cameras.

	SR-3100	SR-3000
σ_{reg} [mm]	12.54	6.31
μ [mm]	9.84	5.11
ς [mm]	7.77	3.69
ϵ [mm]	17.6	8.80
σ_{avg} [mm] see sec. 4.3.1	17.3	9.33

relative error can be found in degradation by multipath: the test scene is a cube corner, multipath effects cannot be neglected (see sec. 4.4.1). Since the camera viewpoints are different, multipath errors are different for each camera.

Overall, the three registration methods performed similarly well on this large overlap dataset. Note that master plane registration provides a better estimation of the rotation in the brick plane wall: the gaps between the bricks are better aligned in figure 7.8 than they were in fig. 7.7 and 7.5.

7.4 Small overlap between point sets

Figure 7.9 shows region near the door of a room imaged by a SR-3100 device C_0 and a SR-3000 device C_1 . The fields of view of the two devices overlap only partially: C_0 is pointed towards the door and the wall next to it, while C_1 is pointed towards the floor in front of the door. The overlap is approximately 33%. For registration errors computation, τ_m is chosen as 100mm. In order to reduce the influence of noise in the comparison, the range and amplitude data were averaged from 50 range images acquired sequentially. This dataset will be used to compare three registration methods: ICP, bundle adjustment on intensity images, and master plane.

7.4.1 Registration from bundle adjustment of checkerboard target

As a reference against other methods, we carried out a registration experiment based on the calibration toolbox provided by Bouguet [17], using a large checkerboard pattern (each square was 117 mm) as a calibration target for bundle adjustment. Figure 7.10 shows a set of calibration images, where the corners have been detected and highlighted. The toolbox used allows to determine *intrinsic* parameters of the camera, such as the focal length f_o or the coordinates (c_x, c_y) of the principal point. In tables 7.6 and 7.7 values obtained with this toolbox are compared to the manufacturer's data for two SR-3X00 devices. In both cases, the agreement is good : the manufacturer's data lies within the uncertainty domains of calibration values. We can note here that

Table 7.6: Intrinsic camera parameters for SR-3100 (sn097027)

Parameter	MESA	calib.
f_o [mm]	8.0 ± n.a.	8.04 ± 0.23
c_x	85.0 ± n.a.	83.5 ± 4.7
c_y	76.7 ± n.a.	80.3 ± 5.4

7.4. Small overlap between point sets

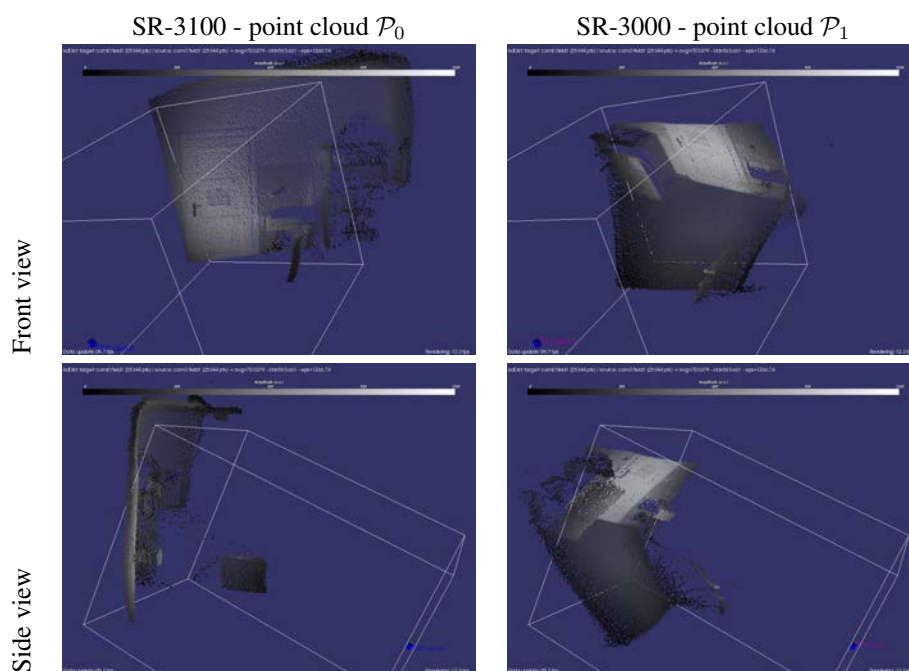


Figure 7.9: Scene with small overlap : office door - Data was averaged from 50 images.

Table 7.7: Intrinsic camera parameters for SR-3000 (sn296012)

Parameter	MESA	calib.
f [mm]	$8.0 \pm \text{n.a.}$	7.98 ± 0.26
c_x	$95.1 \pm \text{n.a.}$	93.8 ± 5.2
c_y	$56.3 \pm \text{n.a.}$	51.6 ± 5.9

Chapter 7. Registration experiments on noisy range images

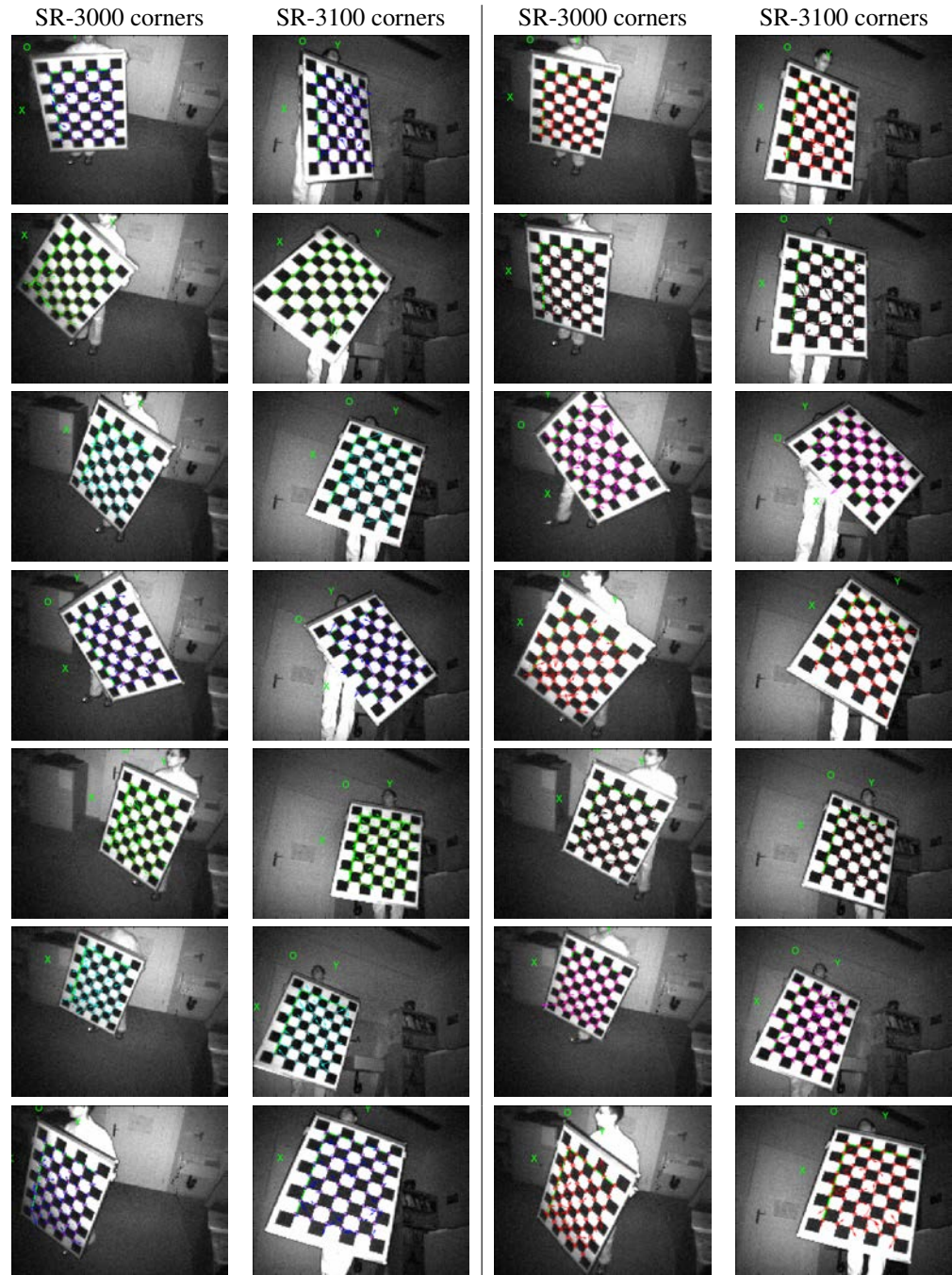


Figure 7.10: Amplitude images used for calibration based on bundle adjustment - Extracted corners with tool by Bouguet [17] - Images 0 to 13

7.4. Small overlap between point sets

the SR-3000 device (sn296012) has an optical center far away from the CCD sensor center.

When the intrinsic parameters are known, bundle adjustment is performed to compute the *extrinsic* parameters of the multi-camera setup. For each target position, the scale and pose of the target object (checkerboard) is used to estimate the position of each camera relative to the target. This estimation is repeated for all target images, allowing to compute a relative displacement of the different cameras that minimizes position errors. Figure 7.11 shows a rendering of the computed relative camera positions.

Unfortunately, the *extrinsic* parameters obtained with this calibration method are not good enough for point cloud registration: see figure 7.12 and table 7.9. Note in particular that the number of matched points W is very low. It appears that the rotation between the point clouds is roughly correct, but the translation is erroneous. In consequence, the point clouds are 'parallel', so that their intersection is very small. Although disappointing, those bad registration results were expected since

- the reliability of the camera position estimation relative to the target is low. This is related to the low lateral resolution of the TOF sensor (176×144): corner feature points used to determine the scaling of the target in the image are not known precisely.
- the camera model parameters computed from the amplitude images are (slightly) different from the parameters used in the SR-3000 driver software for the transformation of the range map into a 3D cloud of points.

7.4.2 ICP registration

Figure 7.5 illustrates a registration failure with ICP. Table 7.9 shows that the distance metric for matched points in the clouds \mathcal{P}_0 and \mathcal{P}_1 isn't significantly different from the value obtained for the unregistered case. The amount of matched points W appears increased, but visual observations indicate that those matches are erroneous. This example clearly illustrates that ICP methods aren't appropriate when the overlap between the points clouds to register is small.

7.4.3 Master plane registration

For each point cloud, the largest plane primitive is found using RANSAC plane detection. 3000 RANSAC iterations are performed, the inlier threshold is set to $thr = 100\text{mm}$. A set of 4 matched point pairs was manually defined in the master plane, which was the wall and door plane in this dataset. The alignment result is illustrated in 7.14. Although the overlap is small, alignment results appear satisfying.

7.4.4 Comparison of registration methods

Table 7.8 provides reference error metrics for individual point clouds in this small overlap dataset. Note that the threshold τ_m was set to 100mm. Table 7.9 compares error metrics for ICP registration (fig. 7.5), bundle adjustment (fig. 7.12) and master plane registration (fig. 7.14) on this dataset. Table 7.9 confirms the visual evaluation result: master plane registration outperforms the two other methods.

W is very low for bundle adjustment. This is consistent with the observation of a 200 mm gap between the two datasets. ICP registration has a higher correspondence

Chapter 7. Registration experiments on noisy range images

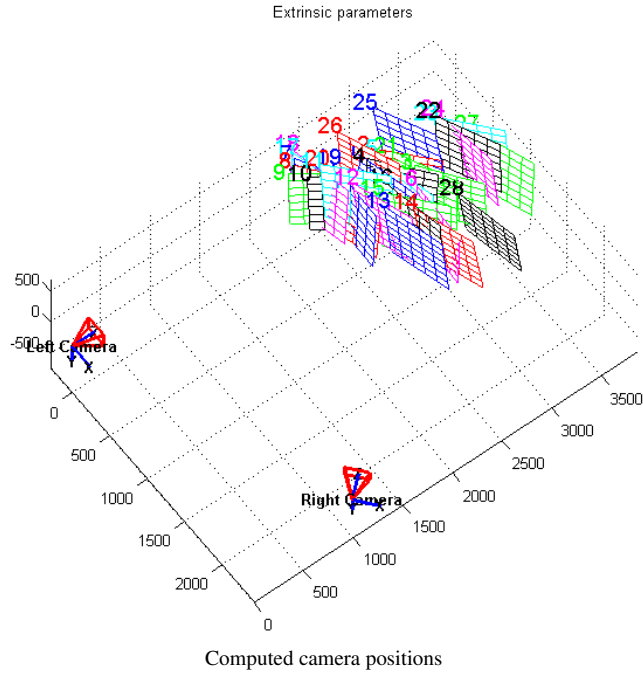


Figure 7.11: Bundle adjustment calibration toolbox results

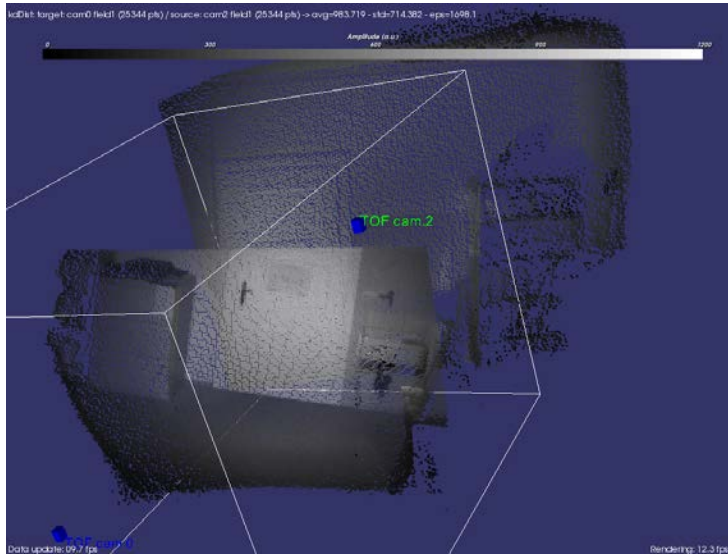
Table 7.8: Nearest neighbor distance metrics - Small overlap dataset - Noise levels for each point cloud produced by the TOF cameras. The threshold τ_m was set to 100mm.

	SR-3100	SR-3000
W [1]	21740	21029
σ'_{reg} [mm]	46.5	44.2
μ' [mm]	39.5	37.0
ζ' [mm]	24.5	24.2
ϵ' [mm]	64.0	62.2
σ_{avg} [mm] see sec. 4.3.1	138	156

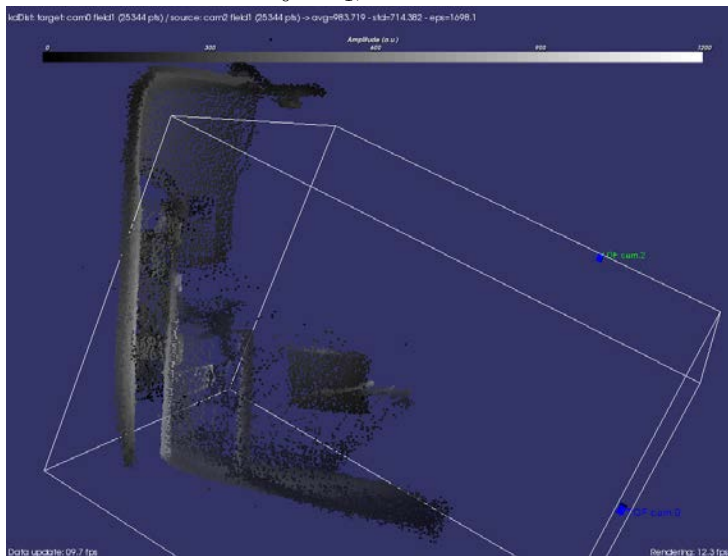
Table 7.9: Small overlap dataset - Nearest neighbor distance metrics, comparison. The threshold τ_m was set to 100mm.

	Initial configuration	ICP registration	Bundle adjustment	Master plane registration
W [1]	2702	7228	1164	8457
σ_{reg} [mm]	58.7	62.4	62.7	42.4
μ' [mm]	51.8	56.6	57.7	36.6
ζ' [mm]	27.6	26.2	24.7	21.3
ϵ' [mm]	79.4	82.8	82.3	57.9

7.4. Small overlap between point sets



$\mathcal{P}_0 + \mathcal{P}_1$, Front view



$\mathcal{P}_0 + \mathcal{P}_1$, Side view

Figure 7.12: Scene with small overlap : office door - Amplitude based bundle adjustment does not provide the correct registration: for the wall and door region, a 200 mm gap is visible in the side view. Moreover, the front view shows that the door frame isn't aligned correctly.

