



HAL
open science

Géo-référencement précis d'acquisition photogrammétrique de “ longues ” scènes d'intérieur

Truong Giang Nguyen

► **To cite this version:**

Truong Giang Nguyen. Géo-référencement précis d'acquisition photogrammétrique de “ longues ” scènes d'intérieur. Traitement du signal et de l'image [eess.SP]. Université Paris-Est, 2018. Français. NNT : 2018PESC2182 . tel-02476477

HAL Id: tel-02476477

<https://theses.hal.science/tel-02476477>

Submitted on 12 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

École doctorale MSTIC
Mathématiques & Sciences et Technologies
de l'Information et de la Communication
Spécialité: Sciences et Technologies de l'Information Géographique

THÈSE DE DOCTORAT

En vue de l'obtention du grade de :
Docteur de l'Université Paris-Est

présentée et soutenue le 19 Décembre 2018 par :

NGUYEN Truong Giang

Géo-référencement précis d'acquisition photogrammétrique de
"longues" scènes d'intérieur

Composition du jury :

Rapporteur	M. Marie-Odile Berger	INRIA Nancy Grand Est
Rapporteur	M. El-Mustapha Mouaddib	Université de Picardie Jules Verne
Examineur	M. Valerie Renaudin	IFSTTAR
Examineur	M. Bertrand Cannelle	HEIG-VD
Directeur	M. Pierrot-Deseilligny Marc	IGN-ENSG
Co-directeur	M. Thom Christian	IGN

REMERCIEMENTS

Je remercie tout d'abord Monsieur Marc PIERROT DESEILLIGNY, mon directeur de thèse, qui m'a donné l'opportunité de poursuivre mes études en tant que doctorant et qui m'a encadré tout au long de ma thèse. Je tiens à remercier également les membres de l'équipe MicMac de l'IGN : Ewelina RUPNIK, Mehdi DAAKIR, Jérémie BELVAUX, Yilin ZHOU et Jonathan LISEIN. Nous avons partagé nos connaissances, de belles expériences et de beaux moments inoubliables. J'adresse aussi tout mes remerciements à l'équipe LOEMI de l'IGN pour son accueil chaleureux. Je remercie Monsieur Christian THOM, le directeur de l'équipe, pour son encadrement. Je remercie Olivier MARTIN et Christophe MAYNARD pour leur partage de connaissances sur le matériel et l'électronique. Je remercie tout particulièrement Monsieur Jean Michael MÜLLER, mon encadrant, pour tout ses efforts qu'il a investi afin de me guider, pour sa patience et pour toutes ses réponses à mes interrogations. Je remercie surtout Monsieur Jean-Philippe SOUCHON pour tout le temps dépensé afin de relire et corriger ce manuscrit. Je remercie aussi les doctorants de l'équipe : Mohamed Amjad LASRI et Ahmad AUDI.

Encore une fois, je tiens à vous faire savoir que je n'oublierai jamais toutes les aides que vous m'avez apporté tant sur le côté professionnel que sur le côté personnel durant ma thèse, pour votre gentillesse et pour les nombreux encouragements que vous m'avez prodigué. Je garde précieusement nos souvenirs passés ensemble dans les brasseries, les fêtes, lors de nos missions et lors de nos nombreux repas.

Je remercie Madame Marie-Odile BERGER et Monsieur El-Mustapha MOUADDIB d'avoir accepté d'être le rapporteur de cette thèse. J'exprime ma gratitude à Madame Valerie RENAUDIN ainsi qu'à Monsieur Bertrand CANNELLE, qui ont bien voulu être les examinateurs.

Je remercie toute l'équipe DIAS à l'ENSG : Jean-François HANGOUËT, Jean-Pierre PAPELARD, Patricia PELLARDI et Antoine PINTE pour leur accueil et de m'avoir accepté dans les stages terrains avec des élèves de l'école. Vous m'avez donné beaucoup de connaissances dans le domaine géomatique. Je remercie spécialement Marc POUPÉE pour le partage des voyages et des loisirs intéressants.

Je tiens à remercier les personnels de l'IGN : Paul et Sylvain de l'équipe Travaux Spéciaux, Laurent CARAFFA, Quoc Dinh NGUYEN de MATIS pour ses conseils.

Je remercie profondément mes parents et ma famille. Ils sont toujours rester à mes côtés, m'ont toujours soutenu et encouragé tout au long de mes études. Je remercie tous mes amis qui m'ont suivi tout au long de ces années en France : Trong Tai, Khoa, Thanh, Ba Tung, Duy Ngan, Phuong Hanh, Trung Truc, Viet Phuong, Thinh, Son, Mai Huong, Vy, Thi Ha, Hong Son, Hoa, Le Hai, Thanh, Truong Phi, Thanh Huyen... Je vous souhaite la santé et le bonheur.

Plus personnellement, je remercie ma copine, Thi-Tuyen PHAN, pour son aide dans la relecture et les corrections d'orthographe, son encouragement, de m'avoir partager sa joie de vivre et surtout, pour son amour qui m'a été essentiel durant la fin de ma thèse.

RÉSUMÉ

Au cours des dernières décennies, plusieurs ruptures technologiques ont favorisé le développement de la photogrammétrie : développement et miniaturisation des appareils photographiques numériques, avènement du drone civil et de la cartographie mobile, apparition de nouveaux algorithmes permettant des traitements 100% automatiques, etc. La photogrammétrie est redevenue une technique de mesure et de surveillance incontournable, surtout dans les contextes où le coût et la légèreté du matériel sont contraignants.

Dans ce contexte, l'IGN et l'équipe LOEMI du laboratoire LaSTIG cherchent à développer des plateformes matérielles et logicielles permettant d'atteindre simultanément productivité et précision de la mesure.

Les utilisations de la photogrammétrie qui nécessitent une grande précision sont par exemple la mesure de déformations ou encore la métrologie industrielle. Sur des acquisitions à larges étendues, notamment les acquisitions linéaires de type corridor, les relevés photogrammétriques sont souvent entachés d'erreurs systématiques, ayant de fortes conséquences sur la précision de localisation du résultat final.

Dans la chaîne métrologique de photogrammétrie, l'extraction des points homologues est une des causes de l'imprécision du résultat final.

Cette thèse consiste à réaliser une méthode permettant d'améliorer la précision du processus de traitement photogrammétrique existant pour réduire les erreurs systématiques.

Nous proposons une méthode de post-traitement du processus photogrammétrique classique. Nous utilisons des modélisations de l'acquisition (poses et calibration des caméras) et de la scènes (maillage 3D) obtenues avec un traitement classique, pour extraire de nouveaux points homologues en optimisant leurs caractéristiques pour la photogrammétrie. Ces caractéristiques sont : la distribution optimale dans l'espace image et objet, la multiplicité des points sur les images et la précision de la mesure. Une seconde itération d'ajustement de faisceaux avec les nouveaux points obtenus permet de raffiner l'orientation externe et la calibration de la caméra et donc d'améliorer la précision de localisation des points 3D finaux.

Les résultats obtenus à partir des jeux de tests issus de différents scénarios montrent l'efficacité et la robustesse de la méthode en diminuant significativement le résidu des points 3D triangulés. Les temps de calcul et le nombre d'itérations sont également étudiés. Les résultats du processus de raffinement convergent dès la seconde itération et montrent que cette méthode ne demande qu'environ 10% du temps d'exécution total de la chaîne classique pour atteindre une amélioration notable.

Mots clefs : point homologue, traitement d'image, estimation de pose, orientation externe, ajustement de faisceaux, photogrammétrie

ABSTRACT

Over recent decades, many technical advances such as the development and miniaturization of electronic components and digital camera, the advent of civil UAV and mobile mapping and the apparition of new algorithms allow 100% automatic processing yield to a strong development of photogrammetry. It becomes an indispensable measurement and surveillance technique, especially in the contexts where budget and compactness of the equipment matters the most.

In this context, IGN and LOEMI team decided to develop the new hardware platforms and software solution allowing enhancements of both productivity and precision.

Besides the visualization and demonstration, certain photogrammetry applications require a good measurement precision, for example, change detection for deformation studies in geosciences or very accurate shape reconstruction in industrial metrology. On large areas acquisitions, especially corridors, the photogrammetric process precision is usually limited due to systematic errors, which yield to an insufficient localization precision of the final result.

In the automatic digital photogrammetric processing chain, tie-points extraction is considered as a first step. Therefore, it is one of the important reasons that cause the imprecision of the final result.

In this research work, we describe a new method, which reduces the systematic errors to enhance the precision of the existing digital photogrammetric processing chain.

We propose a post-processing method for the classical photogrammetric chain. We use the results of the classical chain such as images poses, camera calibration, and mesh of scene. Our method extracts totally new tie-points pack with characteristics optimized for the photogrammetry. These characteristics are: an optimal tie-points distribution in image and object space, a high tie-points multiplicity and a precise points measure on image space. A second bundle adjustment iteration using these new tie-points pack allows refining of external images orientations and camera calibration. In consequence, the localization precision of final triangulated 3D points is enhanced.

Evaluated on many different test scenes, the proposed method shows its efficacy and robustness in improving the obtained 3D points accuracy. Computation time and number of iteration are also discussed in the manuscript. The proposed method converges from the second iteration. It requires additional computation time around 10% of the total time needed by the classical processing chain to reach a significant enhancement.

Keywords : tie-points, digital image processing, pose estimation, orientation refinement, bundle adjustment, photogrammetry.

Table des matières

Table des matières	7
Liste des figures	11
Liste des tableaux	15
1 Introduction	17
1.1 Contexte	17
1.1.1 Brève histoire de la photogrammétrie	17
1.1.2 Photogrammétrie numérique	19
1.2 La photogrammétrie à l'IGN	21
1.2.1 La recherche à l'IGN	21
1.2.2 Chaîne de traitements photogrammétrique à l'IGN	22
1.2.3 Contexte de la thèse	24
1.3 Présentation du sujet de thèse	25
1.3.1 Problématique	25
1.3.2 Objectif de la thèse	26
1.3.3 Description générale de la méthode proposée	26
2 Contributions, activités et organisation du document	29
2.1 Contributions et activités	29
2.1.1 Publications	29
2.1.2 Activités scientifiques et académiques	30
Formations dispensées	30
Formations suivies	30
Travaux terrain	30
Logiciel	30
2.2 Organisation du document	31
3 État de l'art	33
3.1 Introduction	33
3.2 Sources d'erreurs en photogrammétrie	34
3.3 Points homologues	37
3.3.1 Le détecteur et descripteur de points d'intérêt	38
3.3.2 L'appariement	40
Appariement basé caractéristiques visuelles	41
Appariement basé région	41
La corrélation	41
L'appariement par moindres carrés	43
3.4 Conclusion	45

4	Extraction des points homologues SIROP	47
4.1	Introduction	47
4.2	Description détaillée de l'algorithme	48
4.2.1	Sélection des images	48
4.2.2	Rectification des régions d'intérêt	50
4.2.3	Détection des points d'intérêt	52
	Sélection des extrema locaux	52
	Filtrage sur le contraste	53
	Filtrage des motifs répétitifs	54
	Filtrage spatial des points d'intérêt	55
4.2.4	Appariement par la corrélation	55
4.2.5	Filtrage spatial des points homologues	56
4.2.6	Raffinement par moindres carrés	58
4.3	Conclusion	59
5	Résultats	63
5.1	Introduction	63
5.2	Les jeux de données	64
5.3	Méthode d'évaluation	65
5.4	Résultats	66
5.4.1	Les indicateurs d'erreur	66
5.4.2	Couloir	66
5.4.3	Viabon 1	70
5.4.4	Corridor	72
5.4.5	LadyBug	74
6	Conclusions et perspectives	79
6.1	Conclusions	79
6.2	Discussion	80
6.2.1	Plusieurs itérations	81
6.2.2	Temps de calcul	82
6.2.3	Accélération du processus classique	82
6.3	Perspectives	83
6.3.1	Images satellites	83
6.3.2	Optimisation des paramètres de la traitement classique	83
6.3.3	Mixage SIFT/SIROP	84
6.3.4	Modèle 3D théorique	84
A	Modèle de caméra <i>Pinhole</i>/Sténopé	85
A.1	Du repère du monde au repère caméra	86
A.2	Du repère caméra au repère image	87
A.3	La transformation finale	88
B	Modèle de distorsion	89
B.1	Distorsion radiale	90
B.2	Distorsion Fraser	91
B.3	L'objectif Fisheye	91
C	Auto-calibration	95

D	Contraintes photogrammétriques	97
D.1	Contrainte de projection des points homologues	97
D.2	Contrainte de rigidité inter-caméras	98
D.3	Références	100

Liste des figures

1.1	Gauche : une stéréoscopie ; Milieu : première image pris par héliographe ; Droite : le daguerréotype. Source : <i>Wikipedia</i>	18
1.2	Des applications de la photogrammétrie et de reconstruction 3D. Première ligne : Tango (Google), Hololens (Microsoft), VPS (Google), Deuxième ligne : Géoportail (IGN), ITowns (IGN)	19
1.3	Processus de traitement de photogrammétrie numérique.	21
1.4	Illustration des étapes du processus de traitement de photogrammétrie numérique. Gauche : extraction des points homologues, Milieu : calibration, Droite : nuage de points et orientation externe	21
1.5	De gauche à droite : Caméra aérienne V1, V2, Caméra légère Camlight et GéoCube	22
1.6	Une vue générale des outils et des niveaux d'interface de la chaîne MicMac [Rupnik et al., 2017].	24
1.7	Quelques travaux réalisés avec la chaîne MicMac par les étudiants et l'équipe DIAS de l'ENSG	24
1.8	Des scénarios longs linéaires : la cartographie urbaine avec la plateforme Stéréopolis, les égouts de Paris	26
1.9	Diagramme de flux du processus de seconde itération <i>SIROP</i>	27
3.1	La simulation de l'effet de courbure sur plusieurs scénarios et les paramètres internes de la caméra présentés dans [James and Robson, 2014]	35
3.2	Droite : illustration de l'effet de courbure par comparaison entre le traitement qui utilise des images corrigées (undistorted) et originales (original) [Eltner et al., 2016], Bas Gauche : effet causé par la distorsion radiale mal estimée dans la configuration critique [Wu, 2014], Haut Gauche : les différences d'effet de courbure selon les modèles de caméra, et la correction en utilisant un modèle plus sophistiqué (R15 dans la chaîne MicMac) [Tournadre, 2015]	36
3.3	La correction de l'effet de courbure par l'ajout d'images obliques proposée par Gauche : [James and Robson, 2014] et Droite : [Nocerino et al., 2014].	36
3.4	Gauche : les appariements corrects (croix blanches) en comparant <i>SIFT</i> et <i>GSIFT</i> , Droite : <i>CSIFT</i> descripteur plus robuste aux changements d'illumination sur une image couleur.	39
3.5	La distribution de points est plus homogène dans <i>ATiPE</i> [Barazzetti et al., 2010] (Bas) en comparaison avec <i>SIFT</i> (Haut).	40
3.6	La corrélation : Gauche : recherche d'un endroit sur l'image 2 le plus similaire à la fenêtre de l'image 1. Droite : la recherche peut être réduite à la région autour du point d'intérêt.	42

4.1	Les étapes successives de l'algorithme d'extraction de points homologues <i>SIROP</i>	48
4.2	Directions d'observations et résolution minimale.	49
4.3	La sélection de l'image maîtresse selon le critère de résolution optimale.	51
4.4	La rectification de la ROI.	51
4.5	L'erreur de prédiction entre homographie et affinité en fonction de la surface du triangle.	52
4.6	Les étapes de la détection des points d'intérêt. En rouge : effectuées sur toutes les images sélectionnées, En vert : effectuées uniquement sur l'image maîtresse.	52
4.7	La méthode de calcul des scores de contraste CQS_1 et CQS_2	54
4.8	Un exemple de filtrage par le contraste.	54
4.9	Filtrage des motifs répétitifs.	54
4.10	Le filtre de réduction du nombre de points d'intérêt.	55
4.11	La sélection des points appariés potentiels.	55
4.12	Le calcul du score global de corrélation $SGC(P_k)$ pour chaque point homologue multiple P_k dans l'image maîtresse.	57
4.13	Illustration du filtrage spatial. Des points verts sont déjà sélectionnés. Le point jaune est courant. Le point rouge est supprimé.	57
4.14	Étapes du filtre spatial. (Gauche) : tri des points par le score global de corrélation $SGC(P_k)$ représenté par les colonnes vertes, (Milieu) : la région de filtrage en jaune, (Droite) : les points retenus après filtrage	58
4.15	Evaluation de l'apport du raffinement par <i>LSM</i> sur des points <i>SIROP</i> sur l'histogramme d'erreurs de re-projection et sur l'évaluation des paramètres internes.	59
5.1	Exemples d'images acquises lors des expérimentations.	65
5.2	Le processus de traitement des données.	65
5.3	Quelques images du jeu de données <i>Couloir</i> . Caméra photogrammétrique légère IGN CamLight [Martin et al., 2014], longueur focale 18mm.	66
5.4	Le jeux de données <i>Couloir</i> et ses points terrain (rouge).	67
5.5	Comparaison de la distribution des points. Gauche : SIFT, Droite : <i>SIROP</i> , le carré rouge représente les zones lointaines qui sont une limitation actuelle	67
5.6	Partition de la scène en images maîtresses par la condition de la résolution optimale. En haut : chaque couleur est la partie de la scène assignée à une image maîtresse. En bas : un image maîtresse (rouge) et ses secondaires (bleu).	67
5.7	Statistiques sur les points homologues entre les deux processus : <i>SIROP</i> et classique <i>SIFT</i>	68
5.8	Les points dans l'espace objet avec leurs résidus codés en couleur. L'histogramme d'erreur de re-projection dans la gamme de [0,1] pixel est également représenté sur l'échelle de couleur. Haut : <i>SIFT</i> , Bas : <i>SIROP</i>	68
5.9	Les prises de vue avec une grande variation de GSD sont légèrement défavorables à la multiplicité des points <i>SIROP</i>	69
5.10	La distance entre les coordonnées des points terrain - <i>Couloir</i> . De haut en bas : la distance euclidienne, en X et Y et Z. La distance moyenne est encadrée en rouge.	69
5.11	Quelques images du jeu de données <i>Viabon 1</i>	70

5.12	Les points homologues <i>SIFT</i> (Gauche), <i>SIROP</i> (Milieu) et la partition de la scène selon les différentes images maîtresses(Droite).	70
5.13	Les points dans l'espace objet avec leurs résidus codés en couleur. <i>SIFT</i> (Gauche) et <i>SIROP</i> (Droite) (<i>Viabon 1</i>).	71
5.14	Statistiques sur les points homologues entre les deux processus : <i>SIROP</i> et classique sur le jeu de données <i>Viabon 1</i>	71
5.15	Comparaison des résidus en 3D sur les points terrain entre les méthodes <i>SIROP</i> et classique.	72
5.16	Gauche : le jeu de données <i>Corridor</i> et ses points terrain (points rouges). Droite : l'effet de courbure observé sur l'axe <i>Z</i> à l'aide des points terrain le long de la scène.	72
5.17	Points dans l'espace objet avec leurs résidus images (pixels) codés en couleur avec <i>SIFT</i> (Haut) et avec <i>SIROP</i> (Bas) (<i>Corridor</i>).	73
5.18	Statistiques sur les points homologues entre les deux processus : <i>SIROP</i> et classique sur le jeu de données <i>Corridor</i>	73
5.19	Comparaison des résidus en 3D sur des points terrain entre les méthodes <i>SIROP</i> et Classical.	74
5.20	La caméra multi-têtes panoramique LadyBug.	74
5.21	5 images du même bloc d'images de la caméra LadyBug.	75
5.22	Gauche : L'ensemble d'immeubles avec sa taille et l'indication des 4 points terrains utilisés pour la mise à l'échelle. Droite : Nuage de points et les blocs d'images (en rouge).	75
5.23	L'erreur de fermeture de boucle : Gauche : les liens des points homologues pour pouvoir considérer une acquisition en boucle comme un cas linéaire ; Doite : l'estimation de l'erreur de fermeture de boucle avec deux triangulations différentes.	76
5.24	Comparaison des résidus de fermeture de boucle issus des différents tests.	77
6.1	Résidus images et terrains après l'exécution successive du processus <i>SIROP</i>	81
6.2	Évolution des paramètres internes de la caméra après les itérations <i>SIROP</i>	81
6.3	L'extraction des points homologues sur les images diachroniques : Gauche : <i>SIFT</i> Droite : <i>SIROP</i>	83
A.1	Gauche : le caméra sténopé. (Source : <i>Wikipedia</i>) – Droite : Modèle mathématique de la caméra sténopé.	85
B.1	La distorsion radiale et distorsion tangentielle.	89
B.2	Différentes fonctions radiales de correspondance entre l'angle d'entrée θ et la position sur le plan image r	92
B.3	Modèle de caméra avec l'objectif Fisheye	93
C.1	La correction de distorsion d'une image prise par l'objectif Fisheye. Gauche : les vecteurs du résidu de calibration, Milieu : image originale, Droite : image corrigée.	96

Liste des tableaux

4.1	Tableau des valeurs des seuils par défaut dans l'implémentation de la méthode <i>SIROP</i> dans MicMac.	60
5.1	Description des jeux de données expérimentaux.	64
5.2	Tableau de comparaison des résidus de fermeture de boucle entre les 4 méthodes.	77
6.1	Comparaison entre le traitement classique et le processus <i>SIROP</i> sur les résidus 3D, erreur de reprojection et multiplicité des points.	80

Chapitre 1

Introduction

Sommaire

1.1 Contexte	17
1.1.1 Brève histoire de la photogrammétrie	17
1.1.2 Photogrammétrie numérique	19
1.2 La photogrammétrie à l'IGN	21
1.2.1 La recherche à l'IGN	21
1.2.2 Chaîne de traitements photogrammétrique à l'IGN	22
1.2.3 Contexte de la thèse	24
1.3 Présentation du sujet de thèse	25
1.3.1 Problématique	25
1.3.2 Objectif de la thèse	26
1.3.3 Description générale de la méthode proposée	26

1.1 Contexte

L'objectif de cette thèse est d'améliorer la précision des traitements photogrammétriques dans le cadre de l'utilisation du logiciel *open source* MicMac développé à l'IGN, par l'ajout d'une phase finale de traitement au processus classique.

1.1.1 Brève histoire de la photogrammétrie

La photogrammétrie est une technique de mesure sans contact servant à déterminer les coordonnées d'un point 3D à partir d'images.

L'histoire de la photogrammétrie, la transition de l'utilisation des clichés argentiques aux images numériques, les applications intéressantes de la photogrammétrie et les principes de la chaîne de traitement numérique sont présentés dans les parties suivantes.

La photogrammétrie est basée sur le principe de la vision stéréoscopique humaine publié par Charles Wheatstone au $XV^{\text{ème}}$ siècle. Cette technique utilise les photos comme source d'information. Elle est donc fortement liée au développement de la photographie au début du $XIX^{\text{ème}}$ siècle.