Chapter 7. Registration experiments on noisy range images

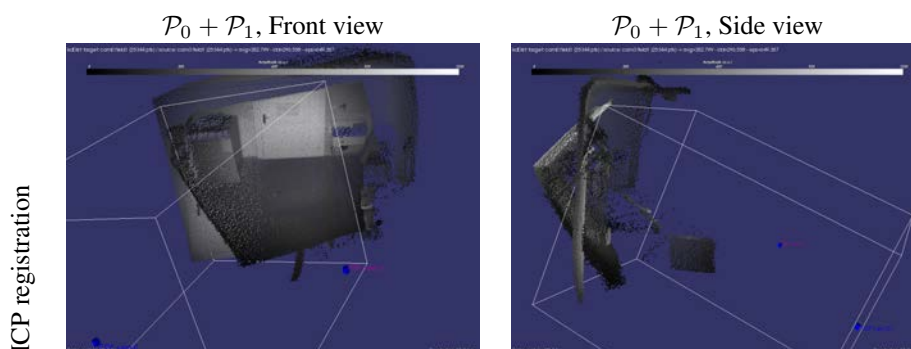


Figure 7.13: Scene with small overlap : office door - ICP registration does not provide the correct registration: the algorithm correctly minimizes the sum of nearest neighbor distances, but since the overlap is small (approx 33%), the result is incorrect.

score W , but visual observation put in evidence that those registration were erroneous. This can be observed in the other metrics (σ'_{reg} to ϵ'), which are similar to values obtained with the initial configuration. In contrast with these two methods, master plane registration decreases error metrics to levels very close to the noise observed for individual cameras.

7.4.5 Error estimation on segmented point subsets

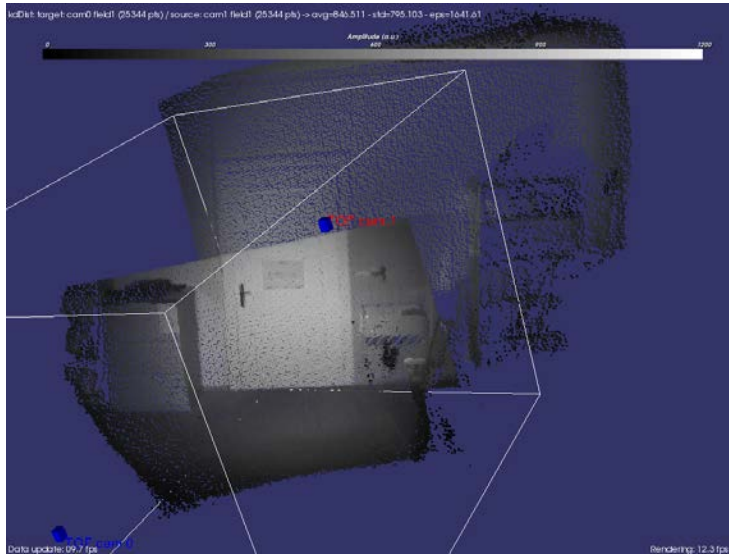
In order to check the registration quality, a new object is introduced in the scene. In figure 7.15, this new object is a person. The error metrics measured with bundle adjustment and master plane registration are reported in table 7.10. The master plane registration technique clearly provides the best results. Figures 7.15 and 7.16 illustrate the registration results. In those illustrations, a segmentation algorithm was used to highlight the person walking in the scene. The segmentation algorithm is based on background subtraction: each point in the clouds is considered as belonging to the foreground subset if the displacement from its background range is larger than 9 times the standard deviation of the range measurement. Subsets $\mathcal{P}_{s,0}$ and $\mathcal{P}_{s,1}$ corresponding to the person are isolated and displayed.

This experiment can be repeated for different positions of the person in the field of view. Figures 7.17 and 7.18 show another example of registration results for this scene with the bundle adjustment, resp. master plane, registration techniques. As before, the master plane registration technique clearly good results; see table 7.11.

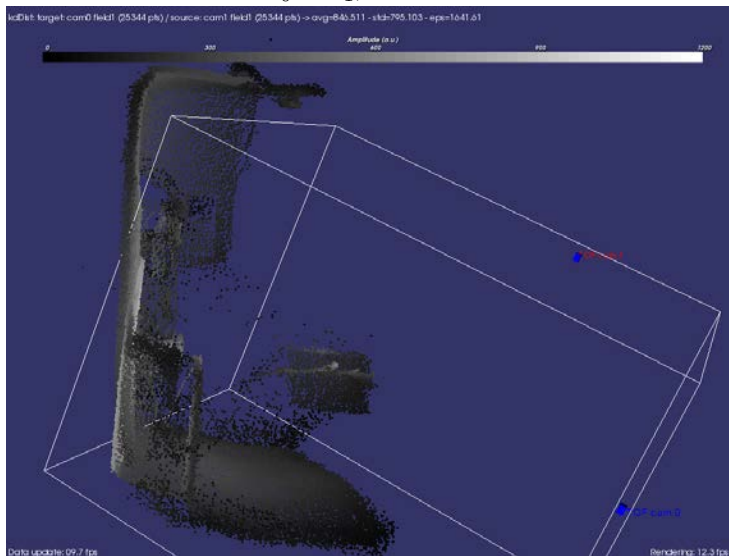
Table 7.10: Small overlap dataset - Nearest neighbor distance metrics, scene showing a person (example 1, see fig. 7.15).

	Initial configuration	Bundle adjustment	Master plane registration
W [1]	5660	2663	10307
σ_{reg} [mm]	56.03	65.77	37.9
μ' [mm]	49.8	60.5	31.8
ζ' [mm]	25.8	25.7	20.6
ϵ' [mm]	75.5	86.3	52.4

7.4. Small overlap between point sets

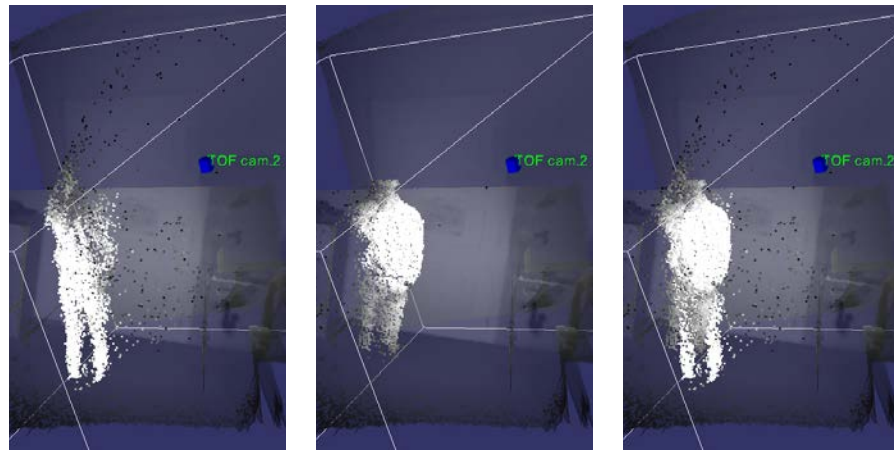


$\mathcal{P}_0 + \mathcal{P}_1$, Front view



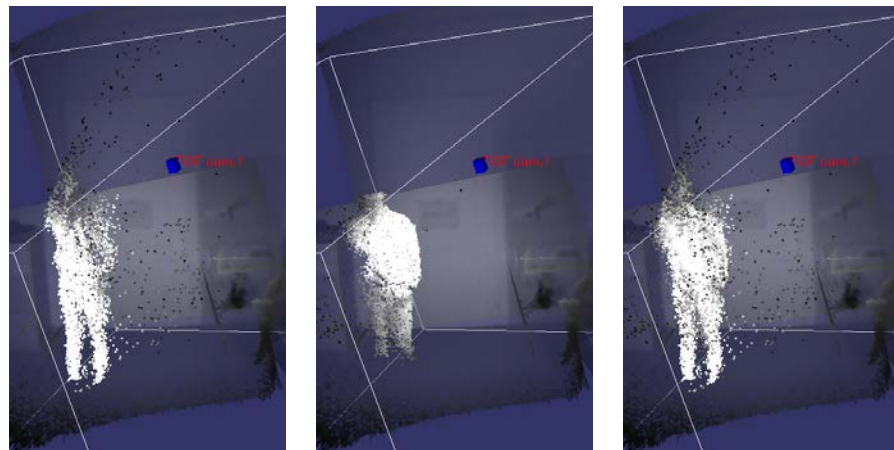
$\mathcal{P}_0 + \mathcal{P}_1$, Side view

Figure 7.14: Master plane alignment of small overlap dataset - RANSAC plane selection: 3000 iterations, inlier threshold $thr = 100\text{mm}$. - 4 alignment points chosen in plane



subset $\mathcal{P}_{s,0}$, 5072pts. subset $\mathcal{P}_{s,1}$, 4691pts. $\mathcal{P}_{s,0} + \mathcal{P}_{s,1}$
Background data $\mathcal{P}_{bg,0}$ and $\mathcal{P}_{bg,1}$ shown as transparent surfaces for better readability.

Figure 7.15: Bundle adjustment registration - Point subsets corresponding to a person, Example 1. Visually, the registration appears crude: the legs are mismatched in the two subsets.



subset $\mathcal{P}_{s,0}$, 5072pts. subset $\mathcal{P}_{s,1}$, 4691pts. $\mathcal{P}_{s,0} + \mathcal{P}_{s,1}$
Background data $\mathcal{P}_{bg,0}$ and $\mathcal{P}_{bg,1}$ shown as transparent surfaces for better readability.

Figure 7.16: Master plane alignment of small overlap dataset - Point subsets corresponding to a person, Example 1. Visually, the registration appears good.

7.4. Small overlap between point sets

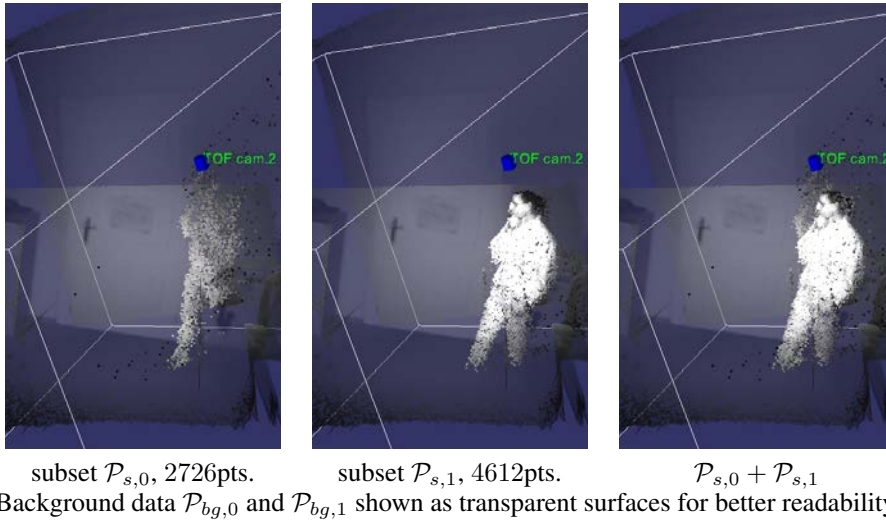


Figure 7.17: Bundle adjustment registration - Point subsets corresponding to a person, Example 2. Visually, the registration appears crude: the upper body is mismatched in the two subsets, as can be seen by comparing the elbow position.

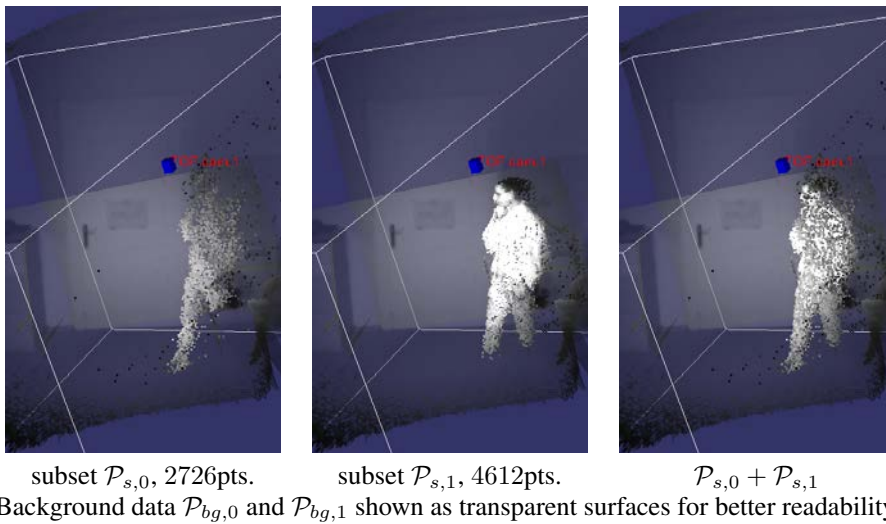


Figure 7.18: Master plane alignment of small overlap dataset - Point subsets corresponding to a person, Example 2. Visually, the registration appears good.

Chapter 7. Registration experiments on noisy range images

Table 7.11: Small overlap dataset - Nearest neighbor distance metrics, scene showing a person (example 2, see fig. 7.16).

	Initial configuration	Bundle adjustment	Master plane registration
W [1]	4625	3782	9909
σ_{reg} [mm]	62.17	68.3	42.1
μ' [mm]	56.3	63.6	35.2
ζ' [mm]	26.3	24.8	23.0
ϵ' [mm]	82.7	88.4	58.2

7.5 Conclusion

The suitability of the registration methods discussed in chapter 6 for integration of SR point clouds has been tested. The results are summarized in table 7.12. The 5 spheres calibration object did not enable successful registration in practical experiments, since the determination of spheres centers from TOF data proved too unreliable. The results obtained with bundle adjustment from the camera calibration toolbox also proved too inaccurate for successful point cloud merging. Here, the reliability of camera extrinsic parameters determination was the main problem; this unreliability was probably caused by the low sensor resolution of SR cameras (176×144).

ICP provides good registration results, but only for datasets with large overlap. The same drawback appears in cube corner planes registration. Finally, the master plane technique allowed registration of datasets even when the overlap was small. This technique assumes only that the overlap regions contains a plane region with an easily matched amplitude pattern. As this situation is often found in surveillance applications, this technique represents a good compromise between the required calibration effort and the quality of the final registration. Experiments have shown that the registration remains valid when new objects are added in the scene, thus allowing integration of point cloud data produced by the cameras in real-time, even if the registration is performed offline.

Table 7.12: Suitability of point cloud registration techniques to SR data

Technique	Large overlap	Small overlap
5 spheres calibration object	(✓)	✗
Bundle adjustment (on intensity images)		✗
ICP	✓	✗
Cube corner planes	✓	✗
Master plane	✓	✓

8

Applications

This chapter focuses on applications enabled by range image registration. The system we consider is a network of real-time range cameras, used for example in human activity monitoring for access control or safety applications. A custom software allowing simultaneous acquisition with multiple time-of-flight (TOF) cameras is presented. Registration procedures are used to combine the different TOF views into a unique 3D scene. Specific examples of usage of a TOF cameras network for access control are presented. In access control, a small security zone, usually close to a door, must be monitored. Section 8.3 illustrates occlusions removal, while section 8.4 illustrates field of view extension.

8.1 Network of TOF cameras

The system we consider is a network of real-time range imaging cameras based on the TOF measurement principle. In the network, cameras are operated from different viewpoints. The viewpoints are not known *a-priori*, but adjusted in order to get the most complete representation of the scene (fig. 8.1). Such a network can be used in human safety applications, and must comply with stringent timing requirements. In the following sections, examples acquired with a minimal TOF camera network are presented. The network is minimal since it contains only two TOF cameras, but serves as a proof of concept for larger networks, which could be realized when TOF cameras become more affordable.

8.2 Software implementation

A custom software application for simultaneous data acquisition with multiple TOF cameras had to be developed to realize a TOF camera network. The C++ software implementation is built on top of the SR-3000 TOF driver and application programming interface. The software is controlled through a graphical user interface (GUI) based on

Chapter 8. Applications

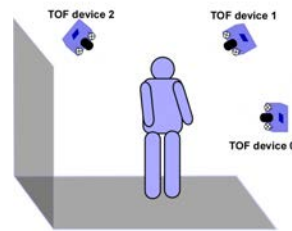


Figure 8.1: Schematic of a multi-camera network. All TOF cameras operate simultaneously. Each TOF device produces ranges images from a different viewpoint. The fields of view overlap partially.

the wxWidgets cross-platform GUI library [107]. The GUI is divided in three main parts (see fig. 8.2):

- The global controller, which keeps a list of the active cameras. This controller is used to launch a synchronous acquisition with all TOF cameras, but also to launch procedures for TOF data registration. Currently, three registration procedures are supported:
 - ICP registration, as described in sec. 6.5,
 - cube corner planes registration, as described in sec. 6.7,
 - master plane registration, as described in sec. 6.8,

The global controller also allows to compute the nearest neighbor distance metrics, introduced in sec. 7.1.1. Finally, this interface allows to control the properties of the 3D display. In particular, the visibility of a specific camera data object can be toggled on or off.