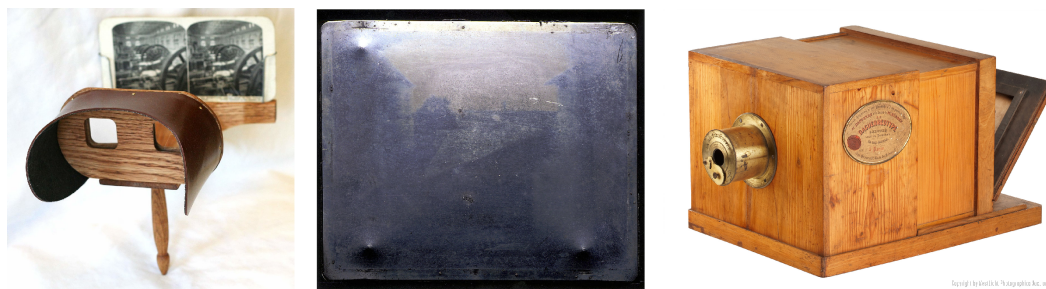


FIGURE 1.1 – **Gauche** : une stéréoscopie; **Milieu** : première image pris par héliographe; **Droite** : le daguerréotype. Source : *Wikipedia*

Le premier travail photogrammétrique a été réalisé par l’officier de l’armée française Aimé Laussedat en 1849 sur la façade de l’Hôtel des Invalides. Il a publié en 1862 une collection des méthodes de mesure à partir de relevés photographiques. Parallèlement, Albrecht Meydenbauer, un ingénieur allemand, a publié la technique de mesure dite “photométrie” en 1867 – aujourd’hui devenue la photogrammétrie.

La restitution photogrammétrique est basée sur la vision stéréoscopique couplée avec un système de mesure permettant de définir les coordonnées de points sur le terrain à partir d’un couple d’images orientées. Selon les caractéristiques des photos utilisées et des données exportées, il existe trois types de systèmes de restitution : analogique (avant 1960), analytique (années 60 à 90) et numérique.

L’apparition des appareils photo numériques et les développements dans les domaines de l’électronique et de l’informatique ont permis une transition essentielle vers la photogrammétrie numérique depuis 1985. L’acquisition d’images numériques et des grandes capacités de stockages de données actuelles permettent de considérer le fait de prendre des photos comme “gratuit” et “illimité”, ce qui change complètement la donne en terme d’acquisition : il devient possible de prendre plusieurs dizaines de photos là où auparavant la modélisation se faisait avec un seul couple, ce qui offre de nouvelles potentialités en terme de robustesse, de précision, de complétion, d’automatisation, etc. La communauté se tourne alors vers le développement de solutions de traitements automatisés sur l’ordinateur, dit photogrammétrie numérique (*digital photogrammetry* ou *Structure from Motion SfM* dans la communauté de vision par ordinateur). Dès les années 1990, une grande vague de solutions de traitements numériques est apparue : MicMac-Apéro, Photomodeler, ShapeCapture, Canoma, Kodak Dimension, 3D Builder Pro, Photo3D, Image Modeler, etc.

Suivant le type d’acquisition, la photogrammétrie peut se diviser en deux grandes catégories : aérienne et terrestre.

La photogrammétrie aérienne est apparue en 1858, grâce à la première image aérienne de Paris prise par Gaspard-Félix Tournachon à une hauteur de 80m depuis une montgolfière. Avec l’apparition des ULM¹, des avions et des drones civils, la photogrammétrie aérienne est aujourd’hui utilisée dans plusieurs domaines comme la cartographie, la surveillance des constructions [Daakir et al., 2017; Tournadre et al., 2015, 2014], le cadastre, la gestion des risques géologiques, l’agriculture [Burgos et al., 2015] ou encore la cartographie forestière [Lisein et al., 2013]. Des études approfondies de l’état de l’art et des applications de photogrammétrie aérienne ont été présentées dans [Lisein, 2016; Tournadre, 2015].

La photogrammétrie terrestre² est appliquée aux divers types de relèvements : les monuments historiques, l’archéologie [Delevoie et al., 2012; Samaan, 2016], les travaux

1. Ultra Léger Motorisé

2. *close-range photogrammetry*

d'architecture [Hanke and Grussenmeyer, 2002; Kennie and McKay, 1987; Werner et al., 2002], l'inspection de mouvements/déformations d'ouvrage [Górski et al., 2010; Luhmann, 2010; Ordóñez et al., 2009], la gestion des risques [Kennie and McKay, 1987; Rieke-Zapp and Nearing, 2005], l'intérieur de structures [Khoshelham and Díaz-Vilariño, 2014], la cartographie urbaine [Bentrah, 2006; Papanoditis et al., 2012; Toth and Paska, 2007], l'agriculture [Aguilar et al., 2010], etc.

La photogrammétrie est aussi très liée à la recherche de méthodes de localisation en intérieur par mesures inertielles ou ondes électromagnétiques, quand le signal GNSS³ est masqué. [Reaudin et al., 2007a,b] présente les idées générales et la conception d'un système de navigation en intérieur basé sur RFID⁴, Bluetooth, MEMS⁵, INS⁶ ou les réseaux UWB⁷. La combinaison entre les données de capteurs inertiels et les techniques de vision est réalisée dans plusieurs applications différentes, par exemple [Aron et al., 2007] l'utilisent pour le suivi d'objets en réalité augmentée.

Aujourd'hui, la photogrammétrie et la reconstruction 3D à grande échelle sont utilisées par des services grand public tels que Google Maps⁸ ou IGN Géoportail⁹. Il s'agit de plateformes Web avec plusieurs types de données comme la cartographie, l'orthophotographie ou le modèle 3D de villes en haute résolution. De plus, le développement de l'électronique et des techniques de géo-référencement direct par GNSS¹⁰ ou par centrale inertielle permet maintenant de réaliser des applications de reconstruction 3D / réalité augmentée en temps réel comme Google Tango / ARCore¹¹, Microsoft Hololens¹² ou Google Visual Positioning System (VPS).

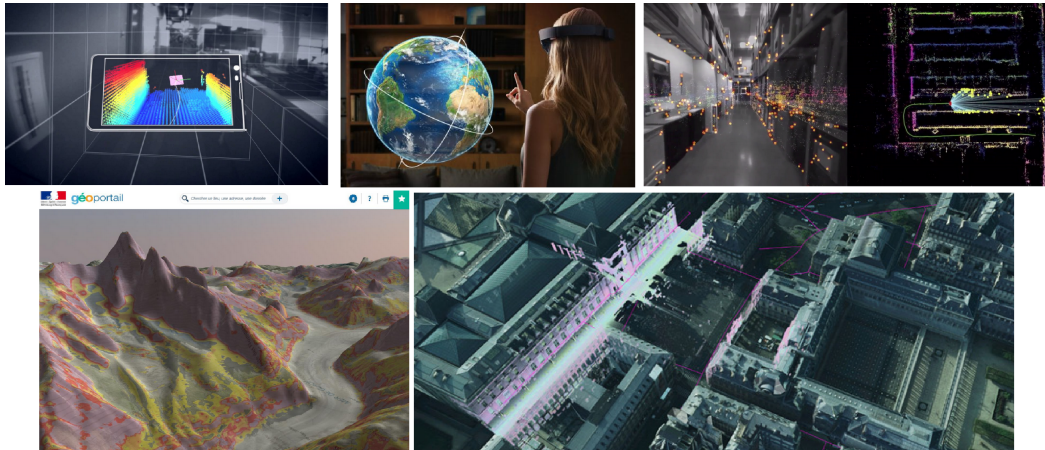


FIGURE 1.2 – Des applications de la photogrammétrie et de reconstruction 3D. **Première ligne** : Tango (Google), Hololens (Microsoft), VPS (Google), **Deuxième ligne** : Géoportail (IGN), ITowns (IGN)

3. Global Navigation Satellite System
4. *Radio-frequency identification*
5. *Microelectromechanical systems*
6. *Inertial navigation system*
7. *Ultra-wideband*
8. <http://maps.google.com/>
9. <https://www.geoportail.gouv.fr/>
10. Global Navigation Satellite System
11. <https://developers.google.com/ar/>
12. <https://www.microsoft.com/en-us/hololens>

1.1.2 Photogrammétrie numérique

Aujourd'hui, la photogrammétrie numérique se fait sur ordinateur avec des méthodes quasiment automatiques. De plus, grâce au développement des algorithmes de reconstruction 3D, la combinaison entre la photogrammétrie et la vision par ordinateur permet de délivrer un modèle 3D satisfaisant avec une faible interaction manuelle. La plupart des étapes dans la chaîne de traitement peuvent être effectuées de manière automatique par les algorithmes de mesure des points homologues, d'auto-calibration, d'appariement dense, etc. Les utilisateurs ayant principalement à acquérir des images et mesurer des points sur le terrain pour le contrôle et le géo-référencement.

Nous pouvons diviser la chaîne de traitements photogrammétrique en cinq grandes étapes :

Extraction des points homologues

Sa fonction est d'extraire sur différentes images des points saillants représentant le même détail terrain en se basant sur des critères de similarité locale¹³. La saisie manuelle sur images argentiques est maintenant remplacée par l'extraction automatique sur images numériques grâce à des algorithmes comme *SIFT* [Lowe, 2004], *SURF* [Bay et al., 2008]. Ces méthodes peuvent fournir un nombre de points homologues assez élevé (de l'ordre de plusieurs milliers lorsque la prise de vue est idéale). L'état de l'art et les informations détaillées de cette étape sont présentés en section 3.3.1.

Synonymes : points de transfert, points homologues, points de liaison.

Calibration interne

Cette étape a pour but de déterminer les paramètres internes du capteur et de l'objectif de la caméra. La longueur focale, le point principal et les coefficients de distorsion pour corriger la déformation d'image causée par l'objectif constituent ces paramètres internes. Un bref rappel des méthodes les plus répandues de calibration et du processus d'auto-calibration est présenté dans l'annexe C.

Synonymes : étalonnage de caméra, calibration, calibration intrinsèque, orientation interne

Orientation externe

Cette étape a pour but de déterminer les paramètres externes de la prise de vue. Il s'agit des paramètres de rotation et de translation pour localiser la position et la direction de la caméra dans l'espace. Ces paramètres sont souvent déterminés en relatif, avec comme observations uniquement les points homologues, sans points d'appuis terrain. La calibration interne et la calibration externe de la caméra peuvent être définies dans le même processus d'auto-calibration.

Synonymes : calibration externe, calibration extrinsèque, estimation de poses, mise en place des images, aéro-triangulation

Compensation / Optimisation

Cette étape est effectuée pour ajuster les paramètres externes/internes de la caméra en utilisant des observations ou des contraintes supplémentaires dans le processus dit d'ajustement de faisceaux [Atkinson, 1996; Triggs et al., 1999].

Appariement dense

Cette étape consiste à appairer de manière dense les images, pour chercher les correspondants de tout les pixels pour en déterminer les coordonnées 3D. Le

13. points de liaisons, points homologues

résultat de cette étape est représenté par une carte de profondeur associée à chaque image et un nuage de points dense 3D dans l'espace objet.

Export - Production

Il s'agit de plusieurs étapes. Nous ne les détaillons pas toutes car elles arrivent en aval du travail réalisé dans cette thèse. Cette dernière étape produit des nuages de points denses, l'orthophoto, des maillages 3D (texturés ou non), des modèles numériques de terrain, etc., à partir des orientations des images et des cartes de profondeur. Dans ce travail nous utilisons le maillage 3D comme modélisation de la scène.

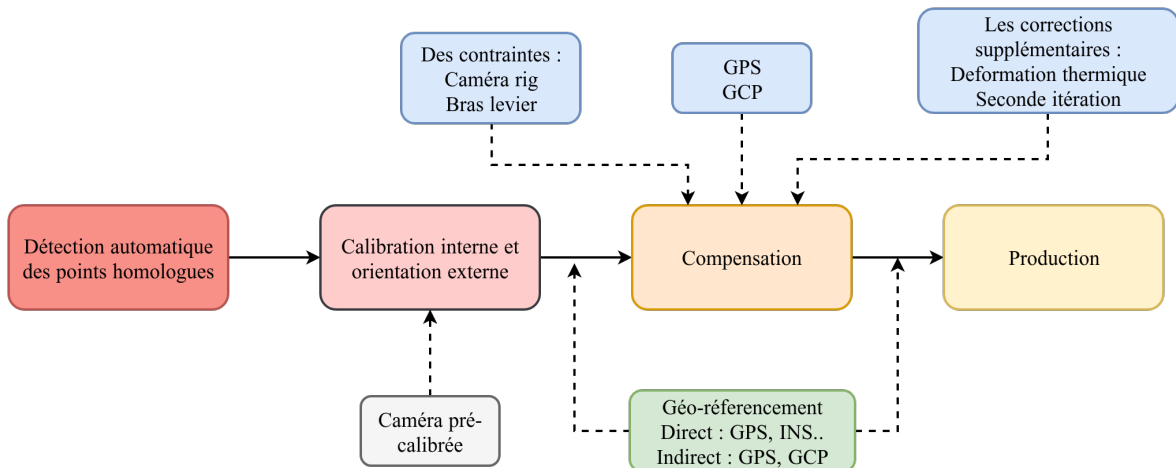


FIGURE 1.3 – Processus de traitement de photogrammétrie numérique.

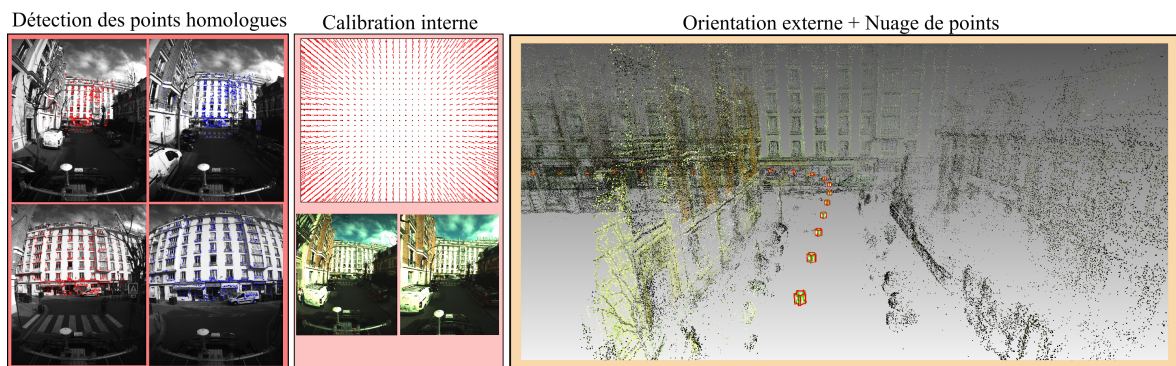


FIGURE 1.4 – Illustration des étapes du processus de traitement de photogrammétrie numérique. **Gauche** : extraction des points homologues, **Milieu** : calibration, **Droite** : nuage de points et orientation externe

Les solutions automatisées les plus utilisées en photogrammétrie et SfM sont présentées dans [Moulon, 2014; Tournadre, 2015].

1.2 La photogrammétrie à l'IGN

1.2.1 La recherche à l'IGN

L'IGN (Institut National de l'Information Géographique et Forestière) est un établissement public à caractère administratif français. L'institut a été fondé le 26 Juin 1940, avec pour principales missions les travaux de vocation pour l'information géographique de la France. Les activités principales de l'IGN aujourd'hui sont :

- la géodésie et le nivellement ;
- l'imagerie aérienne et satellitaire du territoire national ;
- le référentiel à grande échelle : BD ORTHO, BD TOPO, RGE ALTI, BD ADRESSE, BD PARCELLAIRE ;
- l'inventaire forestier ;
- la cartographie ;
- le service web Géoportail¹⁴ ;
- la formation et l'enseignement à l'ENSG¹⁵ ;
- la recherche ;
- la défense ;
- le patrimoine.

La recherche à l'IGN est composée actuellement de cinq équipes : COGIT¹⁶, LAREG¹⁷, LOEMI¹⁸, MATIS¹⁹ et le LIF²⁰. Les quatre premières équipes sont réunies dans le laboratoire LaSTIG²¹. Cette thèse se déroule au sein de l'équipe LOEMI.

L'équipe LOEMI est en charge du développement des instruments spécifiques aux travaux de l'IGN (voir la figure 1.5). Les activités du laboratoire s'articulent autour de deux axes principaux :

- photogrammétrie : développement de la caméra numérique aérienne V1 et V2. Depuis 2009, la caméra photogrammétrique légère Camlight est développée pour l'acquisition par drone et la photogrammétrie terrestre ;
- métrologie : les GéoCubes²² ont pour but de mesurer et de surveiller les déformations.

L'équipe des Travaux Spéciaux du service géodésique et nivellement de l'IGN utilise les instruments de l'équipe LOEMI et la chaîne de traitements MicMac-Apéro pour effectuer les acquisitions photogrammétriques. Les Travaux Spéciaux font toutes sortes de chantiers de métrologie micro-géodésique pour des besoins internes ou externes à l'IGN.



FIGURE 1.5 – De gauche à droite : Caméra aérienne V1, V2, Caméra légère Camlight et GéoCube

14. <https://www.geoportail.gouv.fr/>

15. <http://www.ensg.eu/>

16. Le laboratoire cartographique et géomatique

17. Laboratoire de recherche en géodésie

18. Le laboratoire d'optoélectronique, de métrologie et d'instrumentation

19. Laboratoire de méthodes d'analyses et de traitement d'images pour la stéréorestitution

20. Laboratoire d'inventaire forestier, situé à Nancy

21. Laboratoire des Sciences et Technologies de l'Information Géographique

22. <http://loemi.recherche.ign.fr/pdf/brochureGeocube1.pdf>

1.2.2 Chaîne de traitements photogrammétrique à l'IGN

Toutes les méthodes et recherches développées durant cette thèse sont intégrées à la chaîne de traitement photogrammétrique MicMac [Deseilligny and Clery, 2011; Rupnik et al., 2017]. Le choix de cette solution vient de plusieurs raisons :

- logiciel libre et open source ;
- documentation détaillée destinée aux utilisateurs et aux développeurs ;
- bibliothèque de base de traitements photogrammétriques et de calcul scientifique ;
- paramétrage fin de toutes les étapes de traitement ;
- facilité d'extraction des résultats intermédiaires ;
- peut exécuter des étapes de traitement indépendamment ;
- il est possible de faire des modifications au coeur du logiciel, l'équipe de développement se situant à l'IGN.

Le logiciel MicMac est développé depuis 2003 pour les besoins de l'IGN, au départ par Marc Pierrot-Deseilligny. L'histoire du logiciel est résumée dans les points suivants :

- 2003 : implémentation d'un algorithme d'appariement d'images pour la génération de modèles numériques d'élévation urbains et de l'auto-calibration de la caméra ;
- 2005 : mise en place d'une interface XML, le logiciel s'appelle MicMac ;
- 2007 : dépôt en open source de MicMac ;
- 2008 : développement du module Apéro pour l'orientation d'images ;
- 2010 : sessions de formations et développement d'interfaces simplifiées en ligne de commande ;
- 2012 : projet Culture 3D, migration sous Windows, Mac OSX, distributions binaires, ajout des calculs sur GPU²³ ;
- 2018 : MicMac v2 est démarrée.

MicMac est utilisé comme outil de recherche, de développement et de production photogrammétrique à l'IGN et à l'ENSG, et dans plusieurs laboratoires et projets de collaborations avec MAP-CNRS, CEREMA, INRAP, IPGP, IFFSTAR, Vinci Construction Terrassement SIXENSE MAPPING, etc. La priorité de l'IGN comme celle de MicMac, est la précision de localisation 3D et d'orientation d'image. Il existe plusieurs études de comparaison entre MicMac et les autres principaux solutions photogrammétriques commerciales ou libres dans [Haala, 2014; Moutinho, 2015; Murtiyoso et al., 2018].

Du point de vue de l'utilisateur, MicMac fournit différents niveaux d'accès selon les types d'interfaces. Il y a deux couches d'outils pour l'utilisateur : la couche d'outils de base qui est paramétrée par un fichier XML, et la couche d'interfaces simplifiées qui est paramétrée par des lignes de commande. L'utilisateur expert peut ajuster à très bas niveau des étapes de traitement, par exemple l'ordre de libération des inconnues dans l'ajustement des faisceaux, le choix de l'image initiale dans l'estimation de pose ou la pondération des observations, etc. Une vue générale sur les outils et les interfaces est présentée en figure 1.6.

23. Graphic Processing Unit

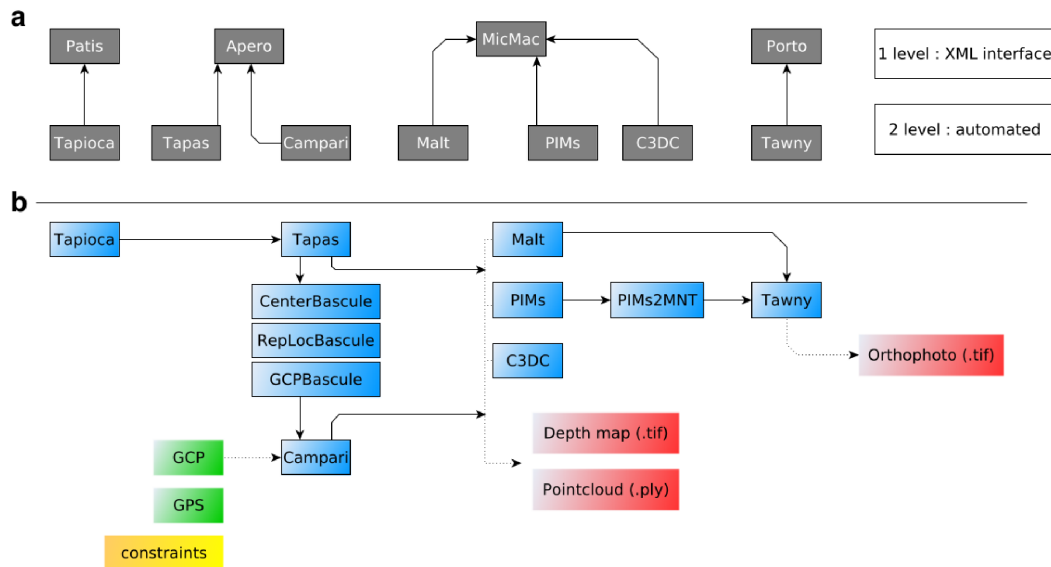


FIGURE 1.6 – Une vue générale des outils et des niveaux d’interface de la chaîne MicMac [Rupnik et al., 2017].



FIGURE 1.7 – Quelques travaux réalisés avec la chaîne MicMac par les étudiants et l’équipe DIAS de l’ENSG

1.2.3 Contexte de la thèse

Cette thèse se déroule dans le cadre de la réponse à un appel d’offre pour la reconstruction 3D et la visualisation virtuelle des égouts de Paris. Plusieurs techniques y compris l’utilisation de photogrammétrie, de lidar mobile, de lidar terrestre et d’autres méthodes sont discutées. En considérant les critères de temps d’acquisition, de complexité, de budget, de facilité à déployer et à entretenir l’instrumentation retenu, l’IGN a décidé d’utiliser une méthode purement photogrammétrique. L’équipe LOEMI du laboratoire LaSTIG développe un système multi caméras pour l’acquisition. Lors de la réponse à cet appel, plusieurs essais ont été effectués sur le terrain pour lesquels une forte dérive avec les méthodes de l’état de l’art a été constatée. L’objectif de cette thèse est de corriger ces effets de dérives qui rendent les résultats inexploitable en terme de précision géométrique.