- A camera controller, supporting the SR-3000 camera series. Each active camera has its own controller instance. The controller allows to change camera settings, such as operation frequency or integration time. The interface also includes displays for the various image data returned by TOF cameras, such as range, amplitude, or z distance. Scattering compensation or range image segmentation can be toggled on or off. Camera transformation matrices produced by registration procedures can be loaded. Finally, parameters can be set for the RANSAC plane extraction procedures (see sec. 6.6), to produce plane primitives used in registration. The feature points used in master plane registration are manually defined by clicking on the image displays.
- The 3D display, based on the VTK visualization toolkit [67]. This display combines the point cloud data from individual cameras into a single 3D scene. The properties of the display can be changed, for example to view only data by a specific camera, or to compare the current range image to previously recorded background data. This 3D rendering is particularly useful to evaluate registration quality.

The main guideline during the development of this software was the necessity to allow real-time operation. This is required to enable interactive setup of a TOF camera

8.2. Software implementation

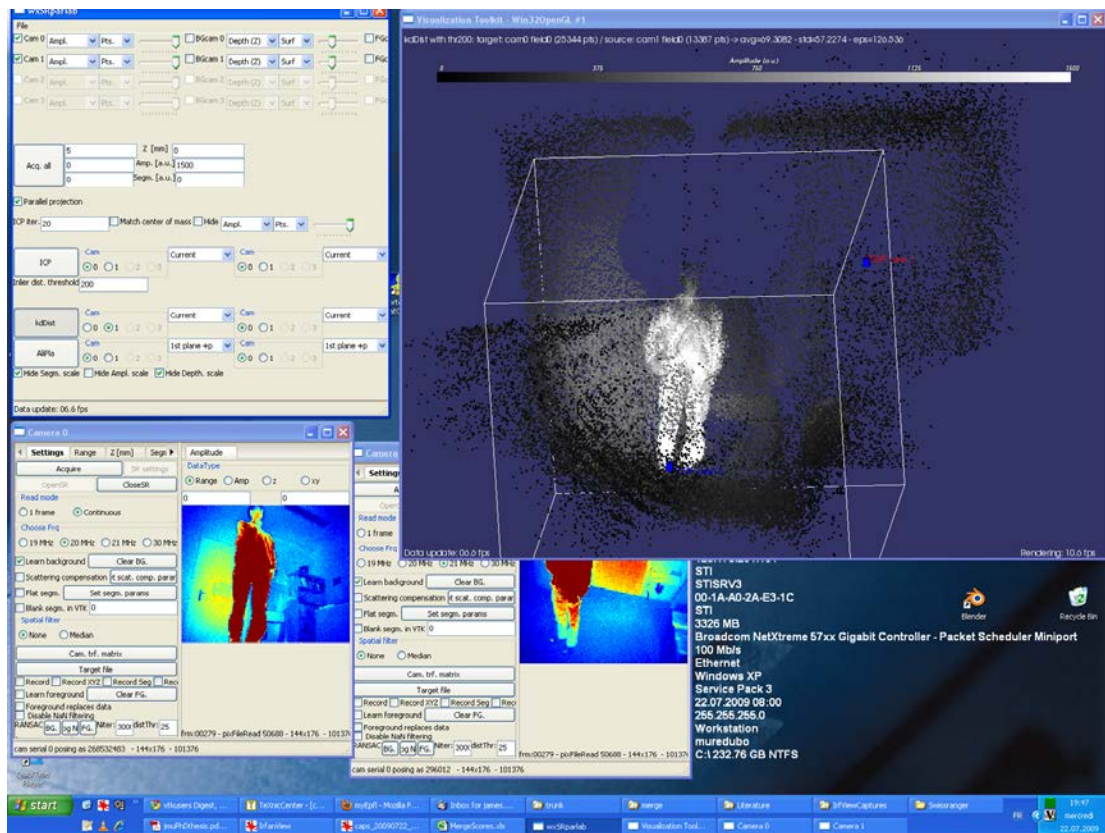


Figure 8.2: Screenshot of custom software for TOF camera network. Top-left corner: global controller. Bottom: 2 instances of the camera controller. Top-right: 3D display.

Chapter 8. Applications

network. In order to allow testing of different camera configurations, the registration of TOF views must be performed easily and quickly from a user's perspective. Registration using ICP takes between 1 second and 10 seconds, depending on the number of iterations authorized for the algorithm. Cube corner planes registration is performed in less than one second, once the plane primitives are defined. The computation time required to define a plane primitive is approximately 200 ms. For master plane registration, the user must currently define point pairs from the 2D displays. This operation takes usually less than a minute. The computation time is smaller than one second.

In its current version, the presented software serves as a proof of concept for a TOF camera network, and can be used as a helper tool to correctly set the positions of the different TOF devices for a given application. The software main functions include:

- simultaneous acquisition of TOF data (tested with up to 2 cameras);
- simultaneous playback of recorded TOF data streams (tested with up to 5 data streams);
- performance comparison of scattering compensation models;
- evaluation of registration accuracy; both visually and by using the nearest neighbor metric;
- range image segmentation.

In our experiments, only two TOF cameras were available for data acquisition. Therefore, only pairwise registration was considered. Pairwise registration is usually sufficient during the setup of a multi-camera system. Higher accuracy approaches, involving global registration and an even distribution of the registration errors across the views, should only be used once the camera positions have been determined and are not expected to change anymore. In the examples presented below, registration was performed using the master plane method presented in section 6.8.

8.3 Occlusion removal

In the context of access control, TOF cameras are used to quickly segment persons in a scene, since the range readings for the persons are significantly different from range reading in an empty scene. But, when multiple persons are allowed in the same scene, occlusions may occur. Occlusions also occur in the vicinity of doors. In the example of figures 8.3 to 8.5, access through a door is monitored. Figure 8.3 presents the empty scene. Since this scene is static, time averaging is used to reduce the noise level. Figure 8.4 presents a snapshot taken when the door is being opened. Note that the person opening the door is occluded in the view of the TOF camera directly in front of the door. The second TOF camera, placed at an oblique angle from the wall, allows to image (and potentially track) the person as soon as the door is opened. In figure 8.4, the person moves away from the door. In that situation, the door is occluded in the view of the TOF camera facing it. Based only on this range image, it is impossible to check that the door is securely closed. Additional data from the second TOF camera allows to perform this check.

A second example of occlusion removal, in the field of safety systems, is illustrated below. Kohoutek[68] proposes to use a TOF camera in safety systems where humans and robots share the same workspace. To prevent injuries, robots should be automatically stopped if a human could be hit during their motion. This requires to check for

8.3. Occlusion removal

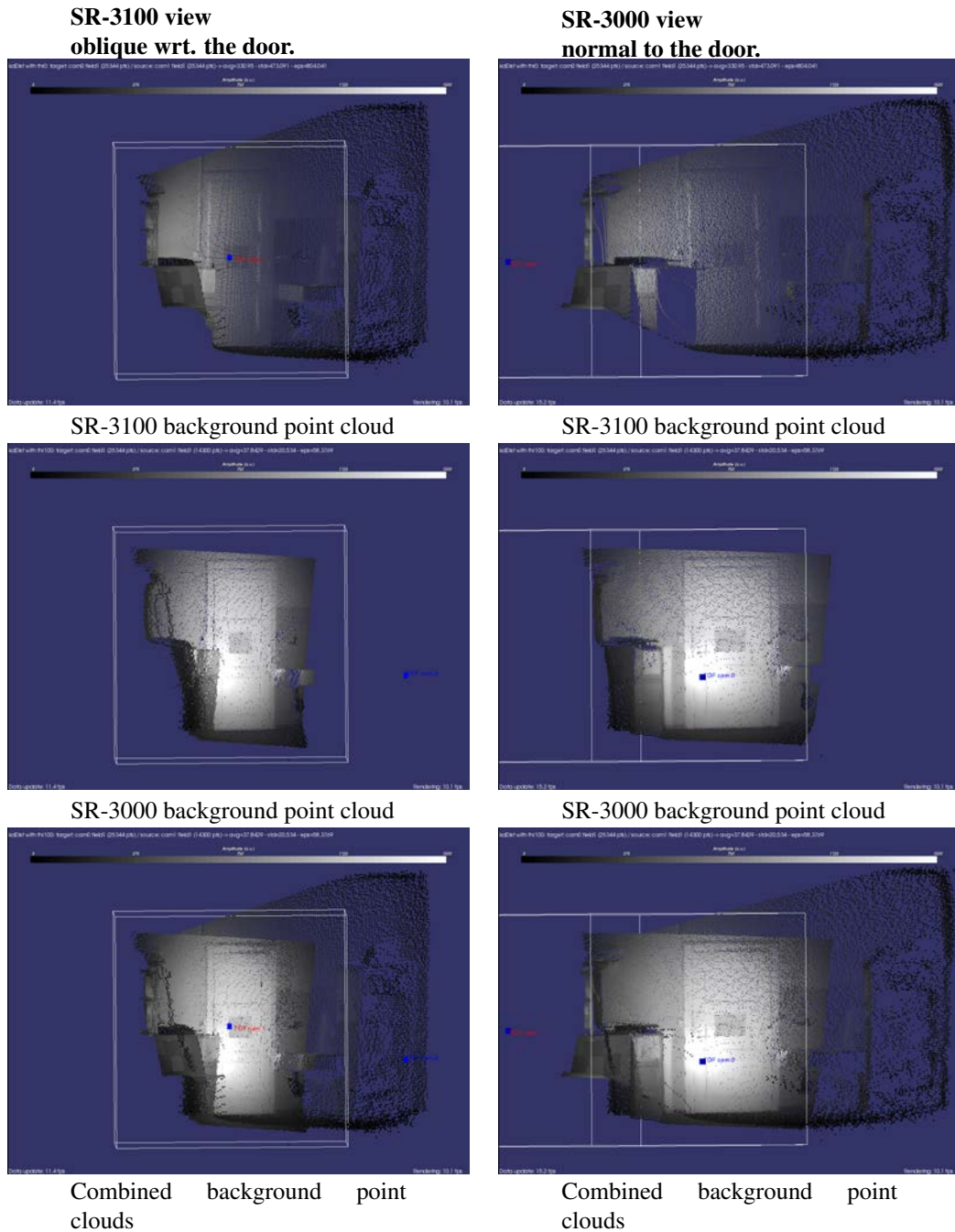


Figure 8.3: Access control: door scene. Empty scene. The SR-3100 has an oblique view to the door, while the SR-3000 is directly in front of the door.

Chapter 8. Applications

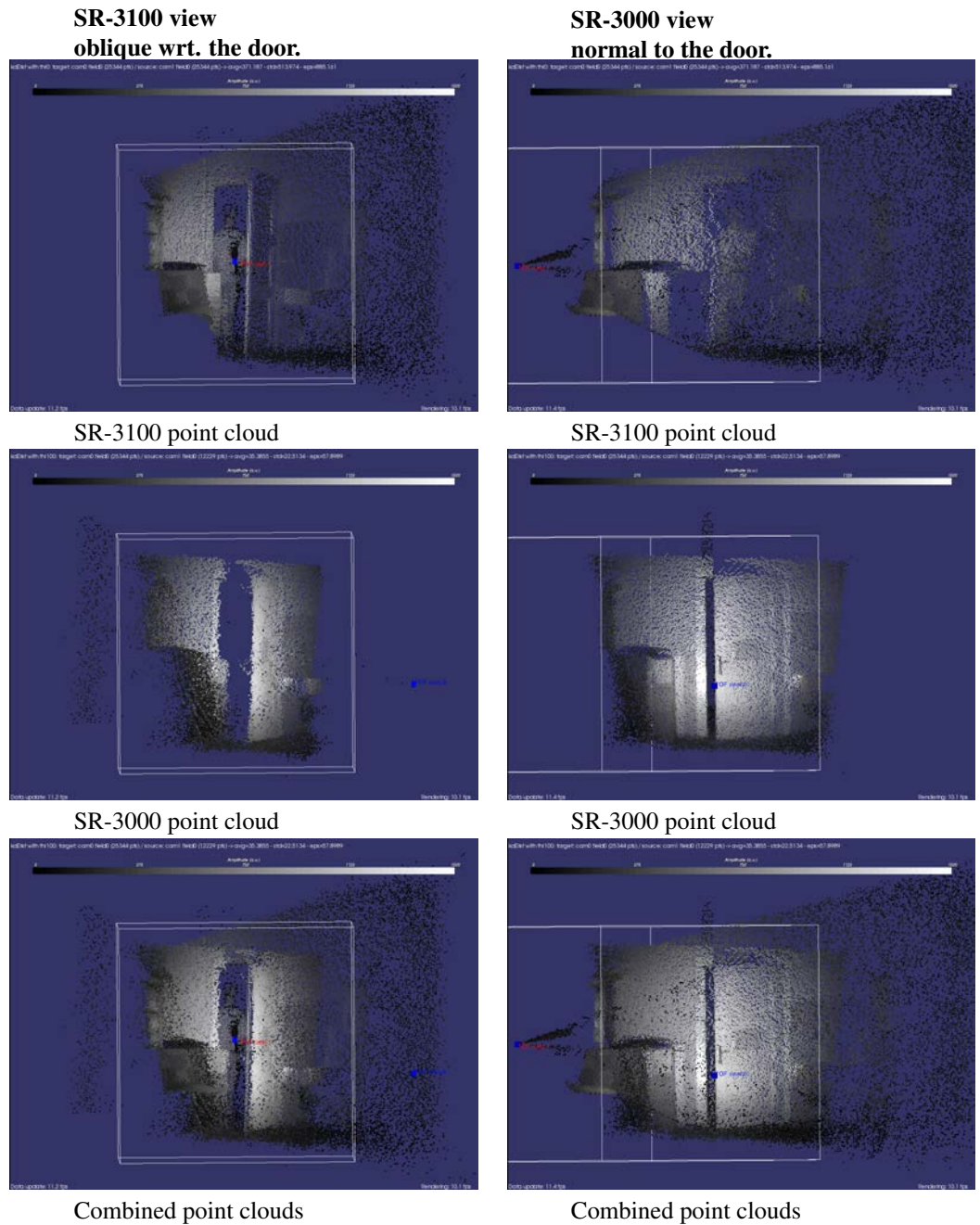


Figure 8.4: Access control: door scene. The door is being opened. The person opening the door is occluded by the door in the SR-3000 view, but can be seen in the SR-3100 view.

8.3. Occlusion removal

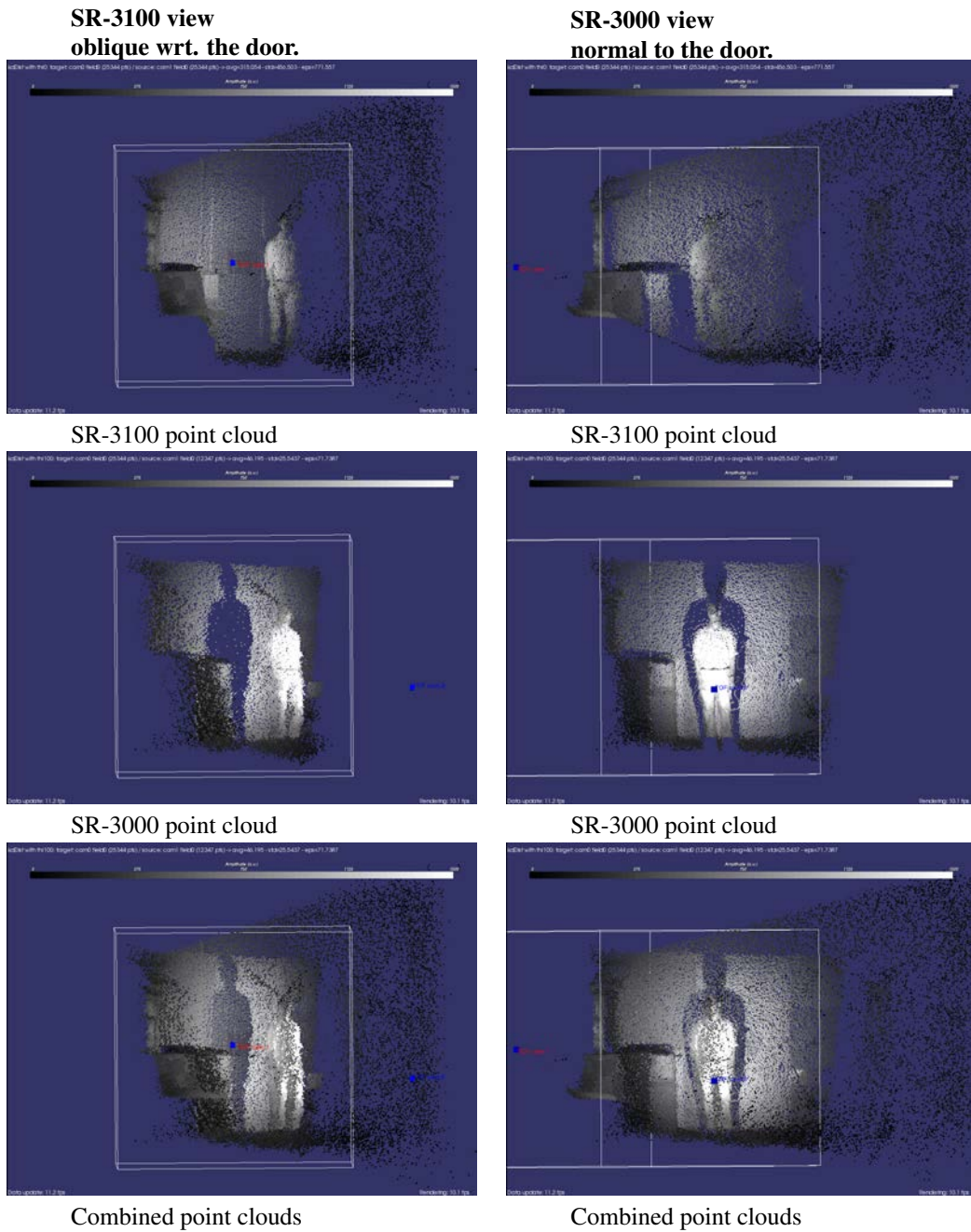


Figure 8.5:]

Access control: door scene. A person stands in front of the door. The person occludes a part of the door in the SR-3000 view, but this area is imaged in the SR-3100 view.

Chapter 8. Applications

human presence in a safety zone around the robot. However, a single viewpoint range imaging system, such as a single TOF camera, isn't sufficient when occlusions occur. Figure 8.6 illustrates the scene considered for our example. In the bottom row images, a cardboard materializes a safety zone. Any human entering this zone should trigger the robot's emergency shutdown. Figure 8.6**b,c,e,f** shows different views of the merged point cloud obtained from range images shot with two TOF cameras. In those images, the projection of the safety zone onto each camera sensor is illustrated by orange lines. When each camera is operated individually, an alarm must be sent for each situation where an object enters the zone delimited by these lines. In figure 8.7, another object occludes the safety zone for rightmost camera (see fig. 8.7**a,d**). This situation would result in a false alarm.

When combining the signal from both cameras, this occlusion doesn't cause a false alarm. The safety zone is now formed by the *intersection* of the projections for each camera.

8.4 Field of view extension

Multiple TOF cameras arranged in a network can be used to increase the system field of view : the merged point cloud in fig. 8.7 describes a larger scene than what could be obtained with a single TOF camera. Figures 8.8 and 8.9 present another example in the context of access control. A door allowing access into a secure zone must be monitored by a TOF system. The maximal distance from the camera to the door is constrained to be less than 3 meters. In that case, a single TOF camera cannot provide an image of the full door frame. With a conventional CCD camera, the optics could be changed to a fish-eye objective to extend the field of view. But active devices such as TOF camera cannot be modified so easily. In that case, it is necessary to use two cameras to monitor the door. Figure 8.8 shows the empty door frame, with the door partially open. This static scene allowed to perform time averaging on 50 frames. Note that the field of view of the two TOF cameras is set so that the full door frame is visible at all times, thus preventing an unauthorized person from sneaking-in unnoticed. In figure 8.9, a person having entered through the door is visible. Combining both TOF point clouds allows to image the person entirely, when only partial views of upper body, respectively legs are visible in individual point clouds.

8.5 Conclusion

Although the study of TOF camera network applications was limited by practical constraints (only two TOF cameras were available for the measurements), the experiments performed allow to validate this concept. Examples in the context of access control and human safety were provided. The software developed can be used as a tool assisting the setup of a multi-camera system based on TOF cameras. The implementation enables fast registration of two TOF views. In situations where scattering degrades the performance of the TOF cameras, scattering compensation algorithms can be used. The examples provided illustrate occlusion removal and field of view extension by a TOF camera network.

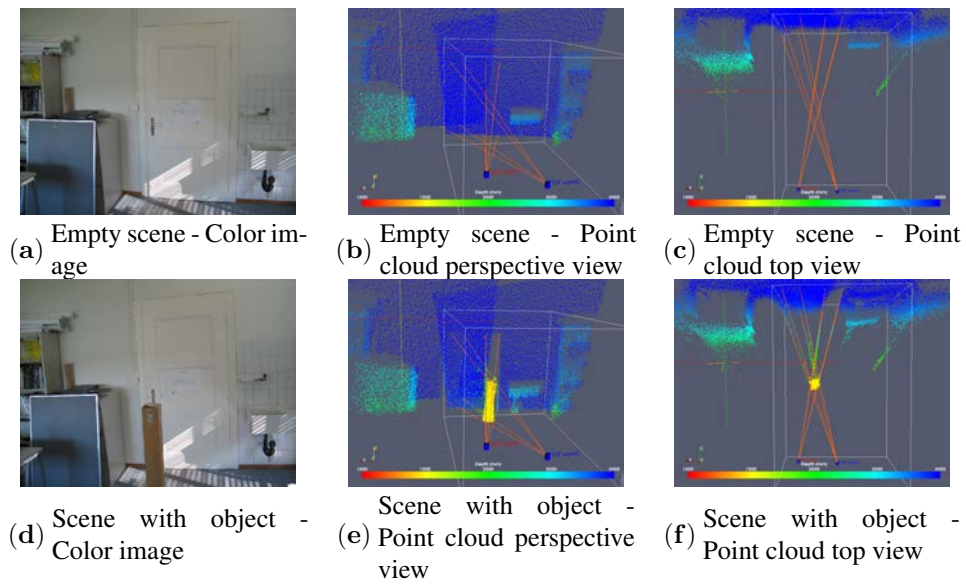


Figure 8.6: Test scene for occlusion removal. - The cardboard materializes a safety zone. Projection of the safety zone onto each camera sensor is displayed with orange lines in the point cloud representation.

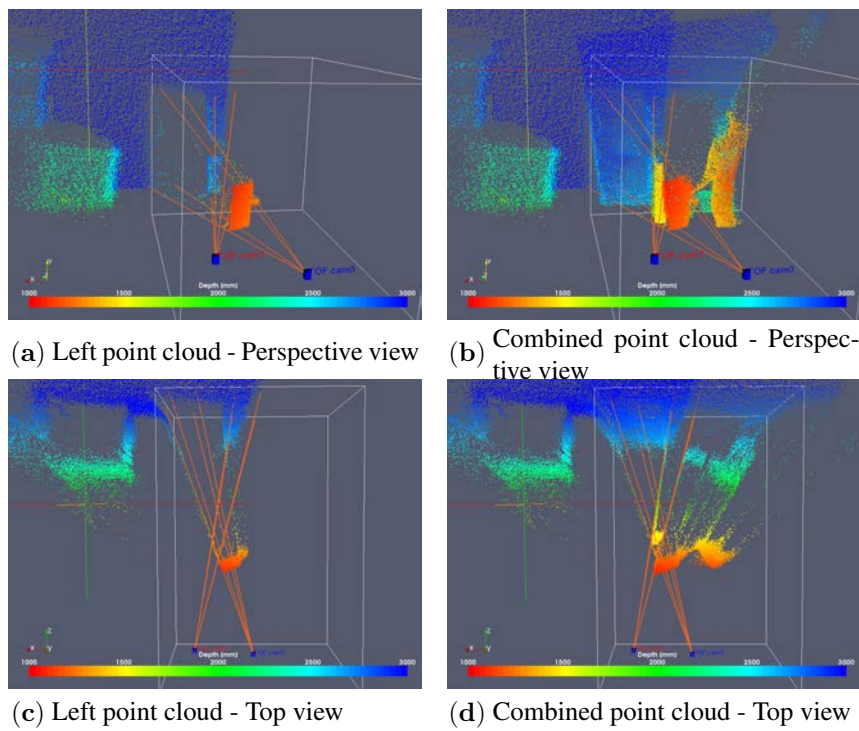


Figure 8.7: Occlusion removal - Target cardbox is occluded in left point cloud (a, c). Using two cameras allows to remove occlusion (b, d).

8.5. Conclusion

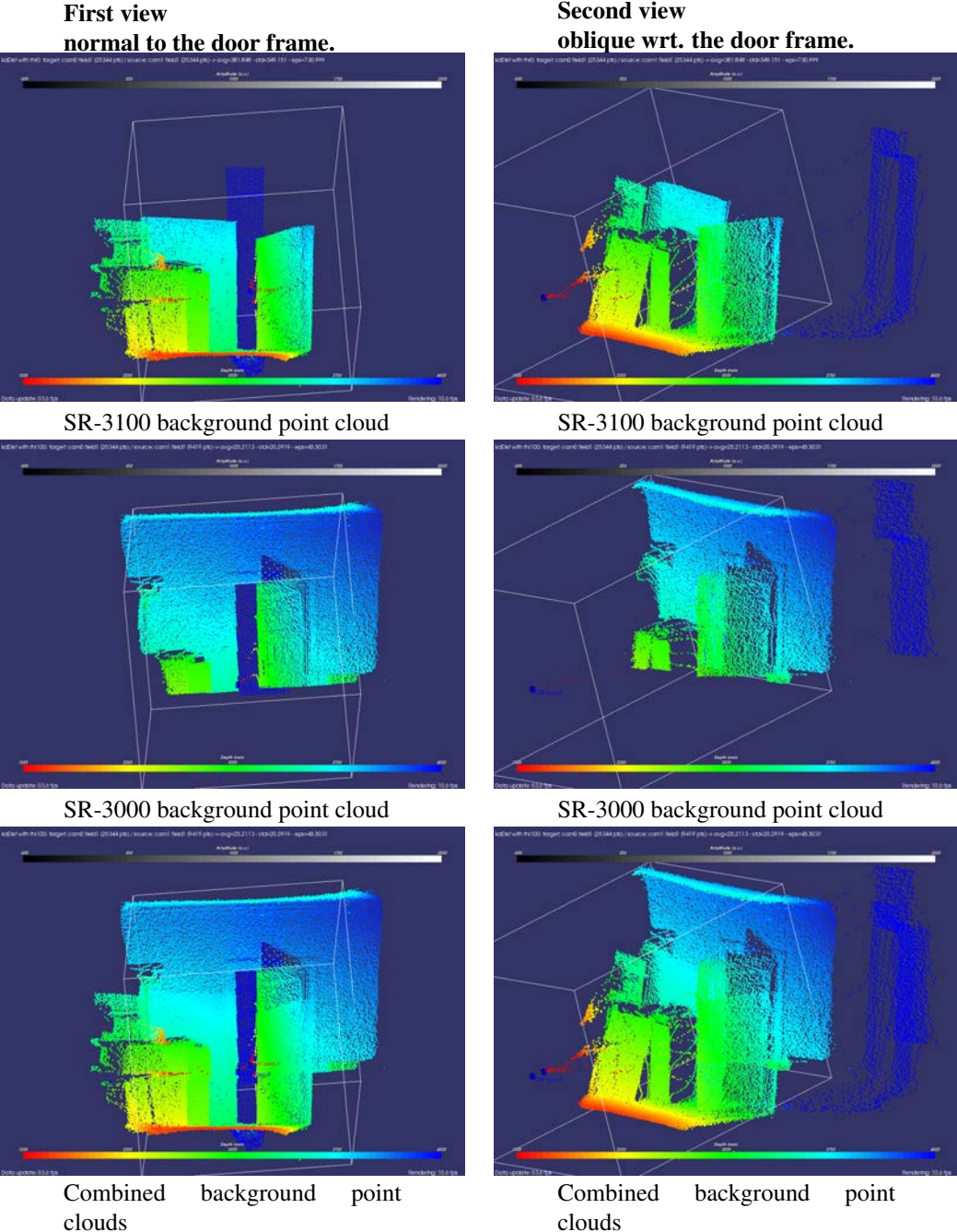


Figure 8.8: Access control: camera close to door. Averaged data for empty door frame. The SR-3100 is pointed towards the ceiling, while the SR-3000 is pointed towards the floor. The color represents depth measured from the SR-3100 camera.

Chapter 8. Applications

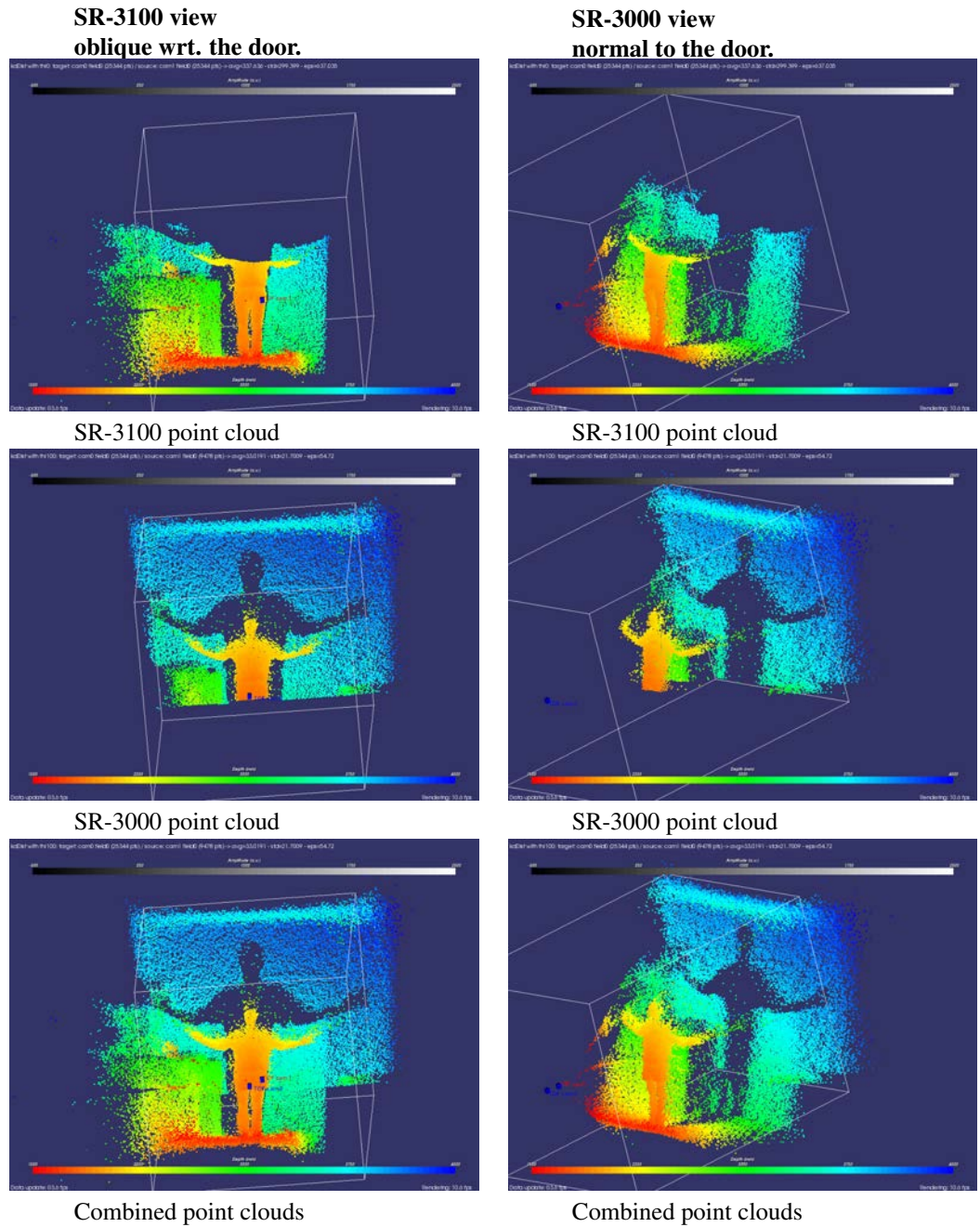


Figure 8.9: Access control: camera close to door. A person is moving in front of the door. The camera pointed towards the ceiling captures the upper body, the camera pointed towards the floor captures the legs. In the combined point cloud, the person is imaged from head to foot.

9

Conclusion

This thesis provides a contribution in the wide and steadily evolving field of 3D vision. Specifically, it considered improvements to two existing range imaging approaches: depth from focus microscopy and TOF cameras. The problem of vision for micro-assembly was studied, and the feasibility of an embedded depth from focus device for micro-assembly was analyzed. Concerning TOF cameras, the contributions of noise and deterministic error sources such as multipath and scattering in commercial systems were analyzed. A method for reduction of scattering related errors was proposed and tested. Finally, the problem of efficient registration of range images was studied with special focus on its application in surveillance systems based on TOF cameras.