Divers essais ont été effectués sur le terrain, et présentent des forte dérives, corrigées grâce aux travaux de cette thèse.

1.3 Présentation du sujet de thèse

1.3.1 Problématique

Avec la mise en place d'une multitude de solutions photogrammétriques ou SfM commerciales ou libres, la reconstruction 3D d'un objet attire de plus en plus les utilisateurs. Cependant, en dehors de la visualisation, le but initial de la photogrammétrie est la mesure des coordonnées des points dans l'espace objet. La précision est donc un élément prioritaire en photogrammétrie dans certaines applications de métrologie, par exemple la mesure de déformation en géosciences ou la reconstruction de haute précision dans l'industrie. Malgré la facilité d'utilisation de certaines solutions photogrammétriques, afin d'obtenir un modèle 3D encore plus précis pour les tâches de mesure et de surveillance spécifiques, le suivi d'un processus d'acquisition strict ainsi que des méthodes de traitement innovantes hors-standard sont nécessaires.

Les mesures 3D obtenues à partir des images résultent de plusieurs étapes successives : l'extraction de points homologues²⁴, l'étalonnage de caméra²⁵, la mise en place d'images²⁶, la mise en correspondance dense et la triangulation 3D. Du fait de la forte relation entre ces étapes, la précision de mesure des points homologues influe sur l'exactitude de la mise en place des images ainsi que sur la calibration et finalement sur la triangulation des points 3D. L'importance des points homologues dans la chaîne de traitement photogrammétrique numérique est connue depuis longtemps et a été démontrée empiriquement grâce à plusieurs études [Calantropio et al., 2018; Honkavaara and Jaakkola, 1998; Lingua et al., 2009].

Un défaut de précision ou une distribution irrégulière des points homologues dégradent l'estimation de poses ainsi que la calibration précise des optiques. Cet effet peut être observé notamment sur les scènes longues linéaires où les erreurs systématiques s'accumulent dans le processus d'estimation de poses et causent des dérives importantes.

La reconstruction 3D des scènes longues s'effectue souvent dans les endroits couverts et les galeries souterraines comme les sous-sols, où des limitations peuvent potentiellement apparaître :

- conditions de lumière médiocres ;
- impossibilité d'ajouter des observations topométriques ;
- indisponibilité du signal GNSS ;
- conditions de prise de vue difficiles ;
- relief important de la scène.

Ces limitations causent les problèmes suivants :

- un manque d'observations sur le terrain pour compenser les erreurs relatives (*outer-constraint*) ;
- une grande part des points homologues mesurés imprécisément ;
- des erreurs systématiques qui résultent en une forte dérive ;
- imprécision de la calibration interne par auto-calibration.

Le besoin d'une méthode pour améliorer la précision de mise en place des images et compenser les erreurs systématiques à partir des contraintes images est une réponse à ces limitations.

24. points homologues, tie-points, points de transferts

25. calibration, orientation interne

26. estimation de poses, orientation externe

1.3.2 Objectif de la thèse

L’objectif de la thèse est d’améliorer la précision de localisation 3D. Il s’agit d’une priorité dans tous les relevés photogrammétriques de l’IGN, surtout dans le cas particulier de la géométrie des longs linéaires. De plus, les résultats peuvent s’appliquer pour tous types d’acquisitions.

Les applications des résultats de cette thèse dans le cadre des travaux de l’IGN sont variées : reconstruction 3D des égouts, des tunnels, du corridor mapping ou localisation de véhicule de cartographie urbaine Stéréopolis [Cannelle, 2013].

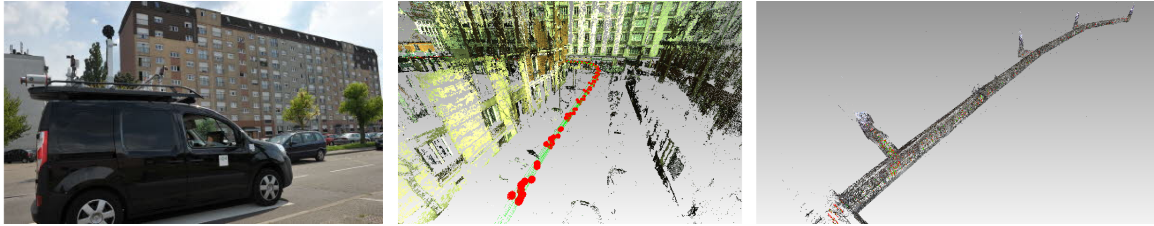


FIGURE 1.8 – Des scénarios longs linéaires : la cartographie urbaine avec la plateforme Stéréopolis, les égouts de Paris

Notre motivation peut alors s’exprimer en deux points principaux :

- correction des erreurs systématiques qui causent souvent la dérive des acquisitions photogrammétriques ;
- amélioration de la précision d’orientation (interne et externe) des images sans utiliser de contraintes externes.

Notre nouvelle méthode est développée et intégrée dans la chaîne de traitement photogrammétrique MicMac de l’IGN.

1.3.3 Description générale de la méthode proposée

Pour limiter et corriger les erreurs de mise en place d’images, plusieurs recherches se sont focalisées sur l’utilisation des points homologues dits “de qualité photogrammétrique”. Il s’agit des points homologues dont l’erreur de reprojection est faible, de bonne distribution dans l’espace image et objet et de multiplicité maximale.

Le travail de recherche dans le cadre de cette thèse s’attelle à améliorer ces aspects en proposant une méthode de raffinement de orientation externe/interne par une seconde itération de traitement. À partir des résultats issus de traitements photogrammétriques utilisant une chaîne de traitement classique, notre méthode génère automatiquement un nouvel ensemble de points homologues²⁷ optimisés pour les calculs photogrammétriques, c’est à dire selon les aspects suivants :

- la précision de mesure image ;
- une haute multiplicité (nombre de fois qu’un point apparaît sur des images différentes) ;
- une distribution optimale dans l’espace image et objet.

Cette méthode d’extraction des nouveaux points homologues est nommée “*Second Iteration to Refine image Orientation with improved Tie-Point – (SIROP)*”. Cette méthode peut donc être considérée comme un post-traitement (voir la figure 1.9) d’amélioration de la précision d’un traitement classique. La méthode est détaillée dans le chapitre 4.

27. paquet de points homologues, tie-point pack

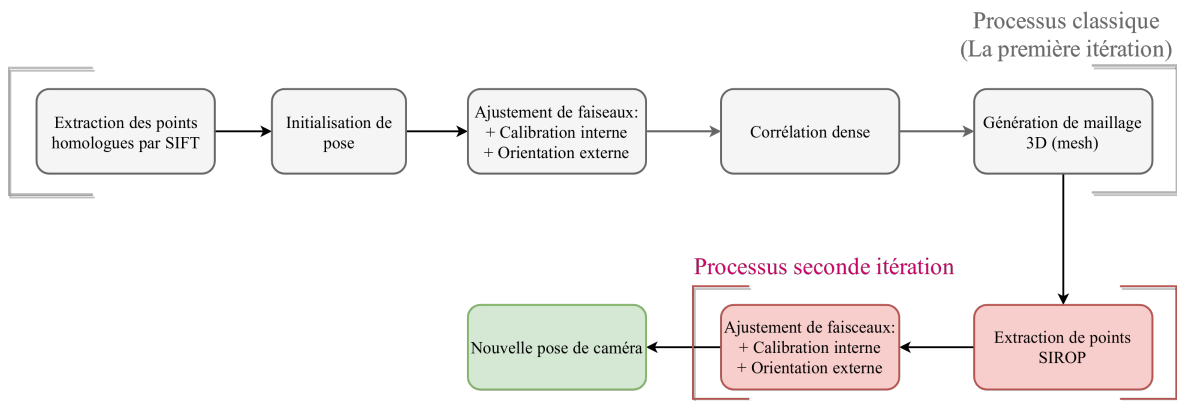


FIGURE 1.9 – Diagramme de flux du processus de seconde itération *SIROP*.

Chapitre 2

Contributions, activités et organisation du document

Sommaire

2.1 Contributions et activités	29
2.1.1 Publications	29
2.1.2 Activités scientifiques et académiques	30
Formations dispensées	30
Formations suivies	30
Travaux terrain	30
Logiciel	30
2.2 Organisation du document	31

2.1 Contributions et activités

2.1.1 Publications

- “Second Iteration of Photogrammetric Processing to Refine Image Orientation with Improved Tie-Points” - Journal MDPI Sensors 2018, Volume 18, Issue 7 – Nguyen Truong Giang, Jean-Michaël Muller, Ewelina Rupnik, Christian Thom, Marc Pierrot-Deseilligny.
- “Géo localisation précise basée image, une approche de type seconde itération du processus photogrammétrique” - Présentation aux 27èmes Journées de la Recherche IGN – Mars 2018 – Marne-la-Vallée.
- “Second iteration of photogrammetric pipeline to enhance the accuracy of image pose estimation” - l’article de conférence en International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 42 – Poster à ISPRS Workshop EuroCOW 2017 – Hannover.
- “Une deuxième itération du processus photogrammétrique pour améliorer la précision de la mise en place des images” - Poster aux 26èmes Journées de la Recherche IGN – Mars 2017 – Marne la Vallée.
- “IGN photogrammetric camera for UAVs : concept and results” - M. Daakir, Y. Zhou, J-P. Souchon, J-M. Muller, G. Nguyen, M. Pierrot-Deseilligny, C. Thom, O. Martin – DGPF Meeting 2017 : UAV Metric Camera session, Würzburg, 8-10 mars 2017.

- “Amélioration de la mise en place par corrélation fine sur l’approximation géométrique de surface” -Présentation au colloque SFPT ”Photogrammétrie et télédétection : vers la convergence” – Mars 2017 – ENSG Marne-la-Vallée.
- “Amélioration de la mise en place d’image” - Présentation à la Journée des Doctorants MSTIC – Juin 2016 - Marne la Vallée

2.1.2 Activités scientifiques et académiques

Formations dispensées

- Tutoriel : “Reconstruction 3D” – cycle d’ingénieur 2ème année en Informatique – ESIGE Marne-La-Vallée – Mai 2018 ;
- Tutorat : “Mise en place d’un supercalculateur lowcost basé sur Raspberry PI” – cycle d’ingénieur 2ème année avec Mehdi Daakir– ENSG Marne-La-Vallée – Octobre 2016 ;
- Tutorat : “Modélisation 3D des sous-sols de l’ENSG” – cycle de lycée professionnelle 3ème année – ENSG Marne La Vallée – Avril 2017 ;
- Tutorat : projet stage de fin d’étude cycle de licence professionnelle ”Qual-Mac, outil Python pour comparer et évaluer la processus photogramétrique par MicMac” – cycle de lycée professionnelle 3ème année – ENSG Marne-La-Vallée – Avril 2017 ;
- Tutoriel : “Reconstruction 3D” – cycle d’ingénieur 2ème année en Informatique – ESIGE Marne-La-Vallée – Avril 2016 ;
- Tutoriel : “Reconstruction 3D” – formation continue PMIC – ENSG Marne-La-Vallée – Juin 2016 ;
- Tutoriel : “Reconstruction 3D” – formation continue PMIC – ENSG Marne-La-Vallée – Octobre 2016 ;
- Tutoriel : “Reconstruction 3D du château de Murol” – Encadrement des élèves Master PPMD – Septembre 2017.

Formations suivies

- “Reconstruction 3D du château de Chambord” – Octobre 2015 ;
- “Photogrammétrie” – cours avec le Master PPMD dispensé par Marc Pierrot Deseilligny – Octobre 2015.

Travaux terrain

- Participation au travail topographique et photogramétrique ”Surveillance de digue”– La digue de La Palière CNR – Septembre 2017 ;
- Participation aux acquisitions des égouts de Paris – Décembre 2015.

Logiciel

- Outil d’extraction de points homologues pour raffiner la mise en place d’images dans la chaîne MicMac/Apéro. (voir partie “New procedure to enhance images orientation precision by TaskCorrel and TiejTri” dans la documentation MicMac [Deseilligny et al., 2018])

2.2 Organisation du document

Ce manuscrit est organisé en 6 chapitres qui présentent la description, les détails et la partie expérimentale d'une méthode d'amélioration de la précision de géo-localisation des acquisitions photogrammétriques. Le chapitre 1 décrit le contexte scientifique dans lequel ce travail est réalisé ainsi que son objectif. Le chapitre 2 liste les contributions et activités scientifiques au cours de la thèse. Le chapitre 3 présente les différentes thématiques dans l'état de l'art directement liées au sujet. Le chapitre 4 détaille l'algorithme en profondeur avec les explications et les différentes étapes afin de le rendre compréhensible auprès des lecteurs et faciliter la qualification de la nouveauté et de la précision de la méthode proposée. Le chapitre 5 présente les résultats expérimentaux effectués sur différents scénarios afin d'évaluer l'efficacité de la méthode. Le chapitre 6 termine ce manuscrit par une conclusion et des perspectives pour le travail présenté.

Chapitre 3

État de l'art

Sommaire

3.1	Introduction	33
3.2	Sources d'erreurs en photogrammétrie	34
3.3	Points homologues	37
3.3.1	Le détecteur et descripteur de points d'intérêt	38
3.3.2	L'appariement	40
	Appariement basé caractéristiques visuelles	41
	Appariement basé région	41
	La corrélation	41
	L'appariement par moindres carrés	43
3.4	Conclusion	45

3.1 Introduction

Notre problématique est la correction d'erreurs d'estimation de pose dans le relèvement photogrammétrique, surtout dans une configuration de scène linéaire longue. Ce problème concerne l'erreur systématique, qui provient principalement des paramètres internes imprécis (la calibration) de la caméra. Cette forme d'erreur peut se propager et s'accumuler dans le processus de l'estimation de poses. Une déformation dans le modèle 3D du relevé photogrammétrique peut en découler, surtout dans l'acquisition de type corridor. Pour cela, nous allons présenter des études sur les sources d'erreurs dans l'estimation de poses.

Le processus de l'estimation de poses et l'auto-calibration de caméra utilise principalement les points homologues comme observations d'entrée, d'où l'influence forte des caractéristiques de ces points homologues sur la qualité du résultat. La méthode que nous avons développée s'attaque à cet aspect du problème en améliorant la qualité des points homologues. Tout d'abord, une partie de cet état de l'art présente les méthodes de détection des points homologues les plus populaires utilisées en photogrammétrie, ainsi que leurs avantages et leurs inconvénients. On présente également les études des caractéristiques des points homologues dans le domaine de la photogrammétrie.

Dans l'implémentation de la chaîne MicMac utilisée lors de nos expérimentations, *SIFT* est choisi comme la méthode d'extraction de points d'intérêts et de descripteurs. Il s'agit d'une méthode de détection et de description des caractéristiques visuelles (point saillant, point d'intérêt) sur image. Cette méthode est robuste aux transformations

géométriques de type : translation, rotation, homothétie. Chaque point saillant est associé à son vecteur de descripteur qui décrit ses caractéristiques. Les descripteurs de ces points saillants sont appariés ensuite sur les différentes images pour créer les points homologues. MicMac utilise *ANN*¹ comme méthode d'appariement de descripteurs. *SIFT* et *ANN* sont les méthodes utilisées pour extraire les points homologues dans le traitement photogrammétrique classique, qui fournissent les entrées pour la méthode proposée. Un état de l'art des méthodes de détection et d'appariement de points saillants basées sur les caractéristiques visuelles² est présenté brièvement dans les parties suivantes.

Notre méthode d'extraction des points homologues *SIROP* utilise l'information 3D qui vient du premier traitement classique pour corriger la différence de géométrie entre les images et prédire la position de la correspondance. L'appariement entre les points *SIROP* est effectué sur les régions re-échantillonnées à la même géométrie. Les détecteurs et descripteurs de points saillants de type invariants à la rotation et à l'échelle comme *SIFT* ne sont pas adaptés à notre objectif. *SIROP* utilise la corrélation et l'appariement par moindres carrés³ pour maximiser la précision de localisation des points. L'état de l'art sur les méthodes d'appariement basées région est aussi présenté.

En conclusion, dans la partie suivante sont présentées les thématiques listées ci-dessous :

1. Les sources d'erreurs et les méthodes pour améliorer la précision de l'estimation de poses.
2. La détection de points d'intérêt.
3. Les méthodes d'appariement basées région.

3.2 Sources d'erreurs en photogrammétrie

Afin de proposer une méthode d'amélioration de la qualité du traitement, on étudie tout d'abord les sources d'erreurs et leur impact sur les résultats obtenus par traitement photogrammétrique. Dans cette partie, des études sur cette thématique sont présentées.

Parmi les sources d'erreurs possibles, les paramètres internes de la caméra issus de sa calibration (point principal, coefficients de distorsion, focale, etc.), le protocole d'acquisition (taille du pixel au sol⁴, la longueur de la ligne de base⁵, le recouvrement, l'angle d'intersection des rayons, l'angle d'incidence, etc.) et la chaîne de traitements peuvent être cités. Les erreurs de mesures associées à la photogrammétrie peuvent être classées dans ces grandes catégories :

- l'erreur systématique causée par l'imprécision des paramètres internes de la modélisation de caméra ;
- l'erreur systématique causée par le protocole d'acquisition (la géométrie du réseau de poses⁶) ;
- l'erreur aléatoire causée par la limite de précision des appareils (bruit de caméra, taille de photosite, etc.).

1. Approximate nearest neighbor search
2. feature based
3. Least Square Matching (LSM)
4. GSD
5. baseline
6. camera network

Portons notre intérêt sur l'étude de l'erreur systématique causée par les paramètres internes de la caméra. En métrologie, l'erreur systématique signifie un biais, ce qui apparaît de manière constante. Dans un système de mesures basé image comme la photogrammétrie, l'erreur systématique causée par l'imperfection de la modélisation de la géométrie de la caméra est principalement due à ces trois composantes :

- modèle de distorsion mal adapté ;
- mauvaise estimation des paramètres de distorsion de l'objectif ;
- la stabilité interne de la caméra.

L'erreur de mesure causée par la distorsion de l'objectif déplace un pixel hors de sa position théorique sur le plan image. Cette distorsion peut être approximée par la fonction de correction de la distorsion radiale ou tangentielle sous la forme d'une fonction radiale de degré élevé (2, 4, 6 ou plus haut – voir l'annexe C). L'estimation imprécise des coefficients de cette fonction de distorsion peut causer la déformation du modèle 3D à la fin du traitement, décrit dans la littérature sous le nom d'effet courbure/*bowl effect*/*verticle dome*. Cet effet est étudié précisément dans [Eltner and Schneider, 2015; James and Robson, 2014; Paul-Henri FAURE and Jacques BEILIN, 2014; Wackrow et al., 2008].

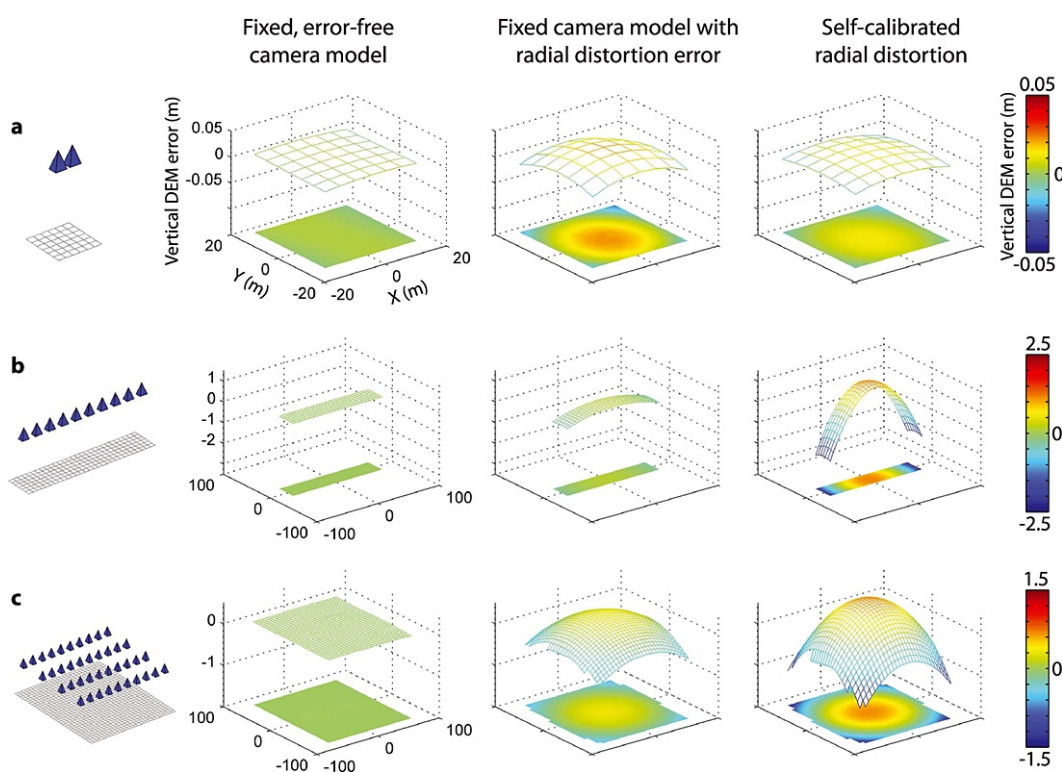


FIGURE 3.1 – La simulation de l'effet de courbure sur plusieurs scénarios et les paramètres internes de la caméra présentés dans [James and Robson, 2014]

Il existe plusieurs études ayant pour but de limiter ou corriger cet effet. Des approches sur le protocole d'acquisition⁷ sont étudiées. [Nocerino et al., 2014] propose de rajouter 2 prises de vues obliques dans des directions opposées toutes les 4 images nadir, quel que soit le type d'acquisition, UAV⁸ ou terrestre. Pour les acquisitions de type UAV, il recommande d'effectuer un vol oblique à hauteur réduite (voir la figure 3.3). [Eltner

7. configuration, network geometry

8. Unmanned aerial vehicle

et al., 2016] utilise les outils de correction de la distorsion d'image comme AgiSoft-Lens ([Kaiser et al., 2014]). Ces images corrigées sont ensuite traitées avec les outils de SfM basiques comme VSfM et Bundler en désactivant toutes les corrections implémentées dans ces outils. [James and Robson, 2014] propose aussi de rajouter des vues obliques (voir la figure 3.3) pour limiter cet effet. [Wackrow et al., 2008] montre aussi qu'une configuration d'images convergentes peut réduire significativement l'erreur systématique causée par l'estimation imprécise des paramètres de la distorsion. Par contre, multiplier les images n'améliore pas forcément la précision du résultat final, comme le montre l'étude de [Micheletti et al., 2015].