9.1 Depth from focus device for micro-assembly

The first part of this thesis focused on the miniaturization potential of a depth from focus microscope device for micro-assembly. The principle of depth from focus range measurement was presented. The essential optical parameters limiting accuracy were discussed. A prototype miniature system was built and allowed to check the validity of the theoretical model experimentally, by comparing the range image accuracy for both the miniature and standard size 3D microscope systems. Although the range accuracy is lower for the miniature system (20 μ m for our prototype), it remains interesting for some assembly applications.

The current technological limitation in depth from focus miniaturization is camera motion: an image stack must be acquired while the camera is positioned at different altitudes relative to the scene, and the steps between camera positions must be precisely controlled. The motors used in the prototype described in this work are too massive to fit in an embedded system. A set of specifications for a micro-camera focus actuator was defined. But its implementation is still an open question.

The study finally showed that processing time is not a limitation for depth from focus imaging. On modern computers, high resolution image stacks can be processed in a few seconds. If the view is restricted to a 256×256 region of interest, continuous

Chapter 9. Conclusion

operation at 10 Hz can be achieved.

Perspective Although the presented study focused on miniaturization aspects, we identified other fields [83] where image processing can contribute to improve depth from focus imaging. One aspect is noise reduction in 3D microscopy. While currently, the sets of parameters used in this process have to be defined manually for each scene image, adaptive algorithms could be developed to automatically reduce noise, using an objective function based on the smoothness of the filtered data. The critical point in this development is to enable the smoothing of large noise fluctuations, while keeping the high spatial frequency information in regions with depth gaps.

9.2 TOF image error compensation

The principles of operations of current TOF cameras were presented. Noise and deterministic error sources caused by multipath and scattering were discussed and compared experimentally for SR-3000 and SR-3100 cameras. Scattering was identified as a major error source for continuous wave TOF cameras.

Regarding scattering, a mathematical model was proposed to describe its effects in TOF imagers: scattering was expressed by a point spread function (PSF) describing the parasitic optical couplings between different pixels. The PSF model was used to design scattering compensation algorithms involving deconvolution of the measured signal from the scattering PSF. Different implementation strategies were compared; timing constraints led to the selection of deconvolution by division in the Fourier domain as the preferred method for scattering compensation. This method is valid with an arbitrary scattering PSF; our experiments indicated that good results are obtained when the PSF is expressed as a sum of gaussian kernels. An ad-hoc optimization procedure was implemented to determine the best model parameters based on a segmented training scene. The scattering compensation implementation was tested on various real-world datasets, and showed in all cases a significant improvement. The improvements were most noticeable for SR-3000 cameras, where scattering is strongest. Provided that the scattering models are adapted, the same algorithm applies also to data produced by SR-3100 devices, where scattering is weakened by anti-reflective coatings inside the cameras; only the scattering model parameters have to be adapted for SR-3100 cameras. In many real-world data sets, errors related to scattering were reduced to the point where they became similar to errors caused by noise.

Perspective The determination of best performing scattering model parameters showed strong variations depending on the device used, but also on the specific scene imaged. This prevented the determination of an universal scattering model for a specific camera series. More complex models for scattering may be required. For example, it may be necessary to include a second scattering PSF describe light coupling between pixels which are quadratic with respect to the light intensity. Note nevertheless that the extent of the PSF found to provide best results indicate that scattering effects spread far away on the sensor, so that some of the information carried by the incoming light is irremediably lost when it falls outside of the light sensor area. Finally, we note that the determination of models parameters, which is currently based on extensive comparison of scattering compensation results, could be improved upon. For example, genetic algorithms could be used to speed up the search of best performing parameters.

9.3 Registration of TOF images

This part of the thesis considered TOF camera networks used for the purpose of improving 3D views by removing occlusions or extending the field of view. It aimed to define efficient point cloud registration strategies, with special focus on application in surveillance systems based on TOF cameras. Different approaches were compared. Standard stereo calibration proved too unreliable on low resolution TOF data. ICP methods appeared limited since they do not handle well configurations with small overlap between the views. A method based on intersections of plane primitives was also considered and tested, but the scenes it can be applied to are rare. Finally, an original alignment procedure was proposed, which we called master plane. The 6 degrees of freedom (DOF) alignment problem is split into two easier problems: 2 rotation (DOF) and one translation DOF are eliminated by matching a common plane region in the two point clouds; the remaining degrees of freedom are then eliminated by affine alignment in the common plane, based for instance on the intensity images for this region. This registration method provides convincing results for real-world TOF data sets. Typically, the alignment error of two views was found to be similar to the noise level of each single image.

Perspective In order to provide a complete solution for the calibration of TOF camera networks, the method must be automated, and its robustness must be improved. If the proposed master plane registration method is used, the affine alignment in the plane could be based on automated alignment techniques developed for 2D images, such as convolution or SIFT matching. Increased automation could also take the form of an algorithmic search of plane matches when multiple plane primitives are found in a scene.

Finally, while this work focused on plane geometric primitives and their combinations for registration, robustness could be improved by considering a larger variety of geometric primitives, such as spheres, cones and cylinders as alignment landmarks.



Depth from focus - Optics

This section presents the most prominent optical considerations to take into account when trying to develop a miniaturized depth from focus system corresponding to embedded systems specifications. The accuracy in depth determination will be at best of the same order of magnitude as depth of field.

A general rule in optics is that image quality increases with the the aperture of the system. In most applications, the aperture is limited by the diameter D of the primary lens. For simplicity, we consider single lens systems in the following discussions. Specificities of multiple lens systems are addressed only when expressly required.

A.1 Image formation and depth of field definition

The depth of field is defined as the maximum distance along the optical axis that a point can occupy such that the blur of its image does not exceed the size ϵ of one pixel. This is illustrated in figure A.1, which represents a single lens image formation system. The system is characterized by :

- f : lens focal length. The range of focal lengths required to image small objects is $0 < f \leq 25 \text{ mm}$.
- d_o : distance from the object to the lens. Generally, for inspection and assembly operations, $f < d_o \leq 5f$.
- ϵ : pixel size. Note that this dimension from the sensor plays a crucial role in the focus analysis, so that sensor and optics cannot be considered separately. Typically pixels measure a few micrometers: $2\mu\text{m} < \epsilon < 10\mu\text{m}$.
- D : entrance pupil diameter, or equivalently k : ratio of the focal length to the diameter (also called f-number $f\#$): $k = \frac{f}{D}$

The image formation condition relates object distance d_o to image distance d_i :

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f} \quad (\text{A.1})$$

A.1. Image formation and depth of field definition

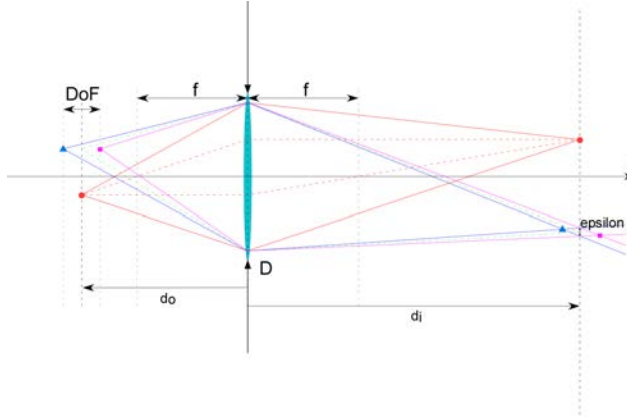


Figure A.1: Image formation system.

This condition must be satisfied to obtain the sharpest achievable image. In this case, the magnification M is determined by the ratio of object distance to image distance:

$$M = \frac{d_i}{d_o} \quad (\text{A.2})$$

The depth of field DoF can be expressed as:

$$DoF = \frac{d_o \cdot f^2 \cdot [2 \cdot \epsilon \cdot k \cdot (d_o - f)]}{f^4 - \epsilon^2 \cdot k^2 \cdot (d_o - f)^2} \quad (\text{A.3})$$

Only objects that lie within the depth of field will be correctly imaged by the vision system. This is a severe limitation for stereo-vision systems. At the same time, this condition can be exploited to reconstruct the 3D image of an object when scanned with a small depth of field (confocal microscopy, depth from focus).

A.1.1 Available magnifications for miniature prototype

As mentioned above, the single objective fitting for microscopic imaging with the Kappa CH-166 camera head has a focal length of 15mm. The standard position of this objective relative to the camera results in a magnification $M = 0.36$. Since higher magnification are desired, 3 special spacer elements were produced, to be added between the objective and imager. Adding a spacer element increases the image distance d_i ; since the focal length f is fixed, the image formation equation requires a reduction in object distance d_o . This results in an effective magnification increase, coming at the cost of a reduced working distance (between sample and objective). The attained magnifications with the 3 spacer elements are reported in table 3.5. The 4 discrete magnification values go from 0.36 to 1.85. Note that, the largest field of view is 6.7×5.1 mm. In the situation with highest magnification, the object distance d_o is relatively short: 40 mm. Higher magnifications would require to put the objective very close to the sample.

Appendix A. Depth from focus - Optics

Table A.1: Available magnifications for miniature imager prototype

Spacers	0	1	2	3
Field of view [mm]	6.7×5.1	3.4×2.6	2.2×1.7	1.3×1.0
Magnification	0.36	0.72	1.10	1.85

A.2 Relation between aperture and depth of field

In a miniature system, the weights and sizes of optical components such as lenses are limited. Equation A.3 can be rewritten to let the diameter D appear explicitly [125]:

$$DoF = \frac{2 \cdot (M + 1) \cdot \epsilon \cdot f \cdot D}{D^2 \cdot M^2 - \epsilon^2} \quad (\text{A.4})$$

Equation A.4 clearly shows that a short depth of field is obtained with short focal length, high magnification, and large entrance pupil diameter. For depth from focus depth measurement, we are interested in having the shortest depth of field. But the entrance pupil diameter is limited by weight considerations in a local sensor. Similarly, reducing the focal length will reduce working distance for the sensor. The curves in figure A.2 show predicted depth of field when D , ϵ and f are fixed, and the optical magnification M is varied to accommodate for different object size into the field of view X . In this example, the pixels size was set to $\epsilon = 3\mu\text{m}$ and the focal length to $f = 15\text{mm}$; the depth of field is plotted for two lens diameters, $D_1 = 7\text{mm}$ and $D_2 = 45\text{mm}$.

For a magnification of $0.3\times$ (corresponding to $9 \times 9\mu\text{m}$ pixel size projected in image plane), the depth of field is less than $200\mu\text{m}$ for the 7mm lens, whereas it would be lower than $30\mu\text{m}$ for a 45mm lens. This stresses the fact that using micro-cameras and micro-objectives increases the depth of field compared to standard sized components, for a given magnification. This effect induces a loss in performance for *depth from focus* systems, but could be an advantage in *stereo-vision* applications.

A.3 Relative depth of field

In many practical applications, the magnification is not specified beforehand, but is adapted to the field that one image must cover. A *relative* depth of field indicator DoF_{rel} is introduced to study the dependence of the depth of field DoF towards the lens diameter D . Note that this relative depth of field indicator involves the sensor size S , which defines the field lateral extension ($Field = \frac{S}{M}$).

$$DoF_{rel} = \frac{DoF}{Field} \quad (\text{A.5})$$

The relative depth of field has been plotted (fig A.3) for a sensor of size $S = 2.5\text{mm}$, which corresponds to the dimension in the CH-166 micro-camera. Analyzing figure A.3 allows us to draw three important conclusions:

- in order to achieve a small relative depth of field ($DoF_{rel} < 10^{-2}$), the magnification A must be high ($A > 3$);
- the slope of the curves decreases for $D > 10\text{mm}$, indicating that going from 7mm to 10mm diameter makes sense, but using lenses larger than 10mm should be considered with caution if weight requirements are severe;

A.3. Relative depth of field

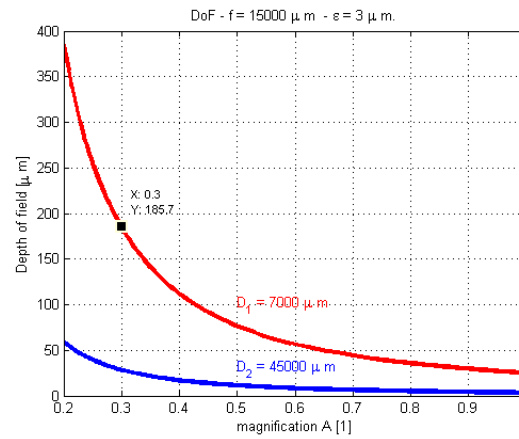


Figure A.2: Depth of field variation with magnification M .

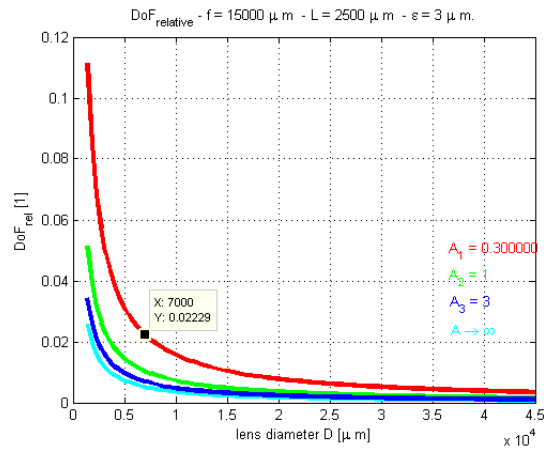


Figure A.3: Relative depth of field variation with lens diameter D , for different magnifications M .

Appendix A. Depth from focus - Optics

- in order to achieve $1\mu m$ vertical resolution within a $1 \times 1 mm$ field (i.e $DoF_{rel} < 10^{-3}$), a lens larger than $36 mm$ would be required, even for very high magnification (see fig. A.4).

A.3.0.1 Depth of field expectations for miniature system

As the depth from focus method requires the shortest possible depth of field, figure A.4 illustrates physical limitations expected for miniature systems. The pixel size can not go lower than $\epsilon = 3\mu m$ for video cameras. An objective with an aperture D larger than $10 mm$ would be incompatible with embedded system weight requirements. Besides, figure A.3 indicates that above $10 mm$ diameter, the effect of an increase in aperture is not highly effective. In theory, using objectives with a shorter focal length could allow to reduce greatly the depth of field (figure A.5). However, this would pose many practical problems due to the reduced working distance: in order to have $A \rightarrow \infty$, we must have $d_o = f$. A microscope objective has a working distance equal to its focal length, but microscopes require complicated (and therefore heavy) illumination schemes. These considerations lead us to the conclusion that it is not possible to fulfill all embedded system requirements with a depth from focus system.

Using the available objective of focal length $f = 15 mm$ and aperture $D = 7 mm$ for the CH-166 micro-camera, the highest realistic magnification is $M \approx 3$ (i.e $d_o \approx \frac{4}{3}f$). This configuration would allow to reach a relative depth of field $DoF_{rel} < 10^{-2}$, i.e. $10\mu m$ vertical resolution for a $1000\mu m$ field.

A.4 Depth of field - Experiments

A small depth of field is wished when the depth from focus method is used. The micro-camera used in the miniature system prototype has a larger depth of field than the reference macroscopic microscope system, as discussed in section A.1. Here we present some experimental results allowing to estimate the depth of field in the miniature system. Figure A.6 presents a subset¹ of an image stack acquired with Kappa CH-166 micro camera. The camera was moved $16 mm$ from first image I_0 (background in focus) to last image I_{511} (top in focus), resulting in a $3.13\mu m$ camera altitude difference between two consecutive images. The dimensions of the visible field (in focus object plane) are $7.52 \times 5.71 mm$. Observing the series of images, we estimate that the optical depth of field spans over approximately 64 images (fig A.7), which represents a depth of field of $200\mu m$. This is consistent with theoretical expectations (see fig. A.2). The value for the relative depth of field ($DoF_{rel} = \frac{DoF}{Field} = \frac{200\mu m}{7500\mu m} \approx 0.027$) is also in good agreement with theoretical expectations(see fig. A.3). As mentioned earlier, the only practical possibility to increase vertical resolution is to increase magnification A . Having $A = 3$ would allow to reach $1\mu m$ lateral resolution, while the depth of field would be reduced to $6\mu m$.

We can note here that the large z-range for this image stack causes severe telecentricity issues (compare visible field in I_0 and I_{511}).

¹17 images out of 511

A.4. Depth of field - Experiments

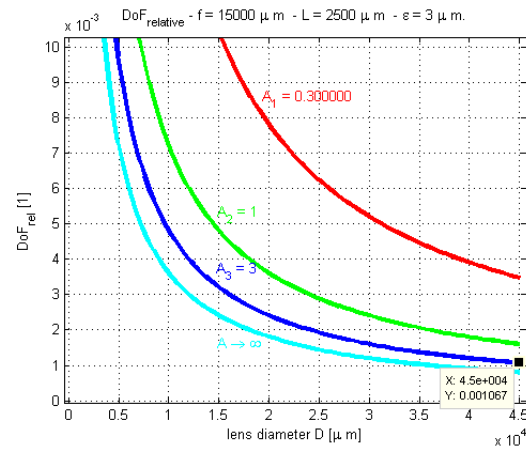


Figure A.4: Relative depth of field variation with lens diameter D , for different magnifications M ; detail.

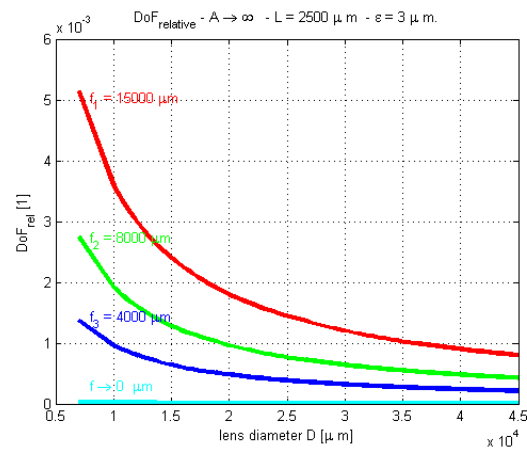


Figure A.5: Relative depth of field variation with lens diameter D for different focal lengths f .

Appendix A. Depth from focus - Optics

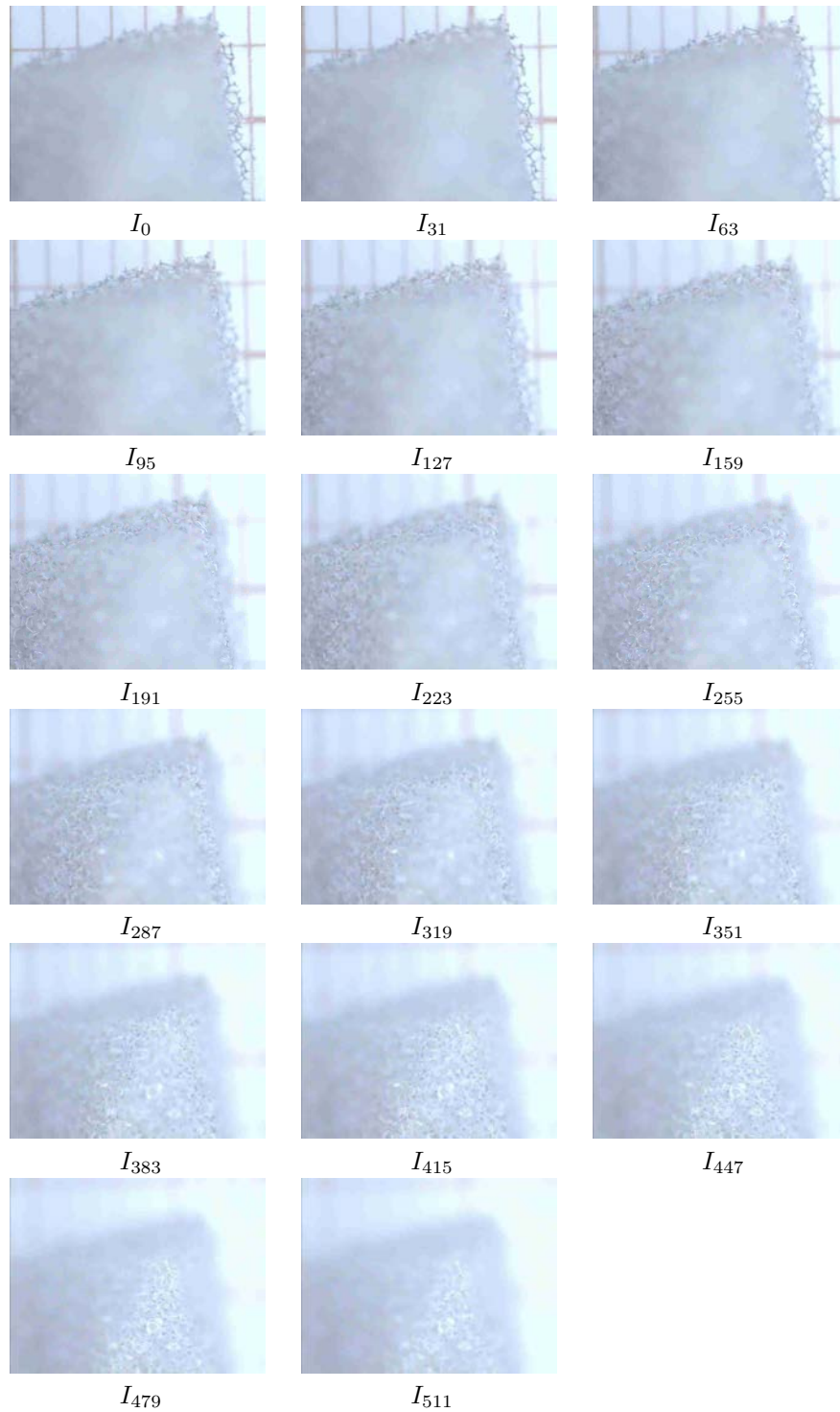


Figure A.6: Series of images with varying focus. Depth step : $3.13 \mu\text{m}$. Visible field : $7.52\text{mm} \times 5.71\text{mm}$. Depth of field: ≈ 64 steps, i.e. $200 \mu\text{m}$.

A.5. Telecentricity

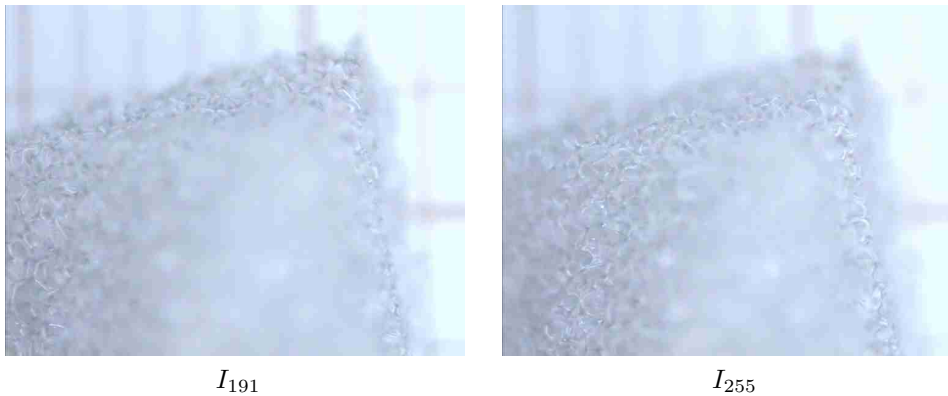


Figure A.7: Effect of $200\ \mu\text{m}$ camera displacement

A.5 Telecentricity

A.5.1 Magnification of blurred regions

Given the expected dimension of target objects in micro-assembly, telecentricity will be a major issue. This topic was only briefly touched in previous IMT works ([127],[53]), which were concerned with objects presenting a much smaller z-range. Figure A.8 illustrates the situation: when the object is moved relative to imaging system, a point P goes from P_1 to P_2 to P_3 . Its image goes from I_1 (blurry image on sensor) to I_2 (sharp image) and I_3 blurry image on sensor). The magnification is different for I_1, I_2 and I_3 . This effect can be observed in figure A.9, were the linear extension of the observable background between first and last image in the stack is close to 10%. This effect is far less problematic when the height of the imaged region is small (fig A.10)

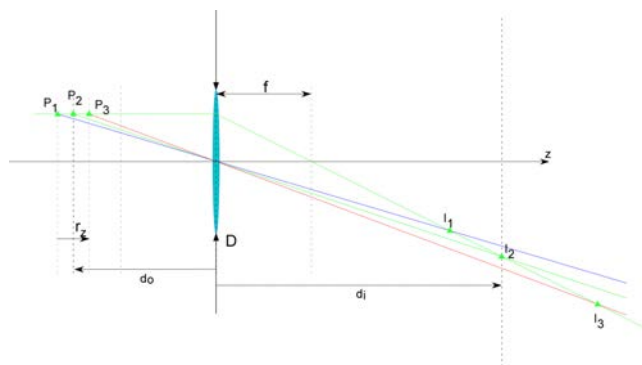
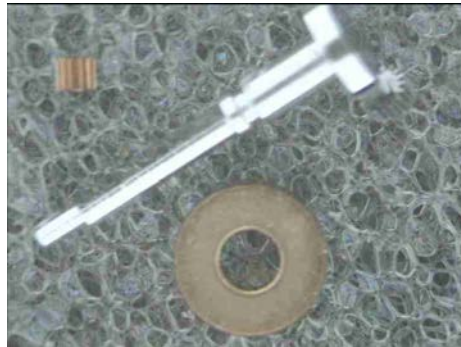
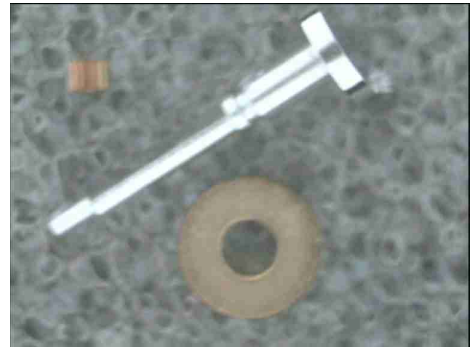


Figure A.8: Telecentricity: blurry regions experience varying magnification when object is moved.

Appendix A. Depth from focus - Optics

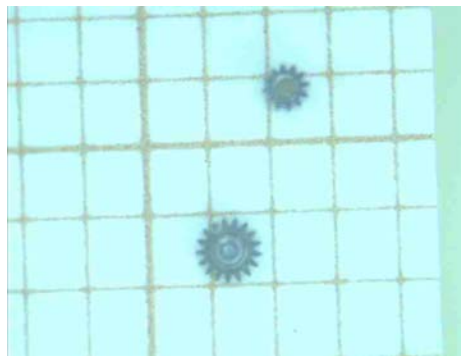


(First image)

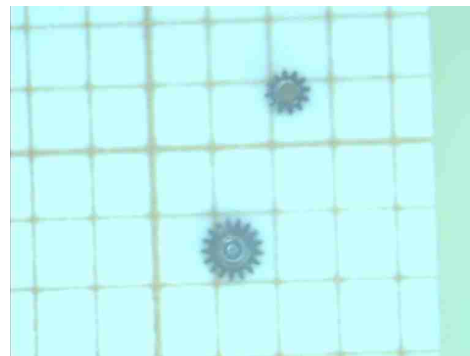


(Last image)

Figure A.9: First and last image in acquired image stack for large object (scene depth 6 mm). The width and the height of the (blurry) background were increased by approximately 10%.



(First image)



(Last image)

Figure A.10: First and last image in acquired image stack for small object (scene depth $< 2.5\text{ mm}$).

A.5.2 Telecentric objectives

Telecentric objectives are commonly used in machine vision applications. Such objectives include a pinhole positioned at the focal point on the image side, so that only light rays parallel to the optical axis are transmitted (fig A.11), effectively destroying the telecentricity effect. Those objectives have two major drawbacks: the intensity of light available at the sensor is greatly reduced, and the maximum dimension of the field is restricted to the lens diameter. Moreover, no telecentric objective exists for Kappa micro-cameras. A custom objective telecentric objective could probably be designed, but this exceeds the scope of this work. Such a design would require to establish a precise set of constraints regarding the objects to be imaged, in order to define design rules for all objective parameters (focal length f , f-number k , working distance d_o , telecentric selectivity, etc.).

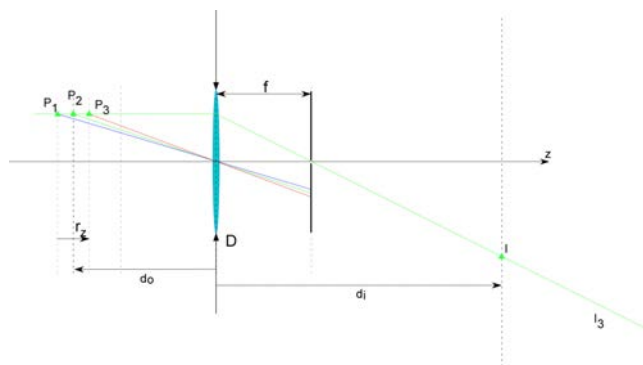


Figure A.11: Telecentric objective: rays that are not parallel to the optical axis are blocked.

B

Bibliography

B.1 Publications list

- James Mure-Dubois & Heinz Hügli, October 2008. **Fusion of Time-of-Flight Camera Point Clouds**. *In: Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*, Marseille 2008. European Conference on Computer Vision 2008. [link](#) [link](#) .
 - James Mure-Dubois & Heinz Hügli, August 2008. **Merging of range images for inspection or safety applications**. *Pages 70660K 1–12 of: Two- and Three-Dimensional Methods for Inspection and Metrology VI*, San Diego. Proc. SPIE, vol. 7066. DOI: [10.1117/12.793629](#) [link](#) .
 - James Mure-Dubois & Heinz Hügli, April 2008. **Automated inspection of microlens arrays**. *Pages 700007 1–9 of: Optical and Digital Image Processing*, Strasbourg. Proc. SPIE, vol. 7000. DOI: [10.1117/12.781015](#) [link](#) .
 - James Mure-Dubois & Heinz Hügli, September 2007. **Optimized scattering compensation for time-of-flight camera**. *Pages 67620H 1–11 of: Two- and Three-Dimensional Methods for Inspection and Metrology V*, San Diego. Proc. SPIE, vol. 6762. DOI: [10.1117/12.733961](#) [link](#) [link](#) .
 - James Mure-Dubois & Heinz Hügli, July 2007. **Time-of-flight imaging of indoor scenes with scattering compensation**. *Pages 117–123 of: Proc. O3D 2008*, Zürich. 8th Conference on Optical 3D Measurement Techniques. [link](#) [link](#) .
 - James Mure-Dubois & Heinz Hügli, March 2007. **Real-time scattering compensation for time-of-flight camera**. *In: Proc. of the ICVS 2007*, Bielefeld. International Conference on Computer Vision Systems 2007. DOI: [10.2390/biecoll-icvs2007-167](#) [link](#) [link](#) .
-

B.1. Publications list

- James Mure-Dubois & Heinz Hügli, October 2006. **Embedded 3D vision system for automated micro-assembly**. *Pages 63820J 1–10 of: Two- and Three-Dimensional Methods for Inspection and Metrology IV*, Boston. Proc. SPIE, vol. 6382. DOI: [10.1117/12.686675](https://doi.org/10.1117/12.686675) [link](#) .
- Heinz Hügli & James Mure-Dubois, October 2006. **3D vision methods and selected experiences in micro and macro applications**. *Pages 638209 1–11 of: Two- and Three-Dimensional Methods for Inspection and Metrology IV*, Boston. Proc. SPIE, vol. 6382. DOI: [10.1117/12.693635](https://doi.org/10.1117/12.693635) [link](#) [link](#) .