L'approche sur les paramètres internes de la caméra (la calibration) est aussi fortement étudiée par Vincent Tournadre [Tournadre, 2015; Tournadre et al., 2015, 2014]. Ces études montrent que l'effet de courbure peut être corrigé en utilisant un modèle de caméra plus sophistiqué (modèle R15 dans la chaîne MicMac – voir Figure 3.2).

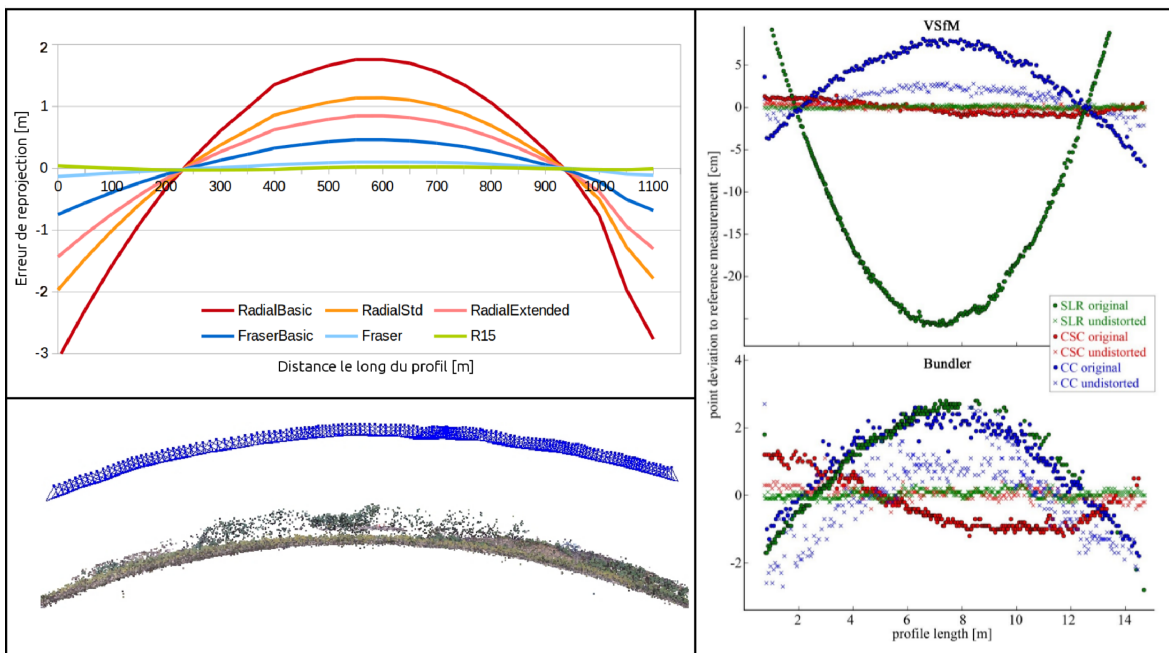


FIGURE 3.2 – **Droite** : illustration de l'effet de courbure par comparaison entre le traitement qui utilise des images corrigées (undistorted) et originales (original) [Eltner et al., 2016], **Bas Gauche** : effet causé par la distorsion radiale mal estimée dans la configuration critique [Wu, 2014], **Haut Gauche** : les différences d'effet de courbure selon les modèles de caméra, et la correction en utilisant un modèle plus sophistiqué (R15 dans la chaîne MicMac) [Tournadre, 2015]

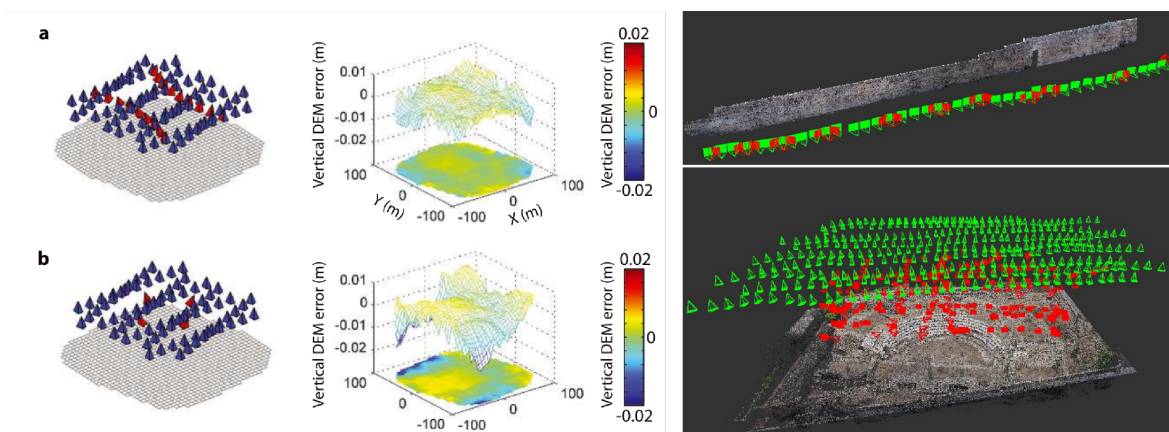


FIGURE 3.3 – La correction de l'effet de courbure par l'ajout d'images obliques proposée par **Gauche** : [James and Robson, 2014] et **Droite** : [Nocerino et al., 2014].

Le choix du modèle de caméra utilisé a également une influence sur la stabilité des paramètres internes (la géométrie interne de la caméra) dans le processus de calibration. La désactivation de toutes les technologies implémentées dans la caméra comme la stabilisation optique/capteur, le zoom, la mise au point automatique lors de l'acquisition, et enfin l'utilisation d'une caméra métrique sont considérés comme des facteurs limitant les erreurs. [Gervais, 2017; Thomas Kraft and Ralf Berger, 2017] présentent des caméras métriques avec des caractéristiques optimisées pour la photogrammétrie ainsi que l'étude de leur efficacité.

L'équipe LOEMI du laboratoire LaSTIG de l'IGN développe une caméra photogrammétrique légère (CamLight) avec un capteur plein format dans un boîtier compact, qui ne pèse que 300g, objectif 35mm inclus [Martin et al., 2014]. Cette caméra est mise en oeuvre dans les travaux de [Audi et al., 2017; Daakir et al., 2017; Tournadre, 2015].

Des approches consistant à rajouter des observations externes comme des GCPs (points terrains) sont aussi étudiées par plusieurs chercheurs [Eltner et al., 2016; Grün, 1978; Nocerino et al., 2013]. [Koci et al., 2017] montrent, comme l'on s'y attendait, que plus le nombre de GCPs utilisés dans l'ajustement de faisceaux est grand, plus le résultat obtenu est précis mais les coûts de travaux sur le terrain augmentent. Il est également parfois impossible de rajouter des GCPs. Par ailleurs, la question sur le nombre et la distribution optimale des GCPs reste ouverte. Les études de Eltner et James [Eltner and Schneider, 2015; James and Robson, 2012, 2014] montrent que plus il y a de GCPs et mieux ils sont distribués dans l'espace de la scène, plus le résultat final de l'ajustement de faisceaux est meilleur en terme de précision.

Considérée comme la première étape de traitement et l'observation principale pour l'estimation de pose (parce qu'il sont issus de calculs 100% automatiques), l'extraction des points homologues joue un rôle très important dans la qualité d'orientation d'image. Plus les points homologues sont précis, meilleure est la mise en place des images. Dans une revue assez exhaustive, [Gruen, 2012] conclut que les méthodes d'appariement les plus précises entre les points sont celles basées région, y compris la cross-corrélation et l'appariement par les moindres carrés. Ces méthodes peuvent atteindre des précisions de l'ordre de 0.1 pixel. Par ailleurs, quelques méthodes d'appariement basées caractéristiques visuelles peuvent fournir une précision sub-pixellaire, mais pas au niveau des précisions atteintes par les méthodes basées région. Une combinaison entre la robustesse des méthodes basées caractéristiques et la précision des méthodes basées régions est très intéressante pour les applications de métrologie [Remondino, 2006].

Il existe dans l'état de l'art des approches qui utilisent la simulation de points de vue pour ajouter des points homologues. Les points supplémentaires peuvent faciliter et augmenter la précision de mise en place des images avec des points de vue éloignés. [Rolin et al., 2015] propose une méthode de simulation de points de vue en générant des images virtuelles afin d'ajouter des points *SIFT*. Cette méthode se montre efficace dans le cas de forts changements de points de vue. Ce type de méthodes nécessitent que la scène contienne des plans. [Soontranon et al., 2013] propose une méthode de détection des points *SIFT* sur les images ortho-rectifiées pour densifier les points détectés. Ces approches ne fonctionnent que là où la scène est sans discontinuité.

3.3 Points homologues

L'approche *SIROP* que nous proposons est basée sur l'amélioration de la qualité des points homologues. Nous présentons dans cette partie un état de l'art de la recherche de points homologues dans le contexte photogrammétrique, y compris la détection de points d'intérêt et leur appariement.

3.3.1 Le détecteur et descripteur de points d'intérêt

Le **détecteur** de points d'intérêt a pour but de choisir des caractéristiques visuelles 2D⁹ bien identifiées dans l'image. Ces points d'intérêt sont ensuite décrits avec le descripteur, qui se construit souvent à partir du voisinage de ces points. Ces descripteurs permettent ensuite d'identifier ou d'apparier les points d'intérêt entre les images pour créer des points homologues. Les points homologues sont les observations principales dans la chaîne de traitement photogrammétrique.

Dans le contexte photogrammétrique, on a besoin que les points homologues soient localisés précisément, bien distribués dans l'espace image et qu'ils aient le plus de multiplicité possible. À cause du changement de prise de vue entre les images acquises, l'invariance à certaines transformations comme la rotation, la translation ainsi que l'échelle est souhaitée pour le détecteur.

Plusieurs types de détecteurs 2D de caractéristiques sont utilisés comme :

- les contours : bords, edges, curves (Canny [Canny, 1986], LSD [Von Gioi et al., 2012], etc.) ;
- les coins/*corners* (Harris [Harris and Stephens, 1988], FAST [Rosten, 2006], etc.) ;
- les régions/*blobs* (*SIFT* [Lowe, 2004], SURF [Bay et al., 2008], KAZE [Alcantarilla et al., 2012], MSER [Matas et al., 2004], GLOH [Mikolajczyk and Schmid, 2005], [Lindeberg, 1998], etc.) ;

[Mikolajczyk and Schmid, 2005; Mikolajczyk et al., 2005] étudie la performance des descripteurs des points dans des conditions différentes de géométrie et de photométrie. Les résultats montrent que *SIFT* et sa variante GLOH fournissent la meilleure performance en termes de nombre de leur points et de multiplicité.

[Lingua et al., 2009] étudie la performance de *SIFT* dans les applications photogrammétriques. Il conclut que *SIFT* peut être considéré comme une solution efficace pour l'extraction automatique de points homologues dans l'application photogrammétrique grâce à son grand nombre de points, sa robustesse aux forts changements de géométrie et de photométrie. Par contre, en terme de précision de localisation des points, ce qui est très important en photogrammétrique, *SIFT* n'est pas le plus précis.

Ce problème est étudié dans l'étude de [Remondino, 2006]. C'est à cause du déplacement du centre de la région détecté en raison du changement de vue. L'étude montre aussi que les détecteurs de type corner comme Förstner [Förstner, 1986], Heitger [Heitger et al., 1998] ou SUSAN [Smith and Brady, 1997] combinés avec un appariement basé région donnent un résultat de mise en place d'images plus précis que les détecteurs de type régions comme *SIFT* ou Harris-Affine. En menant des expérimentations dans le cadre de cette étude, l'auteur montre que les points *SUSAN* peuvent permettre d'obtenir une orientation d'images deux fois plus que pour les points *SIFT*. L'auteur teste aussi une approche de combinaison : *SIFT* pour initialiser les paramètres géométriques (transformation affine) et l'appariement par moindres carrés afin d'affiner la localisation

9. features

des points. Il conclut qu'une amélioration d'un facteur 1.3 est obtenue dans le cas. Par contre, la combinaison de *SIFT* et *LSM* est coûteuse en temps de calcul.

[Jazayeri and Fraser, 2008] évalue 3 détecteurs de points d'intérêt de type corner, FAST, SUSAN et Förstner combinés avec un appariement basé région, pour des applications de reconstruction 3D. L'expérience menée dans l'article montre que parmi ces 3 détecteurs, FAST [Rosten, 2006] est le meilleur en raison de sa rapidité, de sa précision de localisation des points et de sa répétabilité.

En complément de la version originale de *SIFT* [Lowe, 2004], il existe plusieurs variantes développées pour améliorer sa performance ou adapter sa fonctionnalité à des applications spécifiques.

En plus des informations locales utilisées par *SIFT* pour former le descripteur, la version *GSIFT* introduite par [Mortensen et al., 2005] ajoute un vecteur représentant la texture globale dans le descripteur. La version *CSIFT* introduite par [Abdel-Hakim and Farag, 2006] utilise l'information couleur dans la phase de détection des points d'intérêt. [Bay et al., 2008] introduit *SURF* en changeant la méthode de calcul pour améliorer la rapidité de *SIFT*.

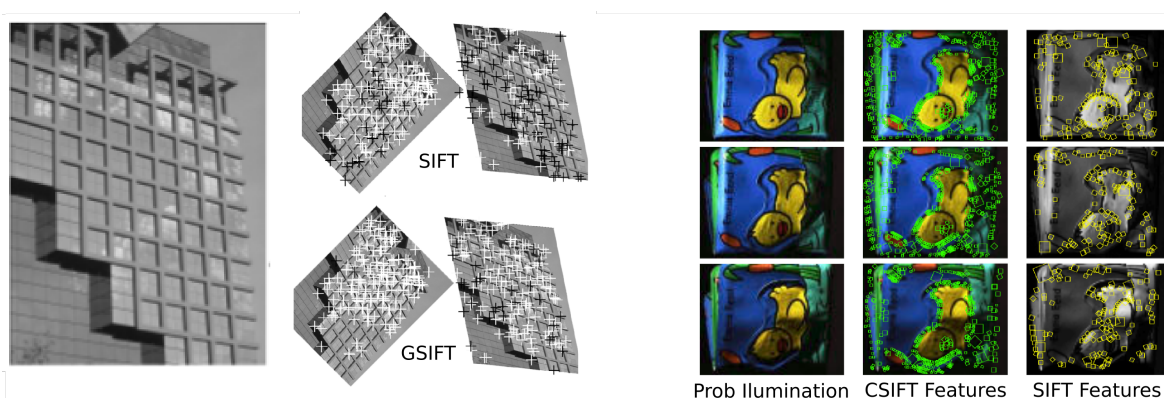


FIGURE 3.4 – **Gauche** : les appariements corrects (croix blanches) en comparant *SIFT* et *GSIFT*, **Droite** : *CSIFT* descripteur plus robuste aux changements d'illumination sur une image couleur.

[Cheng-Yi and Shih-Hong, 2012] propose d'ajouter une valeur d'entropie afin de mieux distinguer les vecteurs descripteurs *SIFT* attribués aux différents points. Il propose aussi de filtrer les points d'intérêt de type *blob* en examinant les valeurs propres de la matrice de Harris. [Sun et al., 2014] propose une implémentation *L2-SIFT* optimisant le calcul pour les grands jeux de données. [Morel and Yu, 2009] prend en compte le changement de géométrie entre les images avec la version *ASIFT*. [Noury et al., 2010; Sur et al., 2011] propose *I-ASIFT*, une méthode d'amélioration de la performance d'*ASIFT* dans les scènes artificielles avec des patterns répétitifs. [Lingua et al., 2009] propose *A²-SIFT*, une version avec la capacité de régler automatiquement le seuil de contraste pour augmenter le nombre de points dans la partie texturée de l'image. Cette idée de modification du seuil de contraste est basée sur les travaux de [Battiatto et al., 2007]. [Tang et al., 2014] propose de supprimer le sous-échantillonnage dans l'espace d'échelle de *SIFT* pour augmenter la précision d'appariement. [Ke and Sukthankar, 2004] propose *PCA-SIFT* en utilisant la compression PCA (*principal component analysis*) pour remplacer l'histogramme 3D (localisation de gradient & orientation) dans le descripteur *SIFT*. [Wu et al., 2013] effectue une évaluation des différentes variantes de *SIFT*. Il conclut que chaque variante a un avantage pour un aspect particulier et qu'il n'y a pas de variante qui améliore de manière remarquable *SIFT*.

[Barazzetti et al., 2010] propose l'outil d'extraction des points *ATiPE* pour les applications photogrammétriques. *ATiPE* ajoute une étape d'élimination des points aberrants en cherchant la meilleure homographie à partir des points *SIFT*. Puis, les couples d'images sont formés et les appariements sont affinés par LSM. Un filtrage spatial final dans l'espace image aide à homogénéiser la distribution spatiale des points. Cette approche montre plusieurs des caractéristiques souhaitées pour les points homologues dans l'application photogrammétrique : la précision de localisation des points, filtrage de points aberrants, une distribution homogène et une grande multiplicité. Cependant, la méthode utilisant des points *SIFT*, elle n'a pas la capacité de combler les lacunes qui apparaissent dans les zones de fortes déformations.

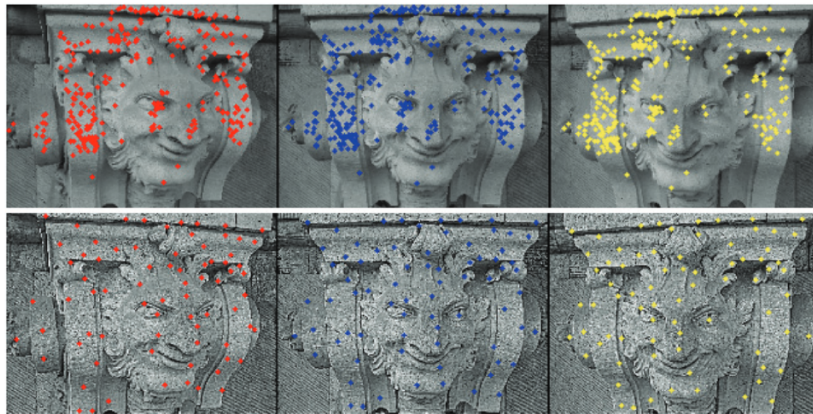


FIGURE 3.5 – La distribution de points est plus homogène dans *ATiPE* [Barazzetti et al., 2010] (**Bas**) en comparaison avec *SIFT* (**Haut**).

Il existe aussi des recherches sur les méthodes de pré-traitement des images avant de détecter les points d'intérêt. [Gaiani et al., 2016] propose un pré-traitement en utilisant le filtre *CBM3D-New* pour diminuer le bruit dans les images. Ensuite, un filtre Wallis est appliqué afin d'augmenter les contrastes locaux. L'article montre qu'avec un seuillage raisonnable, cette méthode permet d'augmenter le nombre de points *SIFT* détectés. [Jazayeri and Fraser, 2008] montre qu'un pré-traitement avec le filtrage Wallis peut augmenter significativement le nombre de points d'intérêt détectés avec FAST et Förstner. Une méthode équivalente est implémentée dans MicMac (nommée SFS). Pour conclure, ces différents traitements servent soit à filtrer les points de liaison, mais dans ce cas-là des lacunes restent, soit à augmenter le nombre de points, mais toujours sans augmenter leur précision. Notre méthode cherche à répondre aux trois critères en même temps.

3.3.2 L'appariement

L'appariement d'images numériques est la technique utilisée pour la recherche de points correspondants entre deux ou plusieurs images. Il existe deux types principaux de méthodes d'appariement : basée région et basée caractéristiques visuelles [Remondino et al., 2008]. Notre extraction de points *SIROP* utilise l'appariement basé région. On a choisi le détecteur de points d'intérêt à maxima/minima locaux en pixel entier et l'appariement entre ces points par corrélation croisée (cross-corrélation) et l'appariement par moindres carrés.

La raison de ce choix est :

- la priorité donnée à la précision de localisation des points ;

- le nombre de points élevé \Rightarrow plus facile d'homogénéiser *a posteriori* la distribution des points ;
- la rapidité de calcul.

La localisation des points comme présentée en section 3.3.1 est meilleure avec les détecteurs de coins comme *Harris*, *FAST*, *SUSAN*, *Förstner* etc. Elle est moins précise avec les détecteurs de type blob comme *SIFT*, *SURF* ([Remondino, 2006]). De plus, l'appariement entre les descripteurs de grande dimension est très coûteux en temps de calcul. D'après [Gruen, 2012], quelques méthodes d'extraction de points homologues de type basées caractéristiques visuelles arrivent à une précision subpixelaire, mais la précision n'est pas toujours suffisante avec les méthodes basées région.

SIROP utilise la prédiction de localisation de points et la correction de la déformation perspective entre les régions d'images grâce au maillage 3D (voir le chapitre 4) de la scène. Il n'est donc pas nécessaire d'utiliser des descripteurs invariants en rotation et à l'échelle comme *SIFT*. Notre priorité est de délivrer un nouvel ensemble de points homologues affinés, adaptés au cahier des charges photogrammétrique, principalement la précision de localisation des points. C'est pourquoi l'appariement type basé région et la détection de caractéristiques visuelles type coins sont choisis.

Appariement basé caractéristiques visuelles

L'appariement basé caractéristiques visuelles a besoin dans un premier temps que les points 2D soient extraits dans les images, puisque ces points se voient assigné un descripteur. Un descripteur est un vecteur de haute dimension, qui sauvegarde les informations extraites dans la région voisine autour de ce point. Le descripteur *SIFT* dans sa version originale est de dimension 64. *GLOH* [Mikolajczyk and Schmid, 2005] est une variante de *SIFT* avec un descripteur de 128 dimensions. L'appariement se fait ensuite entre ces descripteurs. À cause de la haute dimension du vecteur descripteur, leur appariement est une tâche très coûteuse en calcul. L'approche la plus simple et la plus coûteuse est la *brute force*. Cette approche mesure la similarité entre toutes les combinaisons des couples de descripteurs possibles pour choisir le plus similaire. Les méthodes de recherche de plus proches voisins¹⁰ sont considérées comme plus précises et efficaces que la *brute force*. Le meilleur étant *kd-tree* [Bentley, 1975; Friedman et al., 1977]. Par contre, la performance de ce type de méthodes diminue rapidement avec l'augmentation du nombre de dimensions des données. Pour réduire le temps de calcul en conservant la précision de recherche, les méthodes de recherche approximative de plus proches voisins sont proposées [Arya et al., 1998; Beis and Lowe, 1997; MUJA and LOWE, 2009; Silpa-Anan and Hartley, 2008]. Il existe plusieurs revues des méthodes d'appariement basées caractéristiques visuelles dans la littérature [Berrani, 2004; Muja, 2013; Muja and Lowe, 2014]. Dans la chaîne de traitement photogrammétrique MicMac, la méthode de recherche approximative des plus proches voisins (ANN) pour appairer des points *SIFT* est utilisée, avec l'implémentation de [Arya et al., 1998]. Les contributions nouvelles de cette thèse et la méthode proposée ne concernent pas l'appariement basé caractéristiques visuelles. L'état de l'art sur ce point est donc moins détaillé.