B.2 References

- [1] 3DV systems. 2006. *zCam*. <http://www.3dvsystems.com/> (accessed on 30.09.2007). 14, 16, 17, 20
- [2] Aguet, F., Van De Ville, D., & Unser, M. 2008. Model-Based 2.5-D Deconvolution for Extended Depth of Field in Brightfield Microscopy. *Image Processing, IEEE Transactions on*, **17**(7), 1144–1153. DOI: [10.1109/TIP.2008.924393](https://doi.org/10.1109/TIP.2008.924393) [link](#) . 11
- [3] Arun, K. S., Huang, T. S., & Blostein, S. D. 1987. Least-squares fitting of two 3-D point sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, **9**(5), 698–700. 92, 93
- [4] Asyri SA. 2009. *DesktopDelta*. <http://www.asyri.ch/pages/products/desktopdelta.php>. 27
- [5] Atos GmbH. 2005. *PLμ Confocal Imaging Profiler*. URL <http://www.atos-online.de/PLmu-C-k.pdf> (accessed on 17.05.2005). 4
- [6] Bae, K., Belton, D., & Lichti, D. 2008. A Closed-Form Expression of the Positional Uncertainty for 3D Point Clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PP**, 1–1. DOI: [10.1109/TPAMI.2008.116](https://doi.org/10.1109/TPAMI.2008.116) . 24
- [7] Bergevin, R., Soucy, M., Gagnon, H., & Laurendeau, D. 1996. Towards a general multi-view registration technique. *IEEE Trans. Pattern Anal. Mach. Intell.*, **18**(5), 540–547. DOI: [10.1109/34.494643](https://doi.org/10.1109/34.494643) [link](#) . 89
- [8] Bert, J., Dembélé, S., & Lefort-Piat, N. 2006a. Synthesizing a Virtual Imager with a Large Field of View and a High Resolution for Micromanipulation. In: *Proc. 5th Int. Workshop on MicroFactories, IWMF'06*. [link](#) . 27
- [9] Bert, J., Dembélé, S., & Lefort-Piat, N. 2006b. Trifocal Transfer Based Novel View Synthesis for Micromanipulation. *Pages 411–420 of: Proc. Int. Symposium on Visual Computing (ISVC)*. LNCS, vol. 4291. DOI: [10.1007/11919476](https://doi.org/10.1007/11919476) [link](#) . 5
- [10] Bert, J., Dembélé, S., & Lefort-Piat, N. 2006c. Virtual Camera Synthesis for Micromanipulation and Microassembly. *Pages 1390–1395 of: Proc. Int. Conf. on Intelligent Robots and Systems, IEEE/RSJ 2006*. DOI: [10.1109/IROS.2006.281928](https://doi.org/10.1109/IROS.2006.281928) . 27, 28
- [11] Bert, J., Dembélé, S., & Lefort-Piat, N. 2007. Performing Weak Calibration at the Microscale, Application to Micromanipulation. *Pages 4937–4942 of: IEEE Int. Conf. on Robotics and Automation, ICRA'2007*. [link](#) . 5, 7, 24
- [12] Besl, Paul J., & McKay, Neil D. 1992. A Method for Registration of 3-D Shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, **14**(2), 239–256. DOI: [10.1109/34.121791](https://doi.org/10.1109/34.121791) [link](#) . 24, 94
- [13] Bieri, L.-S., & Jacot, J. 2004. Three dimensionnal vision using structured light applied to quality control in production line. *Pages 463–471 of: W. Osten, M. Takeda (ed), Optical Metrology in Production Engineering*. Proc. SPIE, vol. 5457. DOI: [10.1117/12.545039](https://doi.org/10.1117/12.545039) [link](#) . 4, 5

B.2. References

- [14] Bieri, Louis-Severin, & Jacot, Jacques. 2004 (Sept.). Three dimensionnal vision using strucuted light applied to quality control in production line. *Pages 463–471 of: Proceedings of SPIE - Optical Metrology in Production Engineering*, vol. 5457. 6
- [15] Bohme, M., Haker, M., Martinetz, T., & Barth, E. 2008. Shading constraint improves accuracy of time-of-flight measurements. *Pages 1–6 of: Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*. DOI: [10.1109/CVPRW.2008.4563157](https://doi.org/10.1109/CVPRW.2008.4563157) link . 22
- [16] Boissenin, M., Wedekind, J., Selvan, A.N., Amavasai, B.P., Caparrelli, F., & Travis, J.R. 2006. Computer vision methods for optical microscopes. *Image and Vision Computing*, **25**(Oct.), 1107–1116. DOI: [10.1016/j.imavis.2006.03.009](https://doi.org/10.1016/j.imavis.2006.03.009) . 11
- [17] Bouguet, Jean-Yves. 2004. *Camera Calibration Toolbox for Matlab*. http://www.vision.caltech.edu/bouguetj/calib_doc/ (accessed on 10.09.2007). 19, 24, 91, 116, 118
- [18] Bourquin, S., Seitz, P., & Salathé, R. P. 2001. Optical coherence topography based on a two-dimensional smart detector array. *Opt. Lett.*, **26**(8), 512–514. DOI: [10.1364/OL.26.000512](https://doi.org/10.1364/OL.26.000512) . 7, 8
- [19] Büttgen, Bernhard. 2007. *Extending time-of-flight optical 3D-imaging to extreme operating conditions*. PhD Dissertation, Université de Neuchâtel. 17, 22, 48, 49, 51
- [20] Canesta Inc. 2006. *CanestaVision*. <http://www.canesta.com/index.htm>. 13, 17, 20
- [21] Chen, Y., & Medioni, G. 1991. Object modeling by registration of multiple range images. *Pages 2724–2729 of: Proc. of the 1991 IEEE Intl. Conf. on Robotics and Automation*. DOI: [10.1109/ROBOT.1991.132043](https://doi.org/10.1109/ROBOT.1991.132043) link . 24, 25, 94
- [22] Chetverikov, D., Svirko, D., Stepanov, D., & Krsek, Pavel. 2002. The Trimmed Iterative Closest Point Algorithm. *Pages 545–548 of: In International Conference on Pattern Recognition*. DOI: [10.1109/ICPR.2002.1047997](https://doi.org/10.1109/ICPR.2002.1047997) link . 25, 96
- [23] Codourey, A., Rodriguez, M., & Pappas, I. 1997. A Task-oriented Teleoperation System for Assembly in the Microworld. *Pages 235–240 of: Int. Conf. on Advanced Robotics*. DOI: [10.1109/ICAR.1997.620188](https://doi.org/10.1109/ICAR.1997.620188) link . 27
- [24] CSEM. 2006. *SwissRanger*. <http://www.swissranger.ch>. 20, 61
- [25] Danuzer, G. 1996 (July). Stereo Light Microscope Calibration for 3D Submicron Vision. *Pages 101–108 of: Proceeding of the 18th ISPRS Congress*, vol. 31/B5. link . 5, 6
- [26] Danuzer, Gaundez. 1997. *Quantitative stereo vision for the stereo light microscope : an attempt to provide control feedback for a nanorobot system*. Ph.D. thesis, ETH Zurich. [ETH Diss. 22535](https://doi.org/10.1109/ETHD.1997.22535). 5
- [27] Du, Huan, Oggier, Thierry, Lustenberger, Felix, & Charbon, Edoardo. 2005. A Virtual Keyboard Based on True-3D Optical Ranging. *Pages 220 – 229 of: British Machine Vision Conference 2005*, vol. 1. link . 18

Appendix B. Bibliography

- [28] Ens, John E. 1990. *An investigation of methods for determining depth from focus*. Ph.D. thesis, University of British Columbia. 9
- [29] Estana, Ramon. 2006. *MICRON: Miniaturised co-operative Robots advancing towards the Nano Range*. Tech. rept. University of Karlsruhe, Germany. IST2001-33567 http://www.wipr.ira.uka.de/seyfried/MiCRoN/PublicReport_Final.pdf. 11, 12
- [30] Falcao, G., Hurtos, N., Massich, J., & Fofi, D. 2009. *Projector-Camera Calibration Toolbox*. <http://code.google.com/p/procamcalib>. 19
- [31] FARO Technologies Inc. 2009. *Laser Scanner Photon*. <http://laser-scanner.faro.com/us/laser-scanner-photon/> (accessed on 29.06.2009). 18, 20
- [32] Fatikow, S., Seyfried, J., Fahlbusch, S., Buerkle, A., & Schmoeckel, F. 2000. A Flexible Microrobot-Based Microassembly Station. *Journal of Intelligent and Robotic Systems*, **27**, 135–169. DOI: 10.1023/A:1008106331459 [link](#) . 27
- [33] Faugeras, O., & Hébert, M. 1983 (Aug.). A 3D recognition and positioning algorithm using geometrical matching between primitive surfaces. *Pages 996–1002 of: Proc. 7th Intl Joint Conf. on Artificial Intelligence*, vol. PR01991. [link](#) . 99, 100, 101
- [34] Ferreira, A., Fontaine, J.-G., & Hirai, S. 2002. Automation of a Teleoperated Microassembly Desktop Station Supervised by Virtual Reality. *Transactions on Control, Automation, and Systems Engineering*, **4**(1), 2519–2535. [link](#) . 27
- [35] Fitzgibbon, Andrew W. 2003. Robust Registration of 2D and 3D Point Sets. *Image and Vision Computing*, **21**(Dec.), 1145–1153. DOI: 10.1016/j.imavis.2003.09.004 [link](#) [link](#) . 25
- [36] Friedman, Jerome H., Bentley, Jon L., & Finkel, Raphael A. 1977. An Algorithm for Finding Best Matches in Logarithmic Expected Time. *ACM Trans. Math. Softw.*, September, 209–226. DOI: 10.1145/355744.355745 [link](#) . 24
- [37] Frigo, M., & Johnson, S G. 2007. *FFTW*. <http://www.fftw.org/>. 80
- [38] Frigo, Matteo, & Johnson, Steven G. 2005. The Design and Implementation of FFTW3. *Proceedings of the IEEE*, **93**(2), 216–231. special issue on Program Generation, Optimization, and Platform Adaptation. DOI: 10.1109/JPROC.2004.840301 [link](#) . 82
- [39] Fuchs, Stefan, & May, Stefan. 2008. Calibration and registration for precise surface reconstruction with Time-Of-Flight cameras. *Int. J. Intell. Syst. Technol. Appl.*, **5**(3/4), 274–284. DOI: 10.1504/IJISTA.2008.021290 . 18
- [40] Geissler, P., & Dierig, T. 1999. *Depth-from-Focus*. Handbook of Computer Vision and Applications. San Diego Academic Press. 9
- [41] Gelfand, N., Ikemoto, L., Rusinkiewicz, S., & Levoy, M. 2003. Geometrically stable sampling for the ICP algorithm. *Pages 260–267 of: 3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on.* [link](#) . 97
- [42] Guan, L., & Pollefeys, M. 2008. A Unified Approach to Calibrate a Network of Camcorders and ToF Cameras. *In: MMSFAA*. [link](#) . 18, 25

B.2. References

- [43] Guomundsson, S. A., Aanæs, H., & Larsen, R. 2007 (July). Environmental Effects on Measurement Uncertainties of Time-of-Flight Cameras. *Pages 1–4 of: International Symposium on Signals Circuits and Systems - ISSCS*, vol. 1. DOI: [10.1109/ISSCS.2007.4292664](https://doi.org/10.1109/ISSCS.2007.4292664) link . 63, 64
- [44] Gvili, Ronen, Kaplan, Amir, Ofek, Eyal, & Yahav, Giora. 2003. Depth keying. *Pages 564–574 of: Woods, Andrew J., Bolas, Mark T., Merritt, John O., & Benton, Stephen A. (eds), Stereoscopic Displays and Virtual Reality Systems X*, vol. 5006. SPIE. DOI: [10.1117/12.474052](https://doi.org/10.1117/12.474052) link . 16, 18
- [45] Ha, T., Takaya, Y., Miyoshi, T., Ishizuka, S., & Suzuki, T. 2004. High-precision on-machine 3D shape measurement using hypersurface calibration method. *Pages 40–50 of: Machine Vision and its Optomechatronic Applications*. Proc. SPIE, vol. 5603. DOI: [10.1117/12.571001](https://doi.org/10.1117/12.571001) . 4
- [46] Hansen, D.W., Hansen, M.S., Kirschmeyer, M., Larsen, R., & Silvestre, D. 2008 (June). Cluster tracking with Time-of-Flight cameras. *In: Computer Vision and Pattern Recognition (CVPR) Workshops*. DOI: [10.1109/CVPRW.2008.4563156](https://doi.org/10.1109/CVPRW.2008.4563156) . 18
- [47] Harding, Kevin, & Hu, Qingying. 2006. Multiresolution 3D measurement using a hybrid fringe projection and moire approach. *Pages 63820K+ of: Huang, Peisen S. (ed), Two- and Three-Dimensional Methods for Inspection and Metrology IV*, vol. 6382. SPIE. 5
- [48] Heliotis AG. 2005. *Heliotis real-time 3D imaging*. <http://www.heliotis.ch/>. 8, 9, 10
- [49] Hesselbach, J., & Pokar, G. 2000. Assembly of a miniature linear actuator using vision feedback. *Pages 13–20 of: Microrobotics and Microassembly II*. Proc. SPIE, vol. 4194. DOI: [10.1117/12.403698](https://doi.org/10.1117/12.403698) . 26
- [50] Hollis, R., & Gowdy, J. 1998. Miniature Factories for Precision Assembly. *Pages 9–14 of: International Workshop on Micro-Factories*. link . 27, 42
- [51] Huang, Peisen S., & Han, Xu. 2006. On improving the accuracy of structured light systems. *Pages 63820H+ of: Huang, Peisen S. (ed), Two- and Three-Dimensional Methods for Inspection and Metrology IV*, vol. 6382. SPIE. DOI: [10.1117/12.692631](https://doi.org/10.1117/12.692631) . 22
- [52] Hügli, H., & Mure-Dubois, J. 2006. 3D vision methods and selected experiences in micro and macro applications. *Pages 638209 1–11 of: Two- and Three-Dimensional Methods for Inspection and Metrology IV*. Proc. SPIE, vol. 6382. SPIE. DOI: [10.1117/12.693635](https://doi.org/10.1117/12.693635) link link .
- [53] Hüsser, Olivier. 2003 (February). *Traitement d'image basé sur la notion de mise au point*. Tech. rept. IMT 443 HU 02/03. Institut de Microtechnique, Université de Neuchâtel. Available on IMT/Orange: [IMT443_Report](#). 149
- [54] Iddan, G. J., & Yahav, G. 2001. Three-dimensional imaging in the studio and elsewhere. *Pages 48–55 of: Corner, Brian D., Nurre, Joseph H., & Pargas, Roy P. (eds), Three-Dimensional Image Capture and Applications IV*, vol. 4298. SPIE. DOI: [10.1117/12.424913](https://doi.org/10.1117/12.424913) . 15, 18

Appendix B. Bibliography

- [55] Intel Inc. 2000. *Intel Image Processing Library 2.5*. <http://developer.intel.com/software/products/perflib/>. 38, 39, 82
- [56] Jensen, R.R., Paulsen, R.R., , & Larsen, R. 2009. Analyzing Gait Using a Time-of-Flight Camera. In: *Proc. of the 16th Scandinavian Conference on Image Analysis (SCIA)*. DOI: [10.1007/978-3-642-02230-2_3](https://doi.org/10.1007/978-3-642-02230-2_3) . 18
- [57] Jost, T., & Hügli, H. 2002. A multi-resolution scheme ICP algorithm for fast shape registration. *Pages 540–543 of: Proc First Intl. Symp. on 3D Data Processing Visualization and Transmission*. DOI: [10.1109/TDPVT.2002.1024114](https://doi.org/10.1109/TDPVT.2002.1024114) link . 24
- [58] Jost, T., & Hügli, H. 2004 (Oct.). A Multi-Resolution ICP with Heuristic Closest Point Search for Fast and Robust 3D Registration of Range Images. *Pages 427–433 of: Proc. 3Dim2003, 4th Int. Conf. on 3-D Digital Imaging and Modeling*, vol. PR01991. DOI: [10.1109/IM.2003.1240278](https://doi.org/10.1109/IM.2003.1240278) link . 24
- [59] Kahlmann, T., & Ingensand, H. 2007 (July). Increased accuracy of 3D range imaging camera by means of calibration. *Pages 101–108 of: Proc. of the 8th Conference on Optical 3D Measurements*. 17, 22, 91
- [60] Kahlmann, T., Remondino, F., & Ingensand, H. 2006. Calibration for Increased Accuracy of the Range Imaging Camera SwissRanger. *Pages 136–141 of: Proc. of the ISPRS Com. V Symposium*. ISPRS. 18
- [61] Kahlmann, Timo. 2007. *Range imaging metrology : investigation, calibration and development*. PhD Dissertation, ETH Zürich. [Diss. ETH 17392](https://doi.org/10.3929/ethz-b-00017392). 13, 15, 17, 18, 22, 24, 51
- [62] Kallio, R., Q. Zhou, J. Korpinen, & Koivo, H. 2000. Three Dimensional Position Control of a Parralel Micromanipulator Using Visual Servoing. *Pages 103–111 of: Microrobotics and Microassembly II*. Proc. SPIE, vol. 4194. DOI: [10.1117/12.403690](https://doi.org/10.1117/12.403690) . 27
- [63] Kappa. 2006. *Remote Head Micro Camera Family*. URL <http://kappa.de/en/Serie/Micro.Cameras/>. 32
- [64] Kavli, T., Kirkhus, T., Thielemann, J. T., & Jagielski, B. 2008. Modelling and compensating measurement errors caused by scattering in time-of-flight cameras. In: *Proc. SPIE, Vol. 7066, 706604*. DOI: [10.1117/12.791019](https://doi.org/10.1117/12.791019) . 22, 76, 77
- [65] Kim, D.-H., Kim, K., Kim, K.-Y., & Cha, S.-M. 2001. Dexterous Teleoperation for Micro Parts Handling Based on Haptic/Visual Interface. *Pages 211–217 of: 2001 IEEE International Symposium on Micromechatronics and Human Science*. DOI: [10.1109/MHS.2001.965247](https://doi.org/10.1109/MHS.2001.965247) . 27
- [66] Kim, Y.M., Chan, D., Theobalt, C., & Thrun, S. 2008. Design and Calibration of a Multi-view TOF Sensor Fusion System. In: *Computer Vision and Pattern Recognition (CVPR) Workshops*. DOI: [10.1109/CVPRW.2008.4563160](https://doi.org/10.1109/CVPRW.2008.4563160) . 24
- [67] Kitware Inc. 2009. *The Visualization ToolKit (VTK)*. <http://www.vtk.org/>. 96, 111, 128
- [68] Kohoutek, T. 2007. Monitoring of an industrial robot by processing of 3D-range-imaging data measured by the SwissRanger® SR-3000. In: *Proc. of Optical 3-D Measurement Techniques VIII, Zurich, Switzerland*. 17, 18, 23, 130