Appariement basé région

Les méthodes d'appariement basées région comparent les régions autour des points d'intérêt pour les appairer. Elles utilisent directement le niveaux de gris du pixel

10. NN – nearest neighbor search

de l'image originale (ou améliorée). Les méthodes les plus connues sont la corrélation d'images numériques¹¹ et l'appariement par moindres carrés. Ces méthodes peuvent atteindre une précision sub-pixellaire (de l'ordre de 0.1 pixel ou même moins). *LSM* est un processus non-linéaire, pour lequel une très bonne approximation est nécessaire. Les parties suivantes présentent un état de l'art de ces méthodes.

La corrélation Cette technique est très populaire en photogrammétrie pour chercher des points homologues entre les images afin de les orienter. La première méthode d'appariement est développée dans les années 50 par [Hobrough, 1959] avec un circuit électronique corrélateur dans les machines de restitution StereoMat. Avec l'apparition de l'imagerie numérique, plusieurs variantes de corrélation ont vu le jour comme [Briechle and Hanebeck, 2001; Helava, 1978]. La corrélation a pour but de mesurer la similarité entre des régions des images (ou vignettes, imagettes). Il s'agit de fenêtres extraites autour des points d'intérêt de l'image source et autour des points candidats de l'image cible à appairer. Une région de recherche peut être définie autour des points candidats et la fenêtre balaye dans cette région pour chercher l'endroit ayant la meilleure ressemblance au sens de la corrélation (voir la figure 3.6). Cette méthode est rapide en temps de calcul.

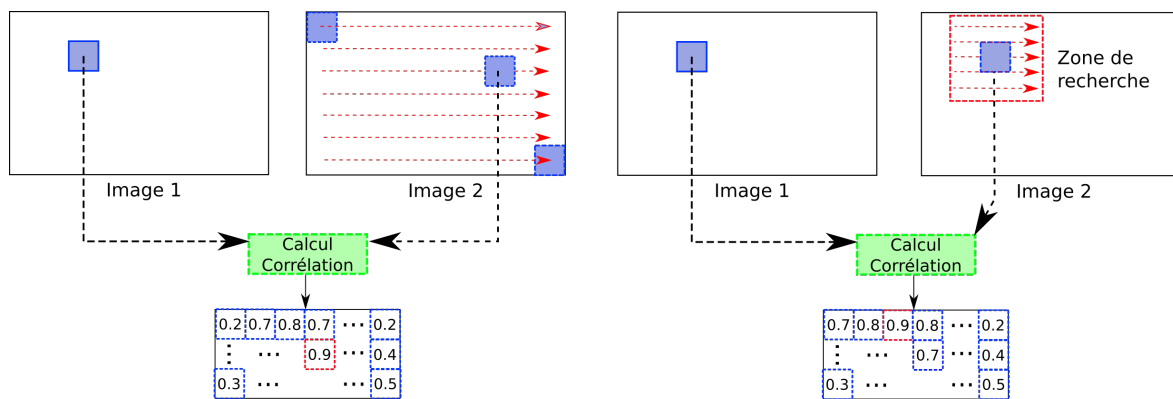


FIGURE 3.6 – La corrélation : **Gauche** : recherche d'un endroit sur l'image 2 le plus similaire à la fenêtre de l'image 1. **Droite** : la recherche peut être réduite à la région autour du point d'intérêt.

Le désavantage de la corrélation comme des techniques d'appariement basées région vient de l'utilisation des niveaux de gris des pixels des images. Si la région est peu texturée ou texturée de manière répétitive, la corrélation peut donner des erreurs. De plus, les occlusions et les déformations entre les images liées à la perspective peuvent mettre la méthode en échec. L'hypothèse d'une région plate est faite dans ce type de méthodes. La taille de la vignette est donc un paramètre important. Plus la vignette est grande, plus elle est sensible à la déformation d'image et aux occlusions. Plus la vignette est petite, plus elle est sensible au bruit dans l'image ([Ehlers, 1985], [Siebert et al., 2007]). La corrélation est également sensible aux problèmes de radiométrie (changement d'illumination, réflexion, etc..).

Il existe un grand nombre de méthodes de calcul pour mesurer la corrélation [Aschwanden and Guggenbuhl, 1992; Chambon and Crouzil, 2003; Roma et al., 2002]. Notons I_1, I_2 deux vignettes dont calculer la corrélation, de taille $(2M + 1) \times (2N + 1)$. Soient leurs pixels associés (x_1, y_1) et (x_2, y_2) . La méthode de mesure la plus simple est :

11. Digital Image Correlation (DIC)

La somme des distances au carré *SSD* :

$$SSD = \sum_{i=-M}^M \sum_{j=-N}^N (I_1(x_1 + i, y_1 + j) - I_2(x_2 + i, y_2 + j))^2 \quad (3.1)$$

La somme des valeurs absolues des distances *SAD* :

$$SSD = \sum_{i=-M}^M \sum_{j=-N}^N |(I_1(x_1 + i, y_1 + j) - I_2(x_2 + i, y_2 + j))| \quad (3.2)$$

Ces fonctions *SSD* et *SAD* comme leurs variantes *ZSSD*, *LSSD*, *NZSSD*, *NSSD*, *NZSAD*, *LSAD* représentent la dissimilarité ou différence entre les deux vignettes ($Z = \text{Zero-mean} / \text{centrée}$, $N = \text{Normalisée}$, $L = \text{Locale}$). Plus la valeur calculée est faible, plus les vignettes se ressemblent. La dénomination Z signifie un centrage des valeurs d'intensité de tous les pixels de la vignette autour de la valeur moyenne. Cette technique permet d'accroître la robustesse des mesures de similarité avec les changements de luminosité.

Une autre fonction de mesure de la corrélation est la corrélation croisée *SCC*. Plus haute est sa valeur, plus les deux vignettes sont similaires. Un facteur de normalisation pour encadrer la valeur de corrélation dans la gamme $[0, 1]$ peut être ajouté dans le cas de la variante Corrélation croisée normalisée *NCC*. Une amélioration de la robustesse face aux changements de luminosité peut être apportée en centrant les valeurs autour de valeur moyenne d'intensité des vignettes. Cette variante s'appelle la corrélation croisée centrée normalisée *ZNCC*. Notons \bar{I}_1 et \bar{I}_2 les valeurs moyennes des niveaux de gris des deux vignettes à corréler. On a alors les formules suivantes :

La corrélation croisée *SCC* :

$$SSC = \sum_{i=-M}^M \sum_{j=-N}^N I_1(x_1 + i, y_1 + j) \cdot I_2(x_2 + i, y_2 + j) \quad (3.3)$$

La corrélation croisée normalisée *NCC* :

$$NCC = \frac{\sum_{i=-M}^M \sum_{j=-N}^N I_1(x_1 + i, y_1 + j) \cdot I_2(x_2 + i, y_2 + j)}{\sqrt{\sum_{i=-M}^M \sum_{j=-N}^N I_1^2(x_1 + i, y_1 + j) \sum_{i=-M}^M \sum_{j=-N}^N I_2^2(x_2 + i, y_2 + j)}} \quad (3.4)$$

La corrélation croisée centrée normalisée *ZNCC* :

$$ZNCC = \frac{\sum_{i=-M}^M \sum_{j=-N}^N (I_1(x_1 + i, y_1 + j) - \bar{I}_1) \cdot (I_2(x_2 + i, y_2 + j) - \bar{I}_2)}{\sqrt{\sum_{i=-M}^M \sum_{j=-N}^N (I_1(x_1 + i, y_1 + j) - \bar{I}_1)^2 \sum_{i=-M}^M \sum_{j=-N}^N (I_2(x_2 + i, y_2 + j) - \bar{I}_2)^2}} \quad (3.5)$$

La corrélation est une technique simple en calcul, facile à implémenter et efficace en performance. Le problème principal de la corrélation est qu'elle ne prend en compte que deux paramètres de translation entre la vignette source et la vignette cible. La rotation, la variation d'échelle et la déformation perspective ne sont pas prises en compte. Les critères suivants sont recommandés lors de l'utilisation de la corrélation : utiliser des images épipolaires (ou rectifiées) et utiliser des fenêtres de petite taille.

Les méthodes de corrélation sont bien adaptées si la vignette contient suffisamment de signal et n'est pas bruitée, et si les distorsions géométriques et radiométriques sont bien contrôlées.

Aujourd'hui, quelques efforts sont faits pour améliorer cette technique. Une combinaison entre la cross-corrélation et *LSM*, afin d'ajouter les paramètres de transformation affine dans la corrélation est introduite par [Zhel'tov and Sibiryakov, 1997].

La corrélation d'images numériques est choisie comme la solution dans plusieurs applications de mesure par les images : [Fayyad and Lees, 2014; Koltsida et al., 2013; Ravanelli et al., 2017; Rosu et al., 2015a,b]. En photogrammétrie, la corrélation est adoptée comme la méthode de base dans plusieurs processus d'appariement dense : [Furukawa and Ponce, 2010; Pierrot-Deseilligny and Paparoditis, 2006; Pons et al., 2007; Remondino et al., 2014, 2013].

L'appariement par moindres carrés L'appariement par moindres carrés est une méthode d'appariement largement étudiée et utilisée dans les applications photogrammétriques ([Afsharnia et al., 2015; Li and Wang, 2014; Rosenholm, 1987]). *LSM* dans sa forme originale est présentée au début par [Förstner, 1982]. Il s'agit d'une méthode développée pour appairer les lignes d'images, donc selon une dimension. [Ackermann, 1984; Gruen, 1985] ont ensuite évolué cette idée en ajoutant les paramètres de transformation géométrique et radiométrique pour l'appariement en deux dimensions, avec une vignette rectangulaire ou carrée.

Notons f et g deux vignettes de taille $p \times q$ à appairer, e est le bruit. On a :

$$f - e = g \quad (3.6)$$

Le pixel de coordonnées (x, y) dans la vignette f est associé avec son pixel correspondant de coordonnées (x', y') dans la vignette g par le modèle de la transformation géométrique et radiométrique :

$$\begin{aligned} f(x, y) - e(x, y) &= r_0 + r_1 g(x', y') \\ x' &= a_0 + a_1 x + a_2 y \\ y' &= b_0 + b_1 x + b_2 y \end{aligned} \quad (3.7)$$

Les six paramètres $(a_0, a_1, a_2, b_0, b_1, b_2)$ sont des composantes dans la transformation affine entre le pixel des coordonnées (x, y) et son correspondant (x', y') entre les deux vignettes :

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} \quad (3.8)$$

et (r_0, r_1) sont les paramètres de luminosité et de contraste dans le modèle radiométrique. Les coordonnées (x, y) et (x', y') peuvent être réelles. L'interpolation bilinéaire est utilisée dans nos implémentations.

L'équation d'observation 3.7 doit être linéarisée du fait de la non-linéarité de la fonction image f et g . En utilisant la série de Taylor, l'équation linéarisée s'exprime :

$$\begin{aligned} f(x, y) - e(x, y) &= g(x', y') + g_x(x', y') da_0 + x' g_x(x', y') da_1 + y' g_x(x', y') da_2 \\ &+ g_y(x', y') db_0 + x' g_y(x', y') db_1 + y' g_y(x', y') db_2 + r_0 + r_1 g(x', y') \end{aligned} \quad (3.9)$$

$$(3.10)$$

Les différentiels partiels $g_x(x', y')$ et $g_y(x', y')$ sont les valeurs de gradient d'image du pixel (x', y') selon les directions x et y :

$$g_x(x', y') = \frac{\partial g}{\partial x} = \frac{g(x' + 1, y') - g(x' - 1, y')}{2} \quad (3.11)$$

$$g_y(x', y') = \frac{\partial g}{\partial y} = \frac{g(x', y' + 1) - g(x', y' - 1)}{2} \quad (3.12)$$

Dans le cas général, les valeurs initiales sont choisies telles que la transformation affine soit égale à l'identité, le facteur de contraste soit égale à 1 et le décalage en luminosité soit égale 0 :

$$a_0^0 = a_2^0 = b_0^0 = b_1^0 = r_0^0 = 0 \quad (3.13)$$

$$a_1^0 = b_2^0 = r_1^0 = 1 \quad (3.14)$$

Chaque pixel (x, y) dans la vignette f forme une équation. Notons \hat{x} le vecteur d'inconnues, A la matrice de design (ou matrice jacobienne), l le vecteur des constantes, v le vecteur de résidus, \hat{x} le vecteur de correction des paramètres à estimer. On a :

$$\begin{aligned} \mathbf{l} + \mathbf{v} &= \mathbf{A}\hat{\mathbf{x}} \\ \text{avec } \mathbf{l}^T &= [\dots \quad f(x, y) - g(x, y) \quad \dots] \\ \hat{\mathbf{x}}^T &= [da_0 \quad da_1 \quad da_2 \quad db_0 \quad db_1 \quad db_2 \quad r_0 \quad r_1] \end{aligned}$$

La solution pour l'équation 3.10 s'exprime :

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{P} \mathbf{A})^{-1} (\mathbf{A}^T \mathbf{P} \mathbf{l}) \quad (3.15)$$

\mathbf{P} est la matrice de poids. Dans nos cas, nous fixons un poids identique pour toutes les équations, d'où $\mathbf{P} = I$. Les valeurs des paramètres à estimer sont corrigées et mises à jour après chaque itération avec le vecteur de correction de paramètres $\hat{\mathbf{x}}$. Les itérations sont terminées si la somme des carrés des résidus est inférieur à un seuil donné.

Il existe plusieurs variantes de *LSM*. La plupart sont basées sur l'ajout de contraintes dans le système d'équations.

[Gruen and Baltsavias, 1986] a évalué en profondeur *LSM* dans le cadre de l'appariement d'images numériques. Il a aussi proposé la méthode *Adaptive Least Square Matching – ALSM*.

En utilisant l'orientation interne et externe des images connues *a priori*, l'auteur ajoute la contrainte géométrique sous la forme d'équations de colinéarité dans le système d'équations d'optimisation par moindres carrés.

Une autre variante qui s'appelle *Geometrically Constrained Multiphoto Matching – MPGC*, [Gruen and Baltsavias, 1988] présente l'implémentation et l'expérimentation de *ALMS*.

[Gruen and Stallmann, 1991] utilise *MPGC* pour apparier les contours d'objets sur des images avec un template synthétique. Cette méthode est utile dans le cas de mesure de la déformation entre l'objet sur une image avec le modèle théorique CAO¹². L'auteur montre aussi son efficacité en l'utilisant comme une méthode d'extraction de contours.

12. Conception assistée par ordinateur

[Baltsavias and Stallmann, 1993] dans la continuité de *ALMS*, propose d'ajouter la contrainte géométrique dérivée de la ligne épipolaire. Cette contrainte est utile dans les cas où l'orientation externe des images est connue, par exemple l'appariement d'images satellite.

[Bethmann and Luhmann, 2011] propose de remplacer la transformation affine par la transformation projective ou un modèle polynomial.

Un résumé de toutes les variantes de MPGC se trouve dans [Gruen, 1996]

3.4 Conclusion

Ce chapitre présente des aspects de l'état de l'art en relation à notre problématique. Il s'agit des sources d'erreurs en photogrammétrie, des thématiques de points homologues comme le détecteur et descripteur de points d'intérêt, l'appariement basé caractéristiques visuelles ou basé région. Les sources d'erreurs en photogrammétrie et les méthodes de correction de l'erreur systématiques fortement liés avec le contexte de la thèse sont étudiées. Les méthodes d'extraction de points homologues les plus populaires ainsi que leurs avantages et désavantages sont également présentés.

De ces différentes études, on constate qu'il est nécessaire d'avoir une méthode d'amélioration de la précision de mise en place des images et de la calibration des caméras en utilisant uniquement des contraintes internes aux images. Les études dans l'état de l'art nous permet d'identifier les critères à optimiser en terme d'extraction des points homologues dans le contexte de la photogrammétrie.

Chapitre 4

Extraction des points homologues SIROP

Sommaire

4.1	Introduction	47
4.2	Description détaillée de l’algorithme	48
4.2.1	Sélection des images	48
4.2.2	Rectification des régions d’intérêt	50
4.2.3	Détection des points d’intérêt	52
	Sélection des extrema locaux	52
	Filtrage sur le contraste	53
	Filtrage des motifs répétitifs	54
	Filtrage spatial des points d’intérêt	55
4.2.4	Appariement par la corrélation	55
4.2.5	Filtrage spatial des points homologues	56
4.2.6	Raffinement par moindres carrés	58
4.3	Conclusion	59

4.1 Introduction

Ce chapitre présente la méthode proposée dans les travaux de cette thèse. Son but est l’extraction de points homologues de haute qualité pour pouvoir affiner la mise en place d’images.

Ces nouveaux points sont extraits à partir de l’information géométrique 3D approchée fournie par le calcul photogrammétrique de l’orientation relative réalisé avec les chaînes de traitement existantes. Ces informations sont utilisées pour déterminer un nouvel ensemble de points homologues de qualité supérieure. La méthode peut fonctionner avec tous les types de scénarios, sans utilisation de contraintes externes. Plus précisément, les principales contributions de cette méthode sont :

- des points homologues sont extraits dans toutes les régions de l’image couvertes par la scène 3D. La méthode peut extraire des points dans des régions de contraste faible, où les méthodes existantes sont peu performantes ;
- les images de chaque triangle de la scène sont rectifiées de la déformation de la projection. Ainsi l’algorithme est adapté à des acquisitions avec une longue

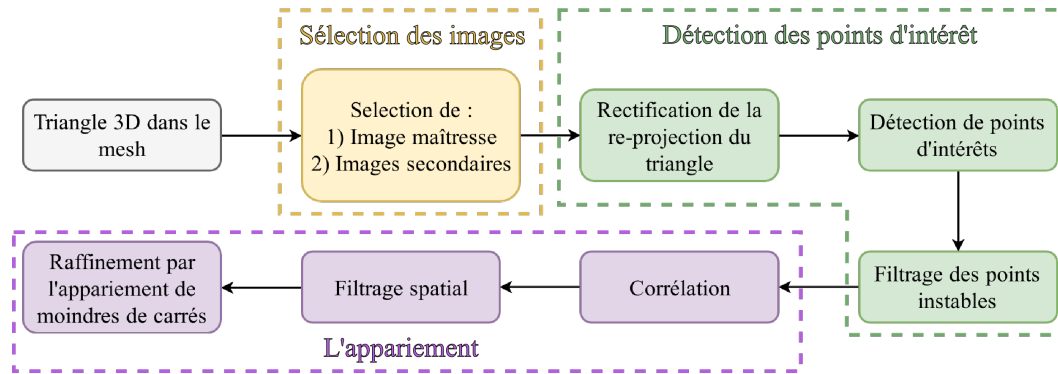


FIGURE 4.1 – Les étapes successives de l'algorithme d'extraction de points homologues *SIROP*.

ligne de base, donc de fortes déformations entre les images ; la méthode n'est pas limitée aux scènes planes comme celle proposée par [Yang et al., 2010] ;

- les points homologues ont une meilleure multiplicité (sauf pour les points à l'infini comme cela sera discuté en section 5.4.2).

La figure 1.9 présente un schéma de la méthode proposée. L'algorithme est implémenté dans la chaîne de photogrammétrie MicMac [Deseilligny and Clery, 2011; Rupnik et al., 2017] sous le nom *SIROP*. La figure 4.1 présente les étapes successives de cet algorithme.

La méthode *SIROP* peut être décrite en trois étapes effectuées pour chaque triangle :

- sélection des images où le triangle est visible : l'image maîtresse est celle qui le voit le mieux, les autres sont les images secondaires ;
- détection et filtrage des points d'intérêts du triangle ;
- appariement par corrélation et affinement par moindres carrés de ces points.

Le traitement sur tous les triangles assure une distribution homogène des points dans l'espace objet.

Les points *SIROP* sont extraits à partir d'une modélisation géométrique de la scène et de son acquisition par le système imageur. La modélisation de la scène est donnée par un maillage 3D (mesh). La modélisation de l'acquisition est donnée par les poses (orientation externe) et la calibration de la caméra (orientation interne). La combinaison de ces modélisations permet de prédire la zone d'appariement et la déformation perspective entre les images afin d'extraire les points *SIROP*. Pour que l'algorithme fonctionne il faut pouvoir majorer l'erreur de prédiction ; ceci est réglé par un paramètre, en pratique la valeur par défaut (10 pixels) est déjà une forte majoration.

On améliore ensuite la précision de l'aérotriangulation en réalisant une seconde itération de traitement photogrammétrique en utilisant les nouveaux points homologues (points *SIROP*).

Chaque étape de ce algorithme est présentée en détails dans la partie suivante.

4.2 Description détaillée de l'algorithme

4.2.1 Sélection des images

Le but de cette étape est de sélectionner une image maîtresse et un ensemble d'images secondaires pour chaque triangle du mesh. Les images candidates retenues doivent satisfaire les deux contraintes :

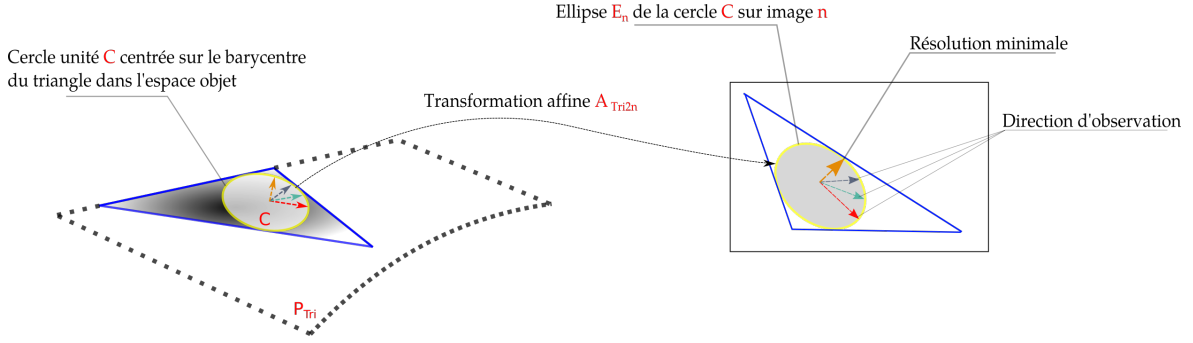


FIGURE 4.2 – Directions d'observations et résolution minimale.

- la visibilité, assurée par un calcul d'occlusion sur le maillage 3D par Z-Buffer ;
- une résolution spatiale minimale : les triangles dont la surface re-projetée sur l'image est inférieure à 100 pixels sont éliminés.

Dans la suite on appellera “triangle 3D” un triangle dans l'espace objet, et “triangle 2D” la projection d'un triangle 3D dans un plan image.

Parmi les images candidates sélectionnées, l'image maîtresse est choisie avec la condition de résolution optimale. En raison de la déformation due à la perspective, la résolution d'un élément de surface (triangle) re-projeté change selon l'angle de la direction d'observation par rapport au plan 3D du triangle.

En toute rigueur, compte tenu des déformations perspectives, la résolution d'une image en un point varie en fonction de la direction. On adopte une stratégie prudente (pessimiste) consistant à définir la résolution en un point comme étant la plus faible obtenue dans toutes les directions possibles de l'image. L'image maîtresse choisie est celle ayant la plus grande résolution minimale.

Pour cela il faut d'abord déterminer mathématiquement ce qu'est la résolution minimale. La résolution minimale est calculée via la déformation d'un cercle dans le triangle 3D par rapport à sa projection sur le plan image. La fonction de projection est approximée par une transformation affine entre 3 sommets du triangle de son plan dans l'espace objet et leurs projections sur le plan image. Dans le cas général, la transformation entre le plan objet et le plan image est une homographie. Cependant, compte tenu de la taille des triangles, nous avons pu vérifier que la différence, à l'intérieur d'un triangle, est négligeable entre homographie et une affinité. Par simplicité on choisit donc une transformation affine. Les 6 paramètres de cette transformation sont calculés à partir des 3 sommets du triangle 3D et de leurs projections.

On note P_n le plan image n , P_{Tri} le plan du triangle dans l'espace objet. C est le cercle unitaire centré sur le barycentre du triangle dans le plan P_{Tri} (voir la figure 4.2 et 4.3).

On note A_{Tri2n} la matrice 2×2 représentant la partie linéaire de la transformation affine de passage entre un triangle 3D et sa projection sur l'image n .

E_n est l'ellipse transformée du cercle C dans le plan image P_n par l'affinité A_{Tri2n} :

$$A_{Tri2n} = \begin{bmatrix} \vec{u}_n & \vec{v}_n \end{bmatrix} \quad (4.1)$$

Pour chaque point $(\cos(\theta), \sin(\theta))$ sur le cercle C , le point correspondant sur l'ellipse E_n s'exprime :

$$\begin{aligned}
 A_{Tri2n} \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} &= \begin{bmatrix} \vec{u}_n & \vec{v}_n \end{bmatrix} \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} \\
 &= \vec{u}_n \cos \theta + \vec{v}_n \sin \theta = \vec{V}_n(\theta)
 \end{aligned} \tag{4.2}$$

La valeur minimale $\min_{\theta}(\vec{V}_n(\theta))$ (l'équation 4.3) représente la résolution minimale obtenue selon toutes les directions d'observation du cercle C dans l'image n . Afin de chercher cette valeur, on développe à partir de l'équation 4.2 :

$$\begin{aligned}
 \vec{V}_n(\theta) &= \vec{u}_n \cos \theta + \vec{v}_n \sin \theta \\
 \|\vec{V}_n(\theta)\|^2 &= (\vec{u}_n \cos \theta + \vec{v}_n \sin \theta)^2 \\
 &= \vec{u}_n^2 \cos^2(\theta) + \vec{v}_n^2 \sin^2(\theta) + 2\vec{u}_n \vec{v}_n \sin(\theta) \cos(\theta) \\
 &= \frac{\vec{u}_n^2}{2}(1 + \cos(2\theta)) + \frac{\vec{v}_n^2}{2}(1 - \cos(2\theta)) + \vec{u}_n \vec{v}_n \sin(2\theta) \\
 &= \frac{\vec{u}_n^2 + \vec{v}_n^2}{2} + \left(\frac{\vec{u}_n^2 - \vec{v}_n^2}{2}\right) \cos(2\theta) + \vec{u}_n \vec{v}_n \sin(2\theta) \\
 &= \frac{\vec{u}_n^2 + \vec{v}_n^2}{2} + (a \cos(2\theta) + b \sin(2\theta))
 \end{aligned} \tag{4.3}$$

On pose $V_0 = \frac{\vec{u}_n^2 + \vec{v}_n^2}{2}$, $L = \sqrt{a^2 + b^2}$ et ϕ t.q. $\cos \phi = \frac{a}{L}$, $\sin \phi = \frac{b}{L}$

On a : $\|\vec{V}_n(\theta)\|^2 = V_0 + L(\cos \phi \cos(2\theta) + \sin \phi \sin(2\theta)) = V_0 + L \cos(\phi - 2\theta)$

La valeur minimale de $\|\vec{V}_n(\theta)\|^2$ est donc égale à :

$$V_0 - L = \frac{\vec{u}_n^2 + \vec{v}_n^2 - \sqrt{(\vec{u}_n^2 - \vec{v}_n^2)^2 + 4(\vec{u}_n \cdot \vec{v}_n)^2}}{2}$$

L'image maîtresse est choisie comme ayant la valeur maximale parmi toutes les résolutions minimales (l'équation 4.3) sur toutes les images.

$$n_{Master} = \max_n(\min_{\theta}(\|\vec{V}_n(\theta)\|^2)) \tag{4.4}$$

Si le nombre des images secondaires est trop important, ce critère de résolution peut être utilisé pour sélectionner le sous-ensemble des N meilleures images secondaires, où N est un paramètre de l'algorithme. Dans les exemples traités dans ce document, on a utilisé toutes les images secondaires pour maximiser la multiplicité des points homologues.

Le résultat du calcul de l'image maîtresse pour chaque triangle est visualisé sur la figure 4.3 sous la forme d'une partition des triangles du maillage 3D.

4.2.2 Rectification des régions d'intérêt

Pour pouvoir corrélérer correctement entre les différentes images d'un même triangle, celles-ci sont d'abord reprojctées dans la géométrie de l'image maîtresse. On appelle région d'intérêt (ROI) la projection d'un triangle objet dans une image (voir figure 4.4).

Une transformation affine à 6 paramètres entre l'image maîtresse et l'image secondaire concernée est utilisée. Celle-ci est déterminée à partir des positions des 3 sommets des triangles dans les deux images.

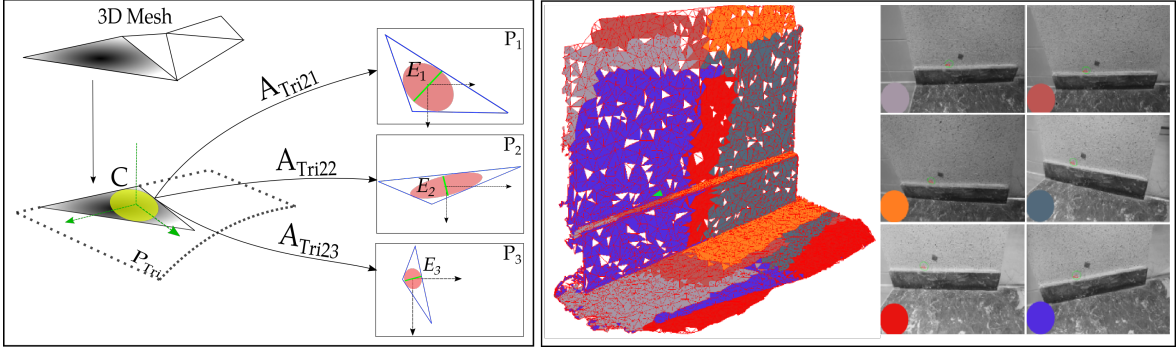


FIGURE 4.3 – La sélection de l'image maîtresse selon le critère de résolution optimale.

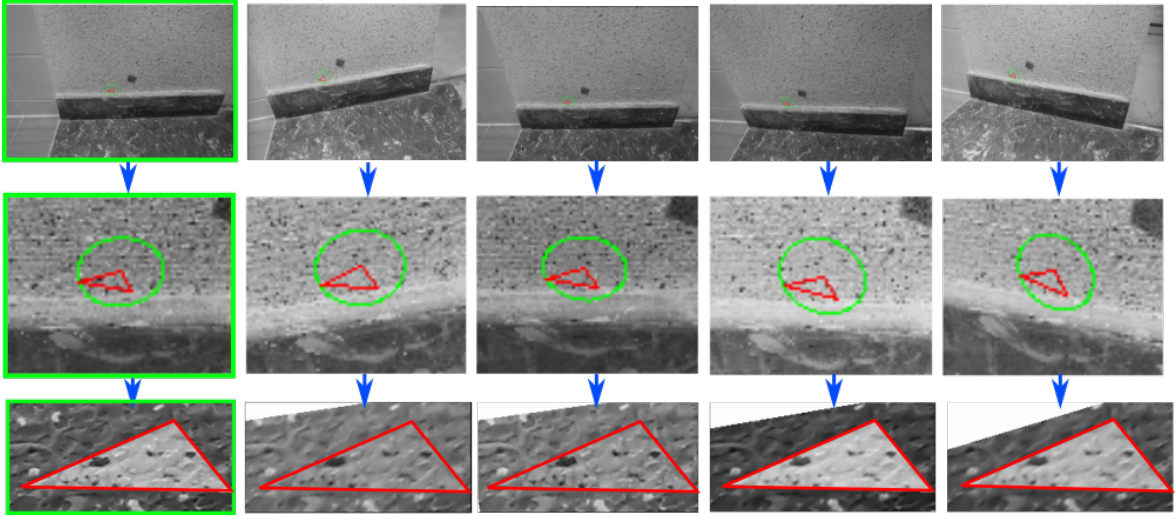


FIGURE 4.4 – La rectification de la ROI.

En appliquant cette transformation, les distorsions perspectives entre les images de différents points de vue sont corrigées.

Notons (x_s, y_s) les coordonnées d'un pixel dans la ROI d'une image secondaire et (x_m, y_m) celles du pixel correspondant dans l'image maîtresse. L'affinité est définie comme suivant :

$$\begin{bmatrix} x_m \\ y_m \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \begin{bmatrix} x_s \\ y_s \end{bmatrix} + \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} \quad (4.5)$$

L'équation 4.5 est appliquée à tous les pixels dans les ROIs des images secondaires pour les redresser dans la géométrie de l'image maîtresse avec une interpolation bilinéaire. La figure 4.4 montre un exemple de rectification de ROIs. Les images de la première colonne concernent l'image maîtresse. La première ligne montre les images entières, la deuxième ligne est un extrait de la région autour du triangle re-projeté, avec l'ellipse affinité. La troisième ligne montre les ROIs redressées.

L'erreur de prédiction des coordonnées des points causée par l'approximation entre homographie et affinité en fonction de la surface du triangle concernée est montrée dans la figure 4.5. Pour un triangle ayant une surface de 9000 pixels, cas le plus défavorable, l'erreur de prédiction n'excède pas 0.08 pixel. Avec une taille de triangle plus standard sur cette scène (environ 2000 pixels correspondant au pic de fréquence), cette erreur n'est que de 0.02 pixel. La figure 4.5 montre aussi que dans 99% des cas, l'erreur de prédiction ne dépasse pas 0.13 pixel.

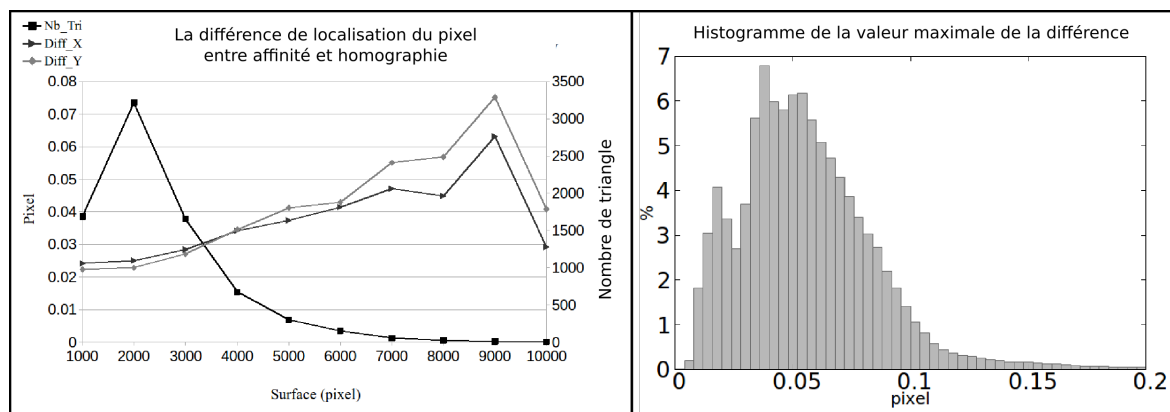


FIGURE 4.5 – L’erreur de prédiction entre homographie et affinité en fonction de la surface du triangle.

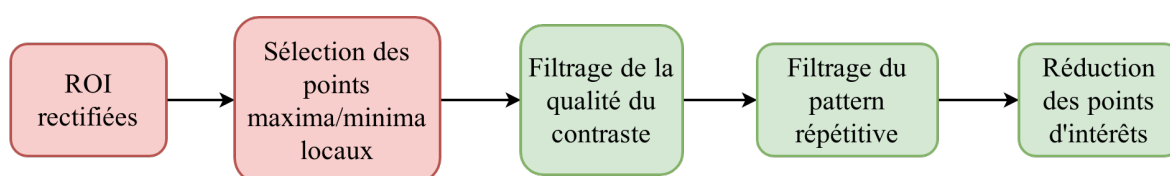


FIGURE 4.6 – Les étapes de la détection des points d’intérêt. **En rouge** : effectuées sur toutes les images sélectionnées, **En vert** : effectuées uniquement sur l’image maîtresse.

4.2.3 Détection des points d’intérêt

La démarche suivie pour la détection des points d’intérêt est d’enchaîner (voir la figure 4.6) plusieurs filtres, en commençant par les plus rapides, puis en rajoutant des critères plus sophistiqués lorsque le nombre de candidats a diminué :

- sélection des extrema sur toutes les images. Cette sélection est faite sans aucun seuil radiométrique ;
- filtrage pour sélectionner des candidats ayant le meilleur potentiel :
 - points ayant suffisamment de contraste ;
 - points non ambigus ;

Ce filtrage n’est fait que sur l’image maîtresse afin de ne pas risquer de perdre des bons candidats pour l’appariement ; ces filtres sont relativement peu sévères dans le but de conserver suffisamment de points. Lors des étapes finales (corrélation et filtrage spatial) le nombre de points est significativement réduit.

Sélection des extrema locaux

Les points correspondant à des maxima/minima locaux en niveau de gris sont détectés dans les ROI. Dans notre implémentation, il s’agit des pixels correspondant à des maxima/minima d’intensité dans une région circulaire de 3 pixels. Cette valeur a été fixée empiriquement de manière à réduire significativement le nombre de candidats potentiels (typiquement il reste 3-4% des pixels de l’image).

Cette étape génère un ensemble de points d’intérêt (points candidats) pour tous les triangles rectifiés sur toutes les images. Ensuite ces candidats sont filtrés, uniquement sur l’image maîtresse, à l’aide de quelques critères tels que la qualité du contraste, le risque de motif répétitif et la distribution spatiale. Le processus de détection des points

d'intérêt est présenté brièvement sur la figure 4.6. Chaque filtrage est présenté en détail dans les parties suivantes.

Une fois les points filtrés, ils sont appariés de manière subpixelaire pour former les points homologues (voir la section 4.2.4).

Filtrage sur le contraste

Le but de ce filtrage est d'éliminer les points candidats de faible contraste, ceux qui risquent d'être mal appariés par la corrélation. L'idée de ce filtre est basée sur les algorithmes FAST [Rosten, 2006] et SUSAN [Smith and Brady, 1997].

Au lieu d'utiliser directement les détecteurs de différences de gaussiennes du type *SIFT*, on a choisi cette approche pour les raisons suivantes :

- rapidité par rapport aux détecteurs de type *SIFT* ;
- appariement sur les ROIs rectifiées : comme présenté dans la section 4.2.2, les déformations perspectives dans les ROIs sont corrigées. Donc, une approche invariante à l'échelle et à la rotation n'est pas nécessaire ;
- il n'est pas nécessaire d'avoir une localisation précise des points dans les images individuelle (de type sub-pixelaire comme dans *SIFT* et la majorité des détecteurs), puisque la précision viendra de l'étape d'appariement par corrélation.

Notre implémentation diffère légèrement de [Rosten, 2006] afin de supprimer moins de points : on utilise deux critères avec des seuils de filtrage très bas. Le filtre étudie la valeur d'intensité d'un point candidat par rapport à ses 24 voisins se situant sur le cercle de rayon $R = 4$. Notons p la valeur d'intensité d'un point candidat, et $(p_1..p_i..p_n)$ les valeurs d'intensité de ses voisins (1.. i .. n) respectivement. La différence en valeur d'intensité d_i est calculée entre chaque voisin i et le point candidat comme $|p - p_i|$ (voir la figure 4.7). Pour pouvoir évaluer la qualité de contraste de chaque point candidat, un score de qualité de contraste (contrast quality score, *CQS*) est calculé. *CQS* est composé de deux termes :

- contraste avec une priorité pour la continuité CQS_1 ;
- contraste global CQS_2 .

Appelons V_d le vecteur des différences de valeurs d'intensité : $V_d = [d_1, \dots, d_{24}]$.

Pour le score CQS_1 , une fenêtre glissante balaye le vecteur V_d . Chaque position de la fenêtre nous donne un score correspondant à la valeur maximale des éléments dans la fenêtre. Le score CQS_1 est la valeur minimale sur tous ces scores.

Le score CQS_2 est le 75^e percentile dans le vecteur V_d , pour calculer cette valeur on utilise un algorithme récursif rapide permettant d'extraire la 18^{ème} sur 24.

Le vecteur est donc trié dans l'ordre croissant et $CQS_2 = V_{d_{sorted}}[18]$ (voir la figure 4.7).

Un seuil est appliqué sur CQS_1 (3 niveaux de gris sur 8 bits) et CQS_2 (5 niveaux de gris sur 8 bits) pour éliminer les points qui ne sont pas assez contrastés. Finalement, une combinaison pondérée des deux termes CQS_1 et CQS_2 nous donne le score de qualité de contraste CQS :

$$CQS = CQS_1 + 2 \times CQS_2 \quad (4.6)$$

Ce score CQS est utilisé dans l'étape de réduction du nombre de points d'intérêt (voir la section 4.2.3) pour déterminer la priorité des points dans le filtre.

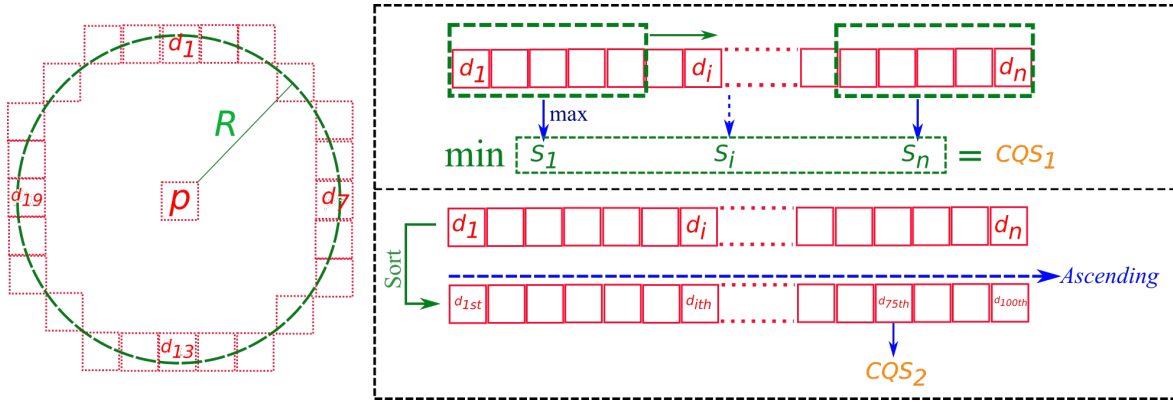


FIGURE 4.7 – La méthode de calcul des scores de contraste CQS_1 et CQS_2 .

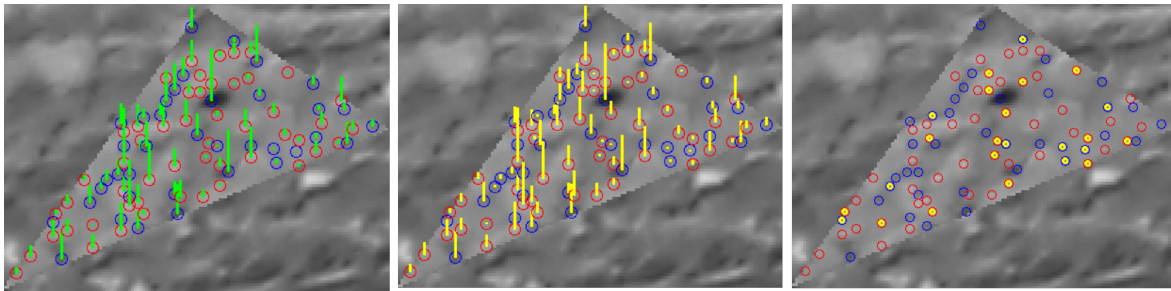


FIGURE 4.8 – Un exemple de filtrage par le contraste.

La figure 4.8 montre le fonctionnement du filtre de contraste : pour les points candidats (bleu pour le minimaux et rouge pour le maximaux locaux), le score CQS_1 est montré à gauche, CQS_2 au milieu et les points rejetés sont en jaune à droite.

Filtrage des motifs répétitifs

Afin d'éliminer les points qui peuvent causer une ambiguïté dans l'appariement par corrélation, on utilise un filtre permettant de rejeter les points qui ont un voisinage qui leur ressemble trop ; en pratique ce filtre supprime surtout les détails linéaires.

Pour chaque point candidat, ses 24 points voisins sur un cercle de rayon 4 sont testés. Un coefficient de corrélation est calculé entre une vignette 7×7 pixels centrée sur le point candidat et chacun des points voisins (voir la figure 4.9).

Le point candidat est éliminé si une corrélation forte est détectée dans au moins une direction. Dans cette implémentation, le seuil d'élimination choisi est de 0.85.

La figure 4.9 présente un exemple de ce filtre. À gauche, le voisinage (une vignette autour d'un point candidat) est représenté en jaune, au centre, un motif répétitif est détecté. À droite, des points éliminés par des filtres sont présentés en jaune.

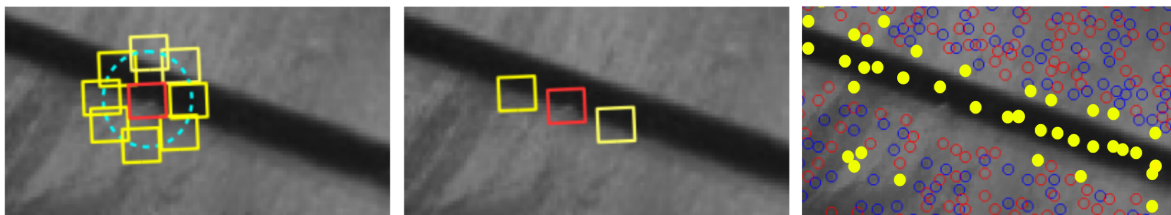


FIGURE 4.9 – Filtrage des motifs répétitifs.