B.2. References

- [69] Kolb, A., Barth, E., Koch, R., & Larsen, R. 2009. Time-of-Flight Sensors in Computer Graphics. *Pages 119–134 of: Proc. Eurographics (State-of-the-Art Report)*. [link](#) . 13, 16, 17, 18
- [70] Kunkel, M. and Schulze, J. 2005 (Apr.). *Non-Contact Measurement of Central Lens Thickness*. Tech. rept. Precitec Optronik GmbH. [link](#) . 4
- [71] Lange, Robert. 2000. *3D Time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology*. PhD Dissertation, University of Siegen. [Diss. Uni Siegen](#). 13, 17, 23, 49
- [72] Leica Geosystems AG. 2009. *Scanstation 2*. http://www.leica-geosystems.com/corporate/en/lgs_62189.htm (accessed on 29.06.2009). 18, 20
- [73] Leica Microsystems. 1995. *M-Series Stereomicroscopes*. URL <http://www.leica-microsystems.com/>. 30
- [74] Lick, Zachee. 2006. *Depth from focus with textured light*. MSc. thesis, University of Neuchâtel. 32
- [75] Lindner, M., & Kolb, A. 2007. Data Fusion and Edge-Enhanced Distance Refinement for 2D RGB and 3D Range Images. *Pages 121–124 of: Proc. of the Int. IEEE Symp. on Signals, Circuits & Systems (ISSCS)*, vol. 1. 91
- [76] Lindner, M., Lambers, M., & Kolb, A. 2008. Data Fusion and Edge-Enhanced Distance Refinement for 2D RGB and 3D Range Images. *Int. J. on Intell. Systems and Techn. and App. (IJISTA)*, **5**(1), 344–354. <http://inderscience.metapress.com/link.asp?id=a03545784444r37v> [link](#) . 24
- [77] Lowe, David G. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, **60**(2), 91–110. DOI: [10.1023/B:VISI.0000029664.99615.94](https://doi.org/10.1023/B:VISI.0000029664.99615.94) [link](#) . 101
- [78] Märzhäuser AG. 2000. *Motorized measuring stage and focus controller*. URL <http://www.marzhauser.com/en/products/measuring-stages/motorized-measuring-stages/mtmot-series.html>. 30
- [79] Mathworks Inc. 2008. *Matlab*. <http://www.mathworks.com>. 60
- [80] Matrox Imaging. 2009. *MIL: Matrox Imaging Library*. <http://www.matrox.com/>. 38, 42
- [81] Mesa Imaging SA. 2006 (June). *SwissRanger SR-3000 manual v1.02*. <http://www.swissranger.ch/customer>. 49, 73, 74
- [82] Mesa Imaging SA. 2007. *SwissRanger SR-3000*. <http://www.mesa-imaging.ch/prodviews.php>. 13, 17, 48, 50, 52
- [83] Mure-Dubois, J., & Hügli, H. 2006. Embedded 3D vision system for automated micro-assembly. *Pages 63820J 1–10 of: Two- and Three-Dimensional Methods for Inspection and Metrology IV*. Proc. SPIE, vol. 6382. SPIE. DOI: [10.1117/12.686675](https://doi.org/10.1117/12.686675) [link](#) . 9, 30, 32, 36, 43, 140

Appendix B. Bibliography

- [84] Mure-Dubois, J., & Hügli, H. 2007a. Optimized scattering compensation for time-of-flight camera. *Pages 67620H 1–11 of: Two- and Three-Dimensional Methods for Inspection and Metrology V*. Proc. SPIE, vol. 6762. SPIE. DOI: [10.1117/12.733961](https://doi.org/10.1117/12.733961) [link](#) [link](#) . 80
- [85] Mure-Dubois, J., & Hügli, H. 2007b. Real-time scattering compensation for time-of-flight camera. *In: Proc. of the ICVS 2007*. International Conference on Computer Vision Systems 2007. DOI: [10.2390/biecoll-icvs2007-167](https://doi.org/10.2390/biecoll-icvs2007-167) [link](#) [link](#) . 22
- [86] Mure-Dubois, J., & Hügli, H. 2007c. Time-of-flight imaging of indoor scenes with scattering compensation. *Pages 117–123 of: Proc. O3D 2008*. 8th Conference on Optical 3D Measurement Techniques. [link](#) [link](#) .
- [87] Mure-Dubois, J., & Hügli, H. 2008a. Automated inspection of microlens arrays. *Pages 700007 1–9 of: Optical and Digital Image Processing*. Proc. SPIE, vol. 7000. SPIE. DOI: [10.1117/12.781015](https://doi.org/10.1117/12.781015) [link](#) .
- [88] Mure-Dubois, J., & Hügli, H. 2008b. Fusion of Time-of-Flight Camera Point Clouds. *In: Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications*. European Conference on Computer Vision 2008. [link](#) [link](#) . 24, 25
- [89] Mure-Dubois, J., & Hügli, H. 2008c. Merging of range images for inspection or safety applications. *Pages 70660K 1–12 of: Two- and Three-Dimensional Methods for Inspection and Metrology VI*. Proc. SPIE, vol. 7066. SPIE. DOI: [10.1117/12.793629](https://doi.org/10.1117/12.793629) [link](#) .
- [90] Niclass, C., Rochas, A., Besse, P.-A., & Charbon, E. 2005. Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes. *IEEE Journal of Solid-State Circuits*, **40**(Sept.), 1847–1854. [link](#) [link](#) . 15, 20
- [91] Niclass, C., Soga, M., & Charbon, E. 2007. 3D Imaging based on Single Photon Detectors. *In: 2nd Symposium on Range Imaging (RIM07)*. Invited Paper, [link](#) . 14, 15, 16, 17
- [92] Noguchi, M., & Nayar, S. K. 1994. Microscopic shape from focus using active illumination. *Pages 147–152 of: Proc. of the 12th IAPR Intl. Conf. on Pattern Recognition*, vol. 1. DOI: [10.1109/ICPR.1994.576247](https://doi.org/10.1109/ICPR.1994.576247) . 32
- [93] Oggier, T., Lehmann, M., Kaufmann, R., Schweizer, M., Richter, M., Metzler, P., Lang, G., Lustenberger, F., & Blanc, N. 2004. An all-solid-state range camera for real-time imaging with sub-centimeter depth resolution (Swissranger). *Pages 534–545 of: Proc. of the SPIE 5249 : Optical Design and Engineering*. SPIE. DOI: [10.1117/12.513307](https://doi.org/10.1117/12.513307) [link](#) . 17
- [94] Oggier, T., Büttgen, B., Lustenberger, F., Becker, G., Rüegg, B., & Hodac, A. 2005. SwissRanger SR3000 and First Experiences based on Miniaturized 3D-TOF Cameras. *In: Proc. of the First Range Imaging Research Day at ETH Zurich*. ETH Zurich. [link](#) . 18, 51
- [95] Point Grey Research. 2006. *Bumblebee*. <http://www.ptgrey.com/products/bumblebee/index.asp> (accessed on 01.09.2006). 19

B.2. References

- [96] Precitec Optronik GmbH. 2005. *Optical Probes*. URL <http://www.precitec.de/images/CHRocodile-Datenblatt-e.pdf> (accessed on 25.07.2006). 4
- [97] Promotion and Display Technology Ltd. 2009. *Minoru*. <http://www.minoru3d.com/>. 19
- [98] Pulnix. 1996. *TM-1001 High resolution progressive scanning camera*. URL <http://www.jai.com/EN/CameraSolutions/Products/Pages/TM-1001.aspx> (last accessed on 20.07.2009). 30
- [99] Rabbani, T., & van den Heuvel, F. 2005. Automatic point cloud registration using constrained search for corresponding objects. *In: 7th Conference on Optical-3D measurement techniques, Vienna, Austria*. <http://doi.ewi.tudelft.nl/live/binaries/7682790e-6925-44f4-8601-b98086ba5e03/doc/Optical3dTahir-1.pdf>. 25, 98, 101
- [100] Robbins, Scott, Murawski, Bryan, & Schroeder, Brigit. 2009. Photogrammetric calibration and colorization of the SwissRanger SR-3100 3-D range imaging sensor. *Optical Engineering*, **48**(5), 053603+. DOI: [10.1117/1.3122002](https://doi.org/10.1117/1.3122002). 24
- [101] Roy, M., Svahn, P., Cherela, L., & Sheppard, C. J. R. 2002. Geometric phase-shifting for low-coherence interference microscopy. *Optics and Lasers in Engineering*, **37**(6), 631–641. DOI: [10.1016/S0143-8166\(01\)00146-4](https://doi.org/10.1016/S0143-8166(01)00146-4). 7
- [102] Rusinkiewicz, S., & Levoy, M. 2001. Efficient Variants of the ICP Algorithm. *Pages 145–152 of: Proc. of the Third Intl. Conf. on 3D Digital Imaging and Modeling*. DOI: [10.1109/IM.2001.924423](https://doi.org/10.1109/IM.2001.924423) link. 24
- [103] Saleh, B., & Teich, M. 1991. *Fundamental of photonics*. Wiley. 73, 80
- [104] Schnabel, R., Wahl, R., & Klein, R. 2007. Efficient RANSAC for point-cloud shape detection. *Computer Graphics Forum*, **26**(2), 214226. DOI: [10.1111/j.1467-8659.2007.01016.x](https://doi.org/10.1111/j.1467-8659.2007.01016.x) link. 25
- [105] Schnabel, R., Wessel, R., Wahl, R., & Klein, R. 2008. Shape Recognition in 3D Point-Clouds. *In: Skala, V. (ed), Proc. of The 16-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2008*. link link see also: link. 25
- [106] Schütz, Christian. 1998. *Geometric point matching of free-form 3D objects*. PhD Dissertation, Université de Neuchâtel. 25, 97, 106
- [107] Smart, J., Roebling, R., & the wxWidgets team. 2009. *wxWidgets cross-platform GUI library*. <http://www.wxwidgets.org>. 128
- [108] Soucy, M., & Laurendeau, D. 1995. A General Surface Approach to the Integration of a Set of Range Views. *IEEE Trans. Pattern Anal. Mach. Intell.*, **17**(4), 344–358. DOI: [10.1109/34.385982](https://doi.org/10.1109/34.385982). 89
- [109] Strang, Gilbert. 1988. *Linear Algebra and Its Applications*. Brooks Cole. third edition. 93, 100, 101

Appendix B. Bibliography

- [110] Sulzmann, T., Carlier, J., & Jacot, J. 1996. Distributed Microscopy: towards a 3D computer graphic-based multi user microscopic manipulation, imaging and measurement system. *Pages 183–191 of: Schenker, P., & McKee, G. (eds), Sensor Fusion and Distributed Robotic Agents. Proc. SPIE, vol. 2905. DOI: 10.1117/12.256328 . 27*
- [111] Tamadazte, B., Dembele, S., Fortier, G., & Le Fort-Piat, N. 2008. Automatic micromanipulation using multiscale visual servoing. *Pages 977–982 of: Automation Science and Engineering, 2008. CASE 2008. IEEE International Conference on. DOI: 10.1109/COASE.2008.4626460 link . 27*
- [112] Tarsha-Kurdi, F., Landes, T., & Grussenmeyer, P. 2007. Hough-transform and extended RANSAC algorithms for automatic detection of 3d building roof planes from Lidar data. *In: ISPRS Proceedings. Workshop Laser scanning, Espoo, Finland. link . 23, 25*
- [113] Technologies, PMD. 2006. *PMDvision. http://www.pmdtec.com/e_index.htm. 13, 17, 20, 51*
- [114] Theuwissen, Albert J.P. 1995. *Solid-state imaging with charge-coupled devices. Kluwer. 32, 49, 50*
- [115] Tomlins, P.H., & Wang, R.K. 2005. Theory, developments and applications of optical coherence tomography. *J. Phys. D: Appl. Phys., 38, 2519–2535. DOI: 10.1088/0022-3727/38/15/002 . 7*
- [116] Torr, P.H.S., & Murray, D.W. 1997. The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix. *International Journal of Computer Vision, 24, 271–300. DOI: 10.1023/A:1007927408552 link . 24*
- [117] Trucco, E., Fusiello, A., & Roberto, V. 2007. Robust motion and correspondence of noisy 3-D point sets with missing data. *Pattern Recognition Letters, 20(9)(Sept.), 889–898. DOI: 10.1016/S0167-8655(99)00055-0 link . 25*
- [118] Tsai, Roger Y. 1987. A versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses. *IEEE Journal of Robotics and Automation, RA-3(Aug.), 323–344. DOI: 10.1109/JRA.1987.1087109 link . 24, 51, 90*
- [119] Turk, Greg, & Levoy, Marc. 1994. Zippered polygon meshes from range images. *Pages 311–318 of: SIGGRAPH 94: Proceedings of the 21st annual conference on Computer graphics and interactive techniques. New York, NY, USA: ACM. DOI: 10.1145/192161.192241 . 24*
- [120] Villaverde, Ivan, & Graña, Manuel. 2009. An Improved Evolutionary Approach for Egomotion Estimation with a 3D TOF Camera. *Pages 390–398 of: Bioinspired Applications in Artificial and Natural Computation. DOI: 10.1007/978-3-642-02267-8_42 . 18, 23*
- [121] von Hansen, W., Michaelsen, E., & Thönnessen, Ulrich. 2006. Cluster Analysis and Priority Sorting in Huge Point Clouds for Building Reconstruction. *Pages 23–26 of: 18th International Conference on Pattern Recognition (ICPR'06), vol. 1. DOI: 10.1109/ICPR.2006.1197 . 25, 98*

B.2. References

- [122] Weyrich, T., Pauly, M., Keiser, R., Heinzle, S., Scandella, S., & Gross, M. 2004. Post-processing of Scanned 3D Surface Data. *Pages 85–94 of: Proc. of Eurographics Symposium on Point-Based Graphics 2004*. <http://graphics.stanford.edu/ma-pauly/Pdfs/PostProcessing.pdf> link . 17, 23
- [123] Willow Garage. 2009. *Open Source Computer Vision Library (OpenCV)*. <http://pr.willowgarage.com/wiki/OpenCV>. 19, 24, 42, 91
- [124] Yahav, G., Iddan, G. J., & Mandelbom, D. 2007. 3D Imaging Camera for Gaming Application. *Pages 1–2 of: Proc. International Conference on Consumer Electronics*. DOI: [10.1109/ICCE.2007.341537](https://doi.org/10.1109/ICCE.2007.341537) . 16
- [125] Zamofing, T., & Hügli, H. 2004a. Applied multifocus 3D microscopy. *Pages 134–144 of: Batchelor, B. G., & Hügli, H. (eds), Two- and Three-Dimensional Vision Systems for Inspection, Control, and Metrology*. Proc. SPIE, vol. 5265. DOI: [10.1117/12.518827](https://doi.org/10.1117/12.518827) link . 11, 30, 34, 38, 40, 44, 144
- [126] Zamofing, T., & Hügli, H. 2004b. Multiresolution reliability scheme for range image filtering. *Pages 98–105 of: Harding, K. G. (ed), Two- and Three-Dimensional Vision Systems for Inspection, Control, and Metrology II*. Proc. SPIE, vol. 5606. DOI: [10.1117/12.580476](https://doi.org/10.1117/12.580476) link . 30, 38, 40
- [127] Zamofing, Thierry. 2003 (July). *BurrTrol Report*. Tech. rept. IMT 447 HU 07/03. Institute of Microtechnology, University of Neuchâtel. Available on IMT/Orange: [IMT447.M3DReport](#). 39, 149
- [128] Zhang, S., & Huang, P. 2004 (June). High-Resolution, Real-time 3D Shape Acquisition. *Pages 1–10 of: 2004 Conference on Computer Vision and Pattern Recognition Workshop*. DOI: [10.1109/CVPR.2004.86](https://doi.org/10.1109/CVPR.2004.86) . 19, 21
- [129] Zhang, Zhengyou. 1994. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, **13**(2), 119–152. DOI: [10.1007/BF01427149](https://doi.org/10.1007/BF01427149) link . 106
- [130] Zhang, Zhengyou. 1999. Flexible Camera Calibration By Viewing a Plane From Unknown Orientations. *Pages 666–673 of: International Conference on Computer Vision (ICCV'99)*. IEEE. DOI: [10.1109/ICCV.1999.791289](https://doi.org/10.1109/ICCV.1999.791289) link . 24, 91