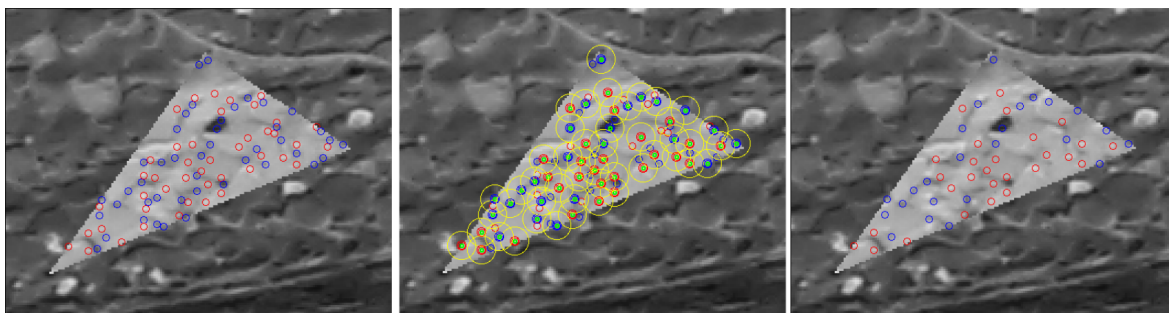


FIGURE 4.10 – Le filtre de réduction du nombre de points d'intérêt.

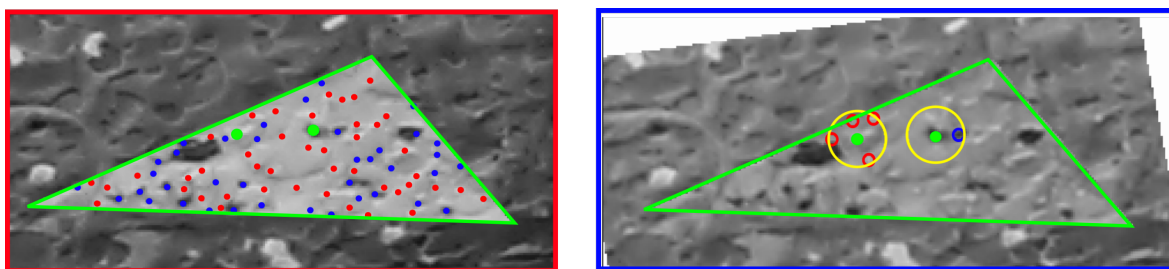


FIGURE 4.11 – La sélection des points appariés potentiels.

Filtrage spatial des points d'intérêt

L'objectif final est d'obtenir une densité fixée de points homologues. Pour réduire suffisamment le nombre de points d'intérêt sans risquer de tomber en dessous de la densité cible, on ne peut filtrer chaque point indépendamment des autres ; on va donc utiliser un filtre spatial qui va conserver localement les meilleurs points (sur le critère de contraste).

Après la phase de détection des points d'intérêt, un filtre est appliqué pour réduire la quantité de ces points. Les points d'intérêt sont triés suivant leur score de contraste CQS afin de former une liste de priorité pour le processus de filtrage. En prenant les points dans l'ordre du score de contraste décroissant, pour chaque point d'intérêt dans le triangle de l'image maîtresse, le filtre forme une région circulaire centrée sur le point (voir la figure 4.10). Seul le point le plus haut en score de contraste dans la région circulaire est retenu.

4.2.4 Appariement par la corrélation

Les points d'intérêt dans l'image maîtresse sont mis en correspondance avec ceux des images secondaires pour créer des points homologues. Comme les ROIs des images secondaires sont rééchantillonnées dans la géométrie de l'image maîtresse, les méthodes d'appariement basées région peuvent être employées pour déterminer les homologues.

La méthode de corrélation utilisée est ZNCC (Zero Mean Normalized Cross-Correlation, voir la section 3.3.2), en plusieurs étapes. Suite au rééchantillonnage, les points d'intérêt sont *a priori* à la même position dans toutes les images corrigées. La zone de sélection des homologues potentiels est une région circulaire de rayon 12 pixels (voir la figure 4.11).

La corrélation entre un point d'intérêt dans l'image maîtresse et chacun des points homologues potentiels dans les images secondaires se déroule en trois étapes avec une approche par raffinement successifs (*coarse to fine*) :

- la corrélation entière en sous-échantillonnant d'un facteur 2 (taille de vignette de corrélation 7×7 pixels) ;
- la corrélation en pixels entiers : tous les pixels dans une vignette dont la taille est fixée à 13×13 pixels sont pris en compte ;
- la corrélation sub pixelaire avec un pas de 0.1 pixel. La taille de la vignette est fixée à 13×13 pixels.

Un seuil de 0.7 est appliqué au score de corrélation lors de la première étape pour éliminer les points trop différents.

Finalement, le point ayant le meilleur score de corrélation est sélectionné comme point homologue. Son score de corrélation est utilisé dans l'étape suivante.

4.2.5 Filtrage spatial des points homologues

Le filtrage spatial a pour but de réduire le nombre de points homologues provenant de la corrélation (4.2.4) tout en maintenant une distribution homogène de ces points dans l'espace image. Cet algorithme est relativement complexe car, pour satisfaire les contraintes de compacité et de qualité photogrammétrique, il doit optimiser simultanément plusieurs critères :

- il faut sélectionner les points les plus fiables (ici les points les mieux corrélés au sens du ZNCC) ;
- il faut privilégier la multiplicité des points (points vus sur le plus grand nombre d'images) ;
- dans l'image maîtresse, il faut garantir une densité spatiale minimale des points sélectionnés correspondant à chaque image secondaire ; formellement cela signifie que l'on essaye de garantir que partout dans l'image il existe à une distance donnée au moins un point sélectionné par image secondaire (voir exemple ci-dessous) ;
- ceci devant être assuré avec un nombre minimal de points.

Pour implémenter ces critères on définit un score $SGC(P_k)$ pour chaque point homologue P_k dans l'image maîtresse qui représente le gain qu'apporte ce point s'il est sélectionné. Ce gain dépend évidemment des points qui ont déjà été choisis, par exemple :

- si un point P_1 est apparié sur les images secondaires $\{I_1, I_2\}$, on notera $S_{init}(P_1) = \{I_1, I_2\}$ les images qui sont vues par P_1 ;
- si on a déjà sélectionné, à une distance inférieure au seuil, un point P_2 voyant les images $\{I_2, I_3\}$, alors I_2 est couverte par P_2 pour P_1 , le gain de P_1 va diminuer car il n'est plus intéressant que pour l'image I_1 , on notera $\bar{S}(P_1) = \{I_1\} = S_{init}(P_1) - S_{init}(P_2)$ les images qui sont non couvertes pour P_1 ;
- si, de plus, un point P_3 , voyant les images $\{I_1, I_4\}$ est aussi déjà sélectionné dans le voisinage, alors P_1 sera considéré sans intérêt et supprimé de la liste des points potentiels ($\bar{S}(P_1) = \emptyset$).

On définit maintenant le gain par :

$$SGC(P_k) = \sum_{I \in \bar{S}(P_k)} \frac{1}{\epsilon + (1 - C(P_k, I))} \quad (4.7)$$

où $C(P_k, I)$ est le score de corrélation du point homologue P_k entre l'image maîtresse et une image secondaire S_t obtenu à l'étape précédente. ϵ est défini à 0.02 pour éviter

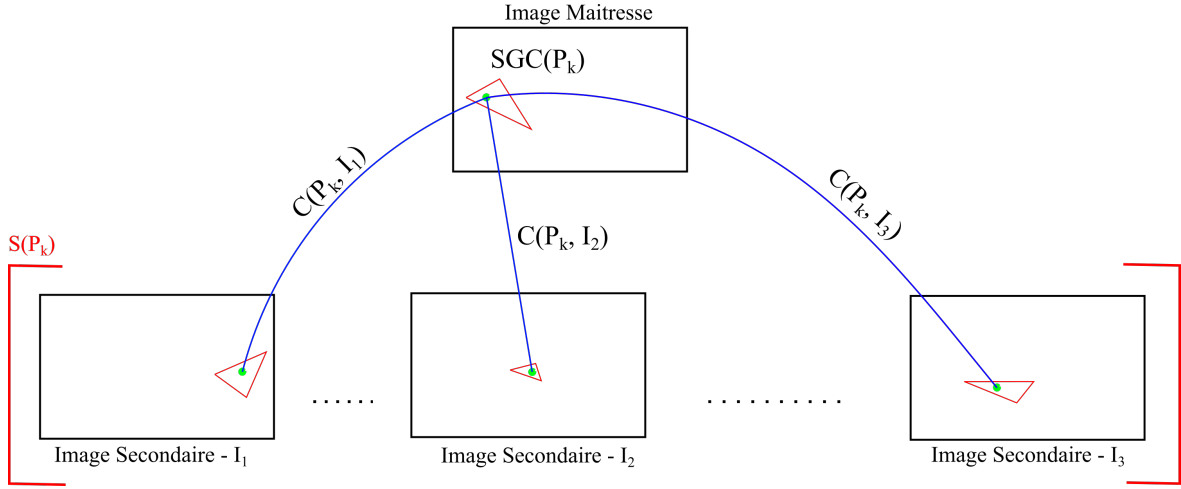


FIGURE 4.12 – Le calcul du score global de corrélation $SGC(P_k)$ pour chaque point homologue multiple P_k dans l'image maîtresse.

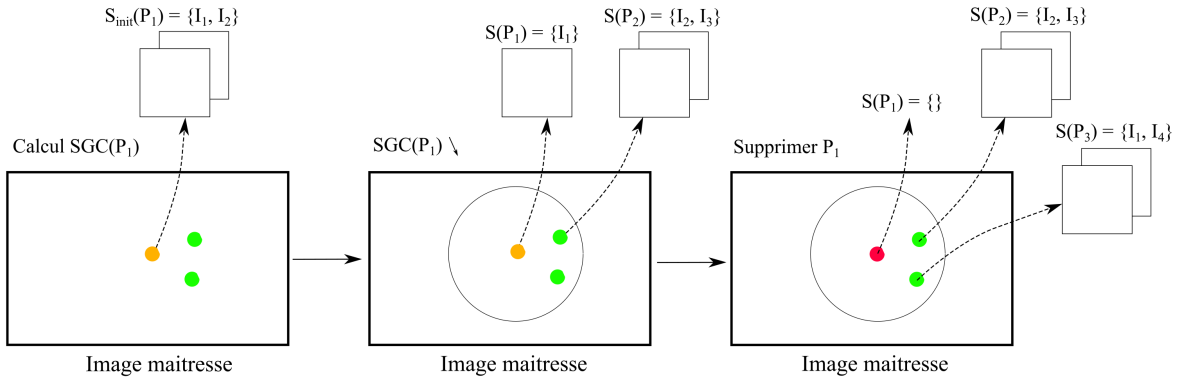


FIGURE 4.13 – Illustration du filtrage spatial. Des points verts sont déjà sélectionnés. Le point jaune est courant. Le point rouge est supprimé.

d'avoir un dénominateur nul. L'équation 4.7 peut être interprétée comme un indicateur de qualité du point, en prenant en compte le score de corrélation et la multiplicité. Les points ayant les meilleurs scores sont retenus (les mieux corrélés et les plus multiples), et sont mis en position prioritaire dans la file d'attente du processus de filtrage.

L'algorithme de sélection des points homologues fonctionne sur le principe général suivant :

- initialement $\bar{S}(P_k)$ toutes les images vues par P_k , initialiser les SGC ;
- soit L la liste des candidats, initialement L est égale à tous les points ;
- tant que $L \neq \emptyset$:
 - extraire de L le point P_{max} de SGC maximal ;
 - pour tous les points P_k dans un disque de rayon R autour de P_{max} , mettre à jour $SGC(P_k)$ et $\bar{S}(P_k) \leftarrow \bar{S}(P_k) - S_{init}(P_{max})$, supprimer P_k de L si $\bar{S}(P_k) = \emptyset$

Pour implémenter rapidement cet algorithme, on utilise une file de priorité (*heap*) pour extraire le meilleur candidat et un indexe spatial (*quadtree*) pour extraire le voisinage d'un point donné.

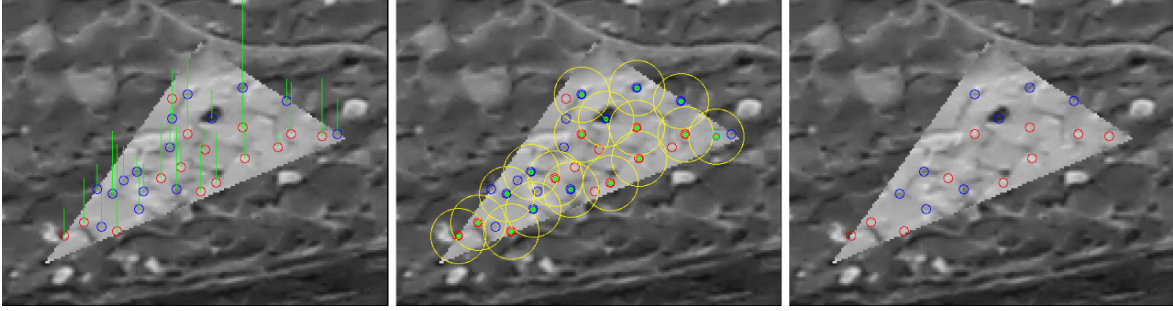


FIGURE 4.14 – Étapes du filtre spatial. (**Gauche**) : tri des points par le score global de corrélation $SGC(P_k)$ représenté par les colonnes vertes, (**Milieu**) : la région de filtrage en jaune, (**Droite**) : les points retenus après filtrage

4.2.6 Raffinement par moindres carrés

Après avoir extrait tous les points homologues par *SIROP*, la méthode d'appariement par moindres carrés (voir la section 3.3.2) est utilisée pour compenser l'erreur d'orientation image non modélisée.

Les incertitudes sur les orientations des caméras et sur le mesh en entrée de *SIROP* font que le rééchantillonnage ne met pas forcément toutes les images dans la même géométrie. Le résultat obtenu par ZNCC peut alors être biaisé, et on utilise le LSM pour s'affranchir de ce problème.

Pour réaliser le raffinement par moindres carrés, les valeurs d'intensité des pixels dans l'imagette centrée sur les points homologues sont utilisées comme des observations. Les 6 paramètres de la transformation affine entre deux triangles projetés (voir les équations de *LSM* dans la section 3.3.2) et le modèle radiométrique (2 paramètres pour le contraste et la luminosité) sont les inconnues dans le modèle de moindres carrés :

$$I_1(x, y) = A \times I_2(a_0 + a_1x + a_2y, b_0 + b_1x + b_2y) + B \quad (4.8)$$

où I_1 , I_2 sont les deux images et x, y parcourent une vignette (ici de taille 7×7) centrée sur les points homologues mis en correspondance par la corrélation. Un modèle linéaire avec le terme de contraste A et luminosité B est adopté pour modéliser la variation de la radiométrie entre deux vignettes. On prend pour valeurs initiales $A = 1$ et $B = 0$.

Le système de l'équation 4.8 est linéarisé en estimant le gradient (voir la section 3.3.2) autour de la transformation affine obtenue à partir des 3 sommets du triangle pour 4 paramètres a_1, a_2, b_1, b_2 . Deux paramètres de translation a_0, b_0 sont initialisés à partir des coordonnées obtenues par la corrélation issue de l'étape précédente (voir la section 4.2.4).

Les paramètres sont estimés après quelques itérations. La condition d'arrêt est soit le nombre maximal d'itérations, soit la convergence de l'erreur entre deux itérations consécutives.

Afin d'évaluer l'avantage de l'utilisation du raffinement par *LSM*, une expérimentation a été effectuée sur 5 images. La figure 4.15 montre une amélioration de 0.02 pixels sur l'erreur de re-projection des points sur ce jeu de données et des changements mineurs sur les paramètres de calibration. Avec une acquisition typique, le temps de traitements pour raffiner des milliers de points homologues est non-négligeable. Par conséquent, dans l'implémentation de *SIROP* dans MicMac, *LSM* est une option non-activée par défaut. Cette étape est également désactivée dans toutes les expériences présentées

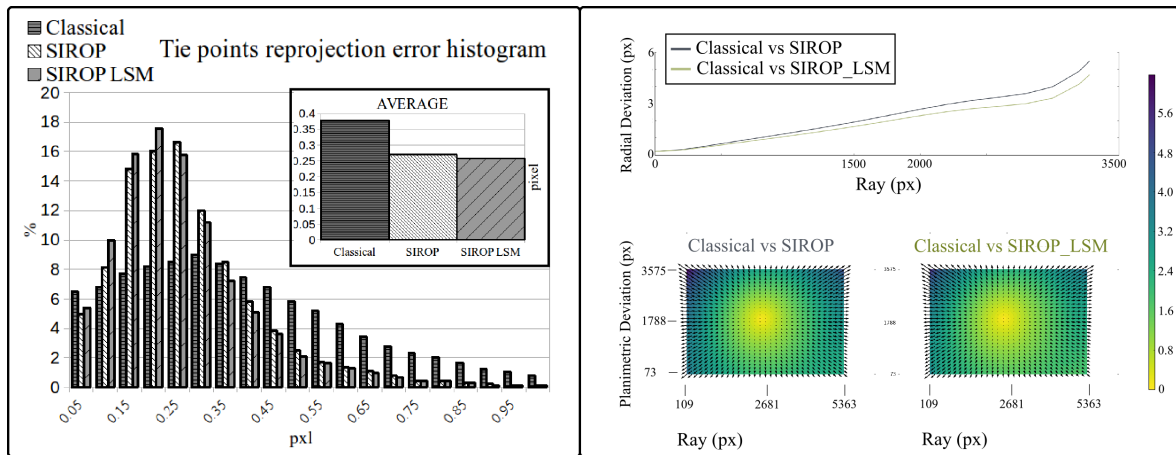


FIGURE 4.15 – Evaluation de l’apport du raffinement par *LSM* sur des points *SIROP* sur l’histogramme d’erreurs de re-projection et sur l’évaluation des paramètres internes.

dans le Chapitre 5.

4.3 Conclusion

Ce chapitre détaille la procédure, l’algorithme et l’implémentation de chaque étape d’extraction des points homologues améliorés *SIROP*. La méthode est composée de plusieurs filtres et étapes successives avec des seuils associés.

Afin de généraliser la méthode à différents types d’images, les seuils ont été choisis pour être tolérants. Le nombre élevé de filtres permet de sélectionner précisément les meilleurs points tout en étant peu sensibles aux caractéristiques des images. Les valeurs des seuils par défaut dans l’implémentation actuelle sont données dans le tableau 4.1. Dans toutes les expériences présentées dans le chapitre 5, les seuils sont fixés aux valeurs par défaut et sont donc identiques pour tous les jeux de données.

Seuil	Nom de la variable	Valeur par défaut
La surface minimale pour éliminer les triangles petits	TT_SEUIL_SURF_TRL_PIXEL	100
Rayon de zone de sélection des extrema locaux	TT_DIST_EXTREMA	3
Rayon de zone de sélection des homologues potentiels	TT_DIST_RECH_HOM	12 pixels
Rayon de zone de filtrage spatial des points d'intérêt	TT_RatioFastFiltrSpatial	6.25 pixels (ratio 1/8 par rapport au TT_DefSeuilDensiteResul)
Seuil de corrélation pour éliminer les points dans le filtrage des motifs répétitifs	TT_SEUIL_AutoCorrel	0.85
Taille de la fenetre de corrélation dans le filtrage des motifs répétitifs	TT_SZ_AUTO_COR	3 pixels (fenetre 7×7)
Rayon de voisinage dans le filtrage sur le contraste	TT_DIST_FAST	4 pixels (24 pixels voisinages)
Seuil d'écart de valeur d'intensité pour contraste global CQS_2	TT_SeuilFastStd	5
Le percentile dans le vecteur V_d pour obtenir le score CQS_2	TT_PropFastStd	0.75 (75 ^e percentile)
Seuil d'écart de valeur d'intensité pour contraste avec une priorité pour la continuité CQS_1	TT_SeuilFastCons	3
Réglage la taille de la fenêtre glissante balaye le vecteur V_d dans la computation du score CQS_1	TT_PropFastConsec	0.6 ($60\% \times 24 = 15pixels$, 24 est le nombre de pixels voisinages)
Seuil de corrélation entière en sous échantillonné d'un facteur 2 pour éliminer les points homologues trop différents.	TT_SEUIL_CORREL_1PIXSUR2	0.7
Taille de la fenêtre de corrélation	TT_DemiFenetreCorrel	6 pixels (fenêtre de 13×13)
Rayon de zone de sélection le voisinage d'un point donné dans le filtrage spatial des points homologues	TT_DefSeuilDensiteResul	50

 TABLEAU 4.1 – Tableau des valeurs des seuils par défaut dans l'implémentation de la méthode *SIROP* dans MicMac.

La méthode proposée est déjà implémentée dans le logiciel photogrammétrique open-source MicMac et peut être utilisée avec les résultats issus des traitements classiques de cette chaîne. Les expériences sur différents jeux de données sont présentées dans le chapitre suivant afin d'analyser l'efficacité de la méthode.

Chapitre 5

Résultats

Sommaire

5.1	Introduction	63
5.2	Les jeux de données	64
5.3	Méthode d'évaluation	65
5.4	Résultats	66
5.4.1	Les indicateurs d'erreur	66
5.4.2	Couloir	66
5.4.3	Viabon 1	70
5.4.4	Corridor	72
5.4.5	LadyBug	74

5.1 Introduction

Ce chapitre présente les expérimentations menées et les résultats obtenus en utilisant notre méthode de raffinement de l'orientation interne/externe par une seconde itération avec les points homologues *SIROP* (voir le chapitre 4). Dans la première partie de ce chapitre, les acquisitions sont présentées, y compris les images acquises, les protocoles d'acquisition, l'ensemble caméras et objectifs utilisés et la vérité terrain. Ensuite, la chaîne de traitements photogrammétrique est présentée en détails ainsi que son paramétrage spécifique. Les résultats sont présentés en comparaison avec le traitement par processus classique afin d'évaluer l'apport de l'utilisation de la seconde itération en terme d'amélioration de la précision. Enfin, les problèmes subsistants de la méthode proposée et les perspectives seront discutés.

Dans toutes nos expérimentations décrites ci-dessous, les paramètres intrinsèques de la caméra sont déterminés par auto-calibration avec un modèle de distorsion choisi. Pour l'estimation de pose et l'ajustement de faisceaux, les seules observations utilisées sont les points homologues. Dans le cas de la cartographie mobile (présenté dans la section 5.4.5), des contraintes de rigidité sur le bloc de caméras¹ sont ajoutées. Les points terrain sont utilisés de la manière suivante :

- calcul de la similitude 3D permettant de passer du relatif à l'absolu (transformation à 7 paramètres) ;

1. contrainte de caméras rigides

- une fois les données géo-référencées, estimation de la précision terrain par calcul de la distance entre les coordonnées des points terrain et leurs estimation par intersection de faisceaux.

Des différents estimateurs, celui-ci est le plus important car il évalue le résultat de la photogrammétrie sur son objectif réel qui est sa capacité à mesurer des coordonnées sur le terrain. Nous avons choisi d'utiliser tous les points, sans séparer appuis et contrôle, car en général il y a beaucoup de points, donc nettement plus de contraintes que de paramètres à estimer ; par ailleurs, le léger biais introduit n'est pas gênant car il est identique pour les deux résultats (1^{ère} et 2^{ème} itérations) que l'on compare. Nous ne faisons aucune compensation sur ces points terrain, car le but est de vérifier à quel point notre méthode peut limiter les dérives, or celles-ci peuvent être complètement masquées après compensation.

5.2 Les jeux de données

Nous avons choisi 4 jeux de données de deux types différents pour analyser l'efficacité de la méthode. Il s'agit de deux acquisitions aéroportées par drone, d'une acquisition en sous-sol de type scène longue et d'une acquisition de type cartographie mobile. La vérité terrain est constituée de points mesurés sur le terrain dans chaque scène, soit par topométrie, soit par mesure GNSS. Ces informations sont détaillées dans le tableau 5.1 :

Jeux de données	<i>Couloir</i>	<i>Viabon 1</i>	<i>Corridor</i>	<i>LadyBug</i>
Type	Terrestre/Intérieur	Drone/Extérieur	Drone/Extérieur	Cartographie
Caméra	IGN Camlight	DJI Phantom 4	IGN Camlight	LadyBug
Focale	18mm	24mm	35mm	18mm
Nombre d'images	179	73	77	385
Taille de la scène	30 × 1.5	140 × 140	437 × 65	60 × 60
• $L(m) \times W(m)$				
Nombre de point terrains	33	7	10	63
• Type	Topométrie	GPS RTK	GPS RTK	Pts Homol
• Incertitude	$\sigma_{XYZ} = [0.8, 1, 0.4]$ mm	$\sigma_{XYZ} = [1, 1, 2]$ cm	$\sigma_{XYZ} = [1, 1, 2]$ cm	sous-pixellaire

TABLEAU 5.1 – Description des jeux de données expérimentaux.

Une attention particulière a été apportée au choix de jeux de données suffisamment différents pour que les évaluations de la méthode puissent être généralisables à différents types de chantiers. Les jeux de données *Couloir* et *Corridor* avec leur géométrie longue et étirée permettent de vérifier la correction de l'effet de courbure apportée par le processus *SIROP* sans utiliser d'observations terrain. Le jeu de données *Viabon 1* est une acquisition par drone classique de type couverture surfacique. Le jeu de données *LadyBug* est une acquisition de type cartographie mobile urbaine, avec un système multi-caméras. Dans tous les cas, même lorsque les acquisitions sont nativement en couleur, les images utilisées pour les traitements sont en niveaux de gris.

La figure 5.1 présente quelques exemples d'images acquises lors de ces missions.



FIGURE 5.1 – Exemples d’images acquises lors des expérimentations.

5.3 Méthode d’évaluation

Le traitement est fait par le logiciel photogrammétrique open-source MicMac. Chaque jeu de données est traité avec deux processus. Le processus classique se déroule avec les points homologues *SIFT* (l’implémentation de [Vedaldi, 2007]). Puis le processus *SIROP* est appliqué aux résultats obtenus pour extraire les nouveaux points homologues.

Sur tous les jeux de données, l’estimation de poses et l’ajustement de faisceaux sont tout d’abord effectués en coordonnées relatives, puis transformés dans le système de coordonnées terrain avec la transformation rigide à 7 paramètres estimée par moindres carrés en utilisant tous les points terrain (transformation de Helmert). Le processus de traitement des jeux de données est présenté sur la figure 5.2.

Dans tout les expériences, les paramètres internes des caméras sont estimés par auto-calibration (voir l’annexe C) en utilisant le modèle Fraser (voir l’annexe B.2), sauf pour la cartographie mobile qui utilise un modèle Fisheye Équidistant (voir l’annexe B.3). Une auto-calibration est effectuée lors du processus classique et également lors de la seconde itération.

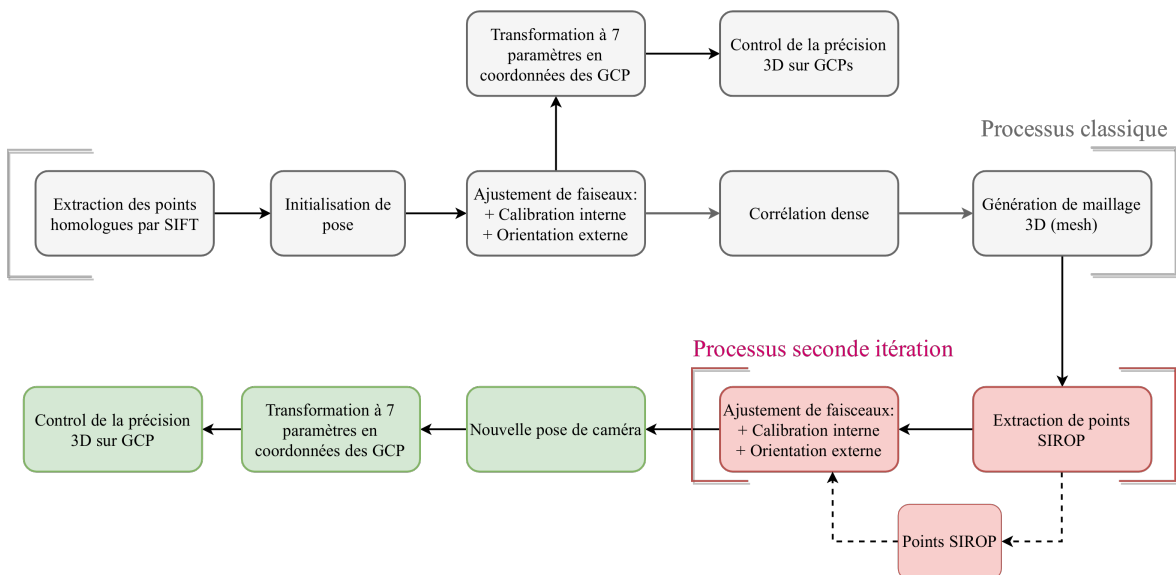


FIGURE 5.2 – Le processus de traitement des données.

Le maillage 3D est généré par la méthode de triangulation de Poisson en utilisant l’implémentation de [Kazhdan and Hoppe, 2013] intégrée dans la chaîne MicMac, à partir de nuages de points denses issus de la méthode [Pierrot-Deseilligny and Paparoditis, 2006]. Ce nuage de points dense est généré avec l’outil C3DC [Rupnik et al., 2017] de la chaîne MicMac avec un facteur de sous-échantillonnage de 4.

Tous les traitements sont faits sur un ordinateur portable avec processeur Core™ i5-5300U CPU @ 2.30GHz 64bits à 4 cœurs et 8Go de mémoire vive. Les calculs ne se déroulent que sur le CPU.

5.4 Résultats

5.4.1 Les indicateurs d'erreur

Chaque jeu de données est traité par le processus présenté sur la figure 5.2. L'ajustement de faisceaux n'utilise que les contraintes internes des images. Comme indiqué précédemment, aucune observation terrain n'est utilisée dans l'ajustement de faisceaux.

L'évaluation des résultats se fait avec plusieurs indicateurs :

- la distance euclidienne entre les coordonnées des points terrain issus de la triangulation d'images et les valeurs de ces coordonnées mesurées sur le terrain (d) ;
- l'écart-type des résidus de l'ajustement de faisceaux (en pixels) ;
- la carte des résidus des points homologues codés en couleur en géométrie image ;
- l'histogramme des résidus des points homologues et leur multiplicité ;
- les résidus des points terrain en géométrie image.

5.4.2 Couloir

Ce jeu de données est acquis dans un couloir souterrain d'une longueur de 30m à l'ENSG², Marne-la-Vallée. Ce jeu de données est représentatif de la problématique de modélisation 3D des égouts qui est à l'origine du financement de la thèse par l'IGN. Il s'agit d'une scène de géométrie longue, en intérieur, avec des murs faiblement texturés à motif répétitif.

Les 33 points terrain sont répartis le long du couloir. Un aperçu des images de ce jeu de données se trouve à la figure 5.3. La distribution des points terrain est présentée sur la figure 5.4.



FIGURE 5.3 – Quelques images du jeu de données *Couloir*. Caméra photogrammétrique légère IGN CamLight [Martin et al., 2014], longueur focale 18mm.

2. École nationale des sciences géographiques

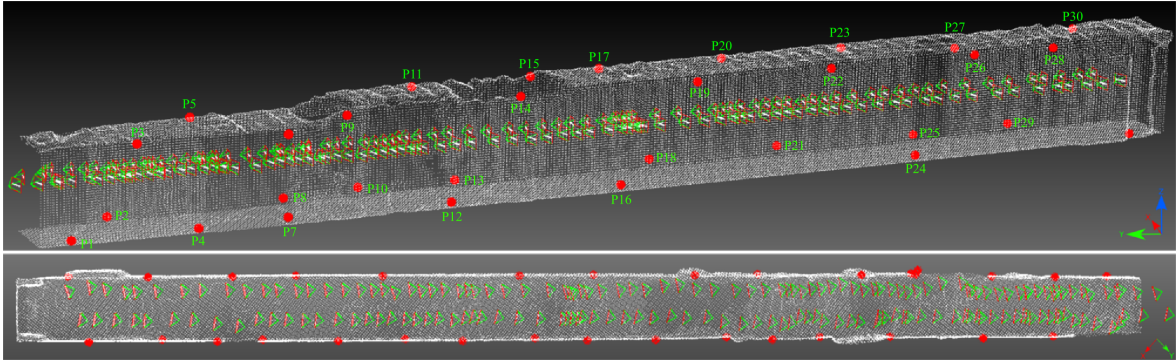


FIGURE 5.4 – Le jeux de données *Couloir* et ses points terrain (rouge).

La figure 5.5 illustre un des avantages de l'extraction des points par *SIROP* : les points *SIROP* sont mieux distribués dans tout l'espace de l'image.

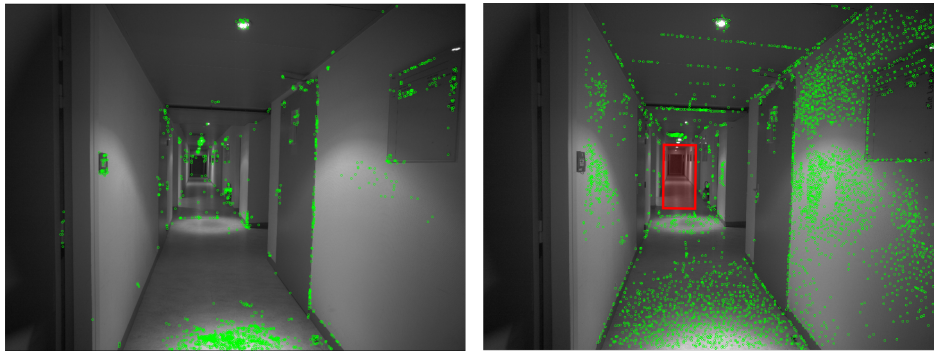


FIGURE 5.5 – Comparaison de la distribution des points. **Gauche** : SIFT, **Droite** : *SIROP*, le carré rouge représente les zones lointaines qui sont une limitation actuelle

Le critère de résolution optimale pour le choix de l'image maîtresse est illustré en figure 5.6. Chaque couleur dans cette figure représente une partie de la scène associée à une image maîtresse distincte.

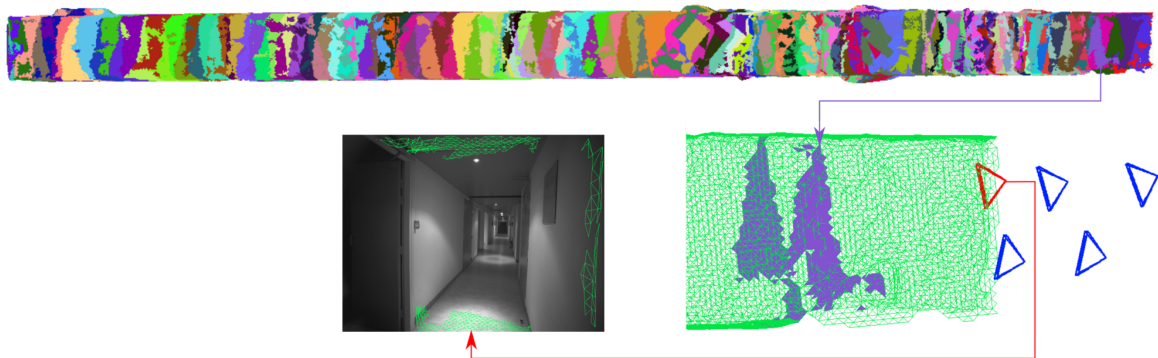


FIGURE 5.6 – Partition de la scène en images maîtresses par la condition de la résolution optimale. **En haut** : chaque couleur est la partie de la scène assignée à une image maîtresse. **En bas** : un image maîtresse (rouge) et ses secondaires (bleu).

SIROP réduit aussi les résidus des points homologues (voir figures 5.7 et 5.8) en moyenne d'un facteur 2. Plus précisément, le résidu moyen des points chute de 0.45 pixel à 0.25 pixel.

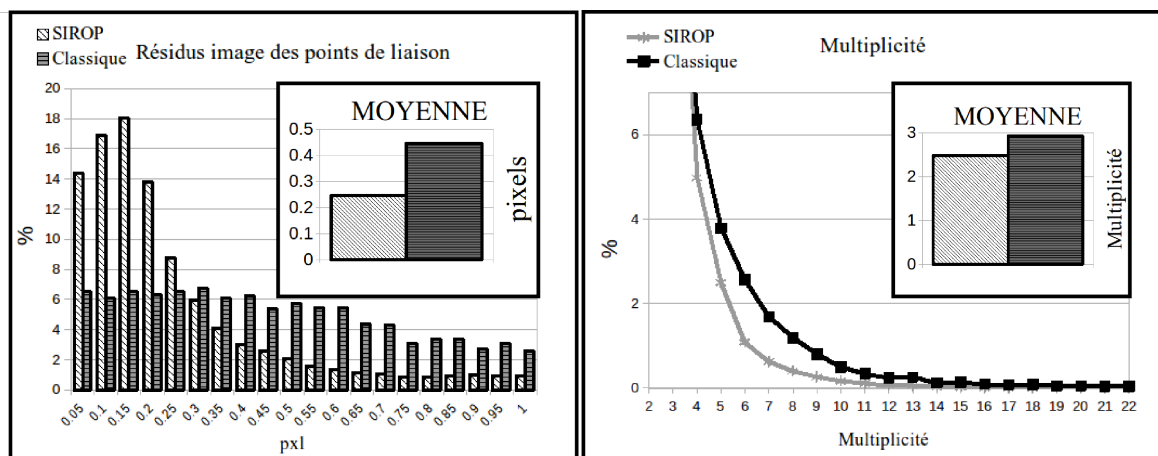


FIGURE 5.7 – Statistiques sur les points homologues entre les deux processus : *SIROP* et classique *SIFT*.

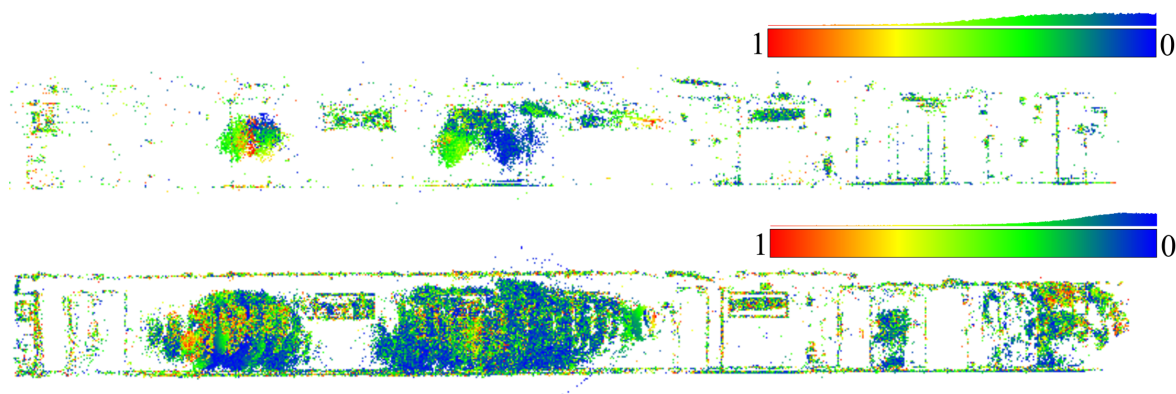


FIGURE 5.8 – Les points dans l’espace objet avec leurs résidus codés en couleur. L’histogramme d’erreur de re-projection dans la gamme de $[0,1]$ pixel est également représenté sur l’échelle de couleur. **Haut** : *SIFT*, **Bas** : *SIROP*.

Par contre, ce jeu de données présente une géométrie légèrement défavorable pour la multiplicité des points *SIROP*. Dans cette scène longue, une grande variation du GSD³ est observée au sein de chaque image, une variation causée par la direction de visée fuyante de la caméra vers le fond du couloir. Par conséquent, l’algorithme choisit dans chaque image la partie la plus résolue de la scène pour extraire des points, c’est-à-dire la partie en champ proche et de ce fait ne couvre pas tout l’espace image (voir la figure 5.6). Les points de la partie lointaine de la scène, au fond du couloir, sont ignorés (voir la figure 5.9). La multiplicité moyenne des points homologues est passée de 2.9 à 2.5. Les solutions envisagées pour résoudre cette difficulté sont discutées dans les perspectives (voir la section 6.3.3).

La Figure 5.10 présente la distance euclidienne entre les coordonnées déterminées par photogrammétrie et celles de la vérité terrain de tous les points terrain pour les deux processus : classique et *SIROP*. On peut constater que le résidu moyen est réduit d’un facteur 1.2, 2.1 et 3.7 respectivement en X , Y et Z .

Le contrôle avec les points terrain est montré sur la figure 5.10. Le résidu moyen des points terrain diminue d’un facteur 2, de 1.33 cm à 0.63 cm.

3. taille de pixel au sol

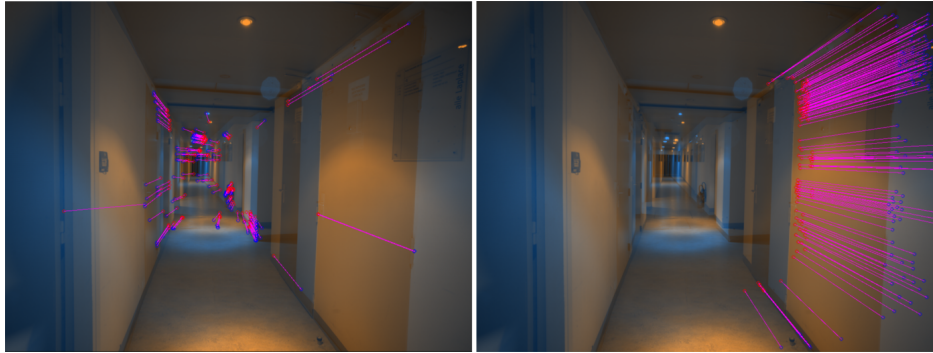


FIGURE 5.9 – Les prises de vue avec une grande variation de GSD sont légèrement défavorables à la multiplicité des points *SIROP*

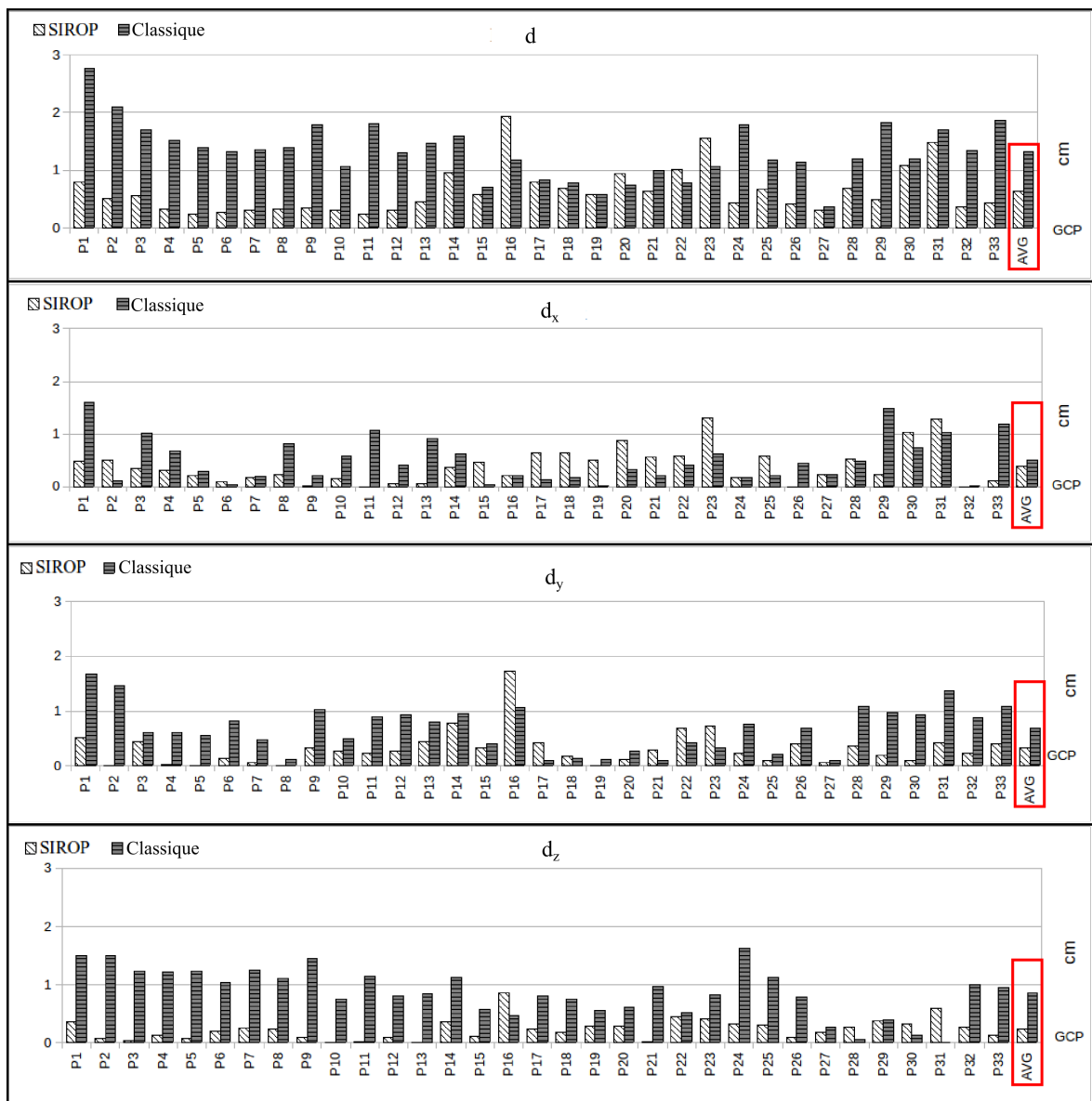


FIGURE 5.10 – La distance entre les coordonnées des points terrain - *Couloir*. De haut en bas : la distance euclidienne, en X et Y et Z. La distance moyenne est encadrée en rouge.

5.4.3 Viabon 1

Le jeu de données *Viabon 1* est un chantier typique d'acquisition par drone. Il s'agit d'une surface rectangulaire relativement plane avec cependant quelques bâtiments qui créent du relief et permettent ainsi de limiter la corrélation entre les paramètres internes et externes. De la végétation est présente dans les images. Les 7 points terrain de ce jeu de données sont relevés par GNSS RTK. Le drone DJI Phantom 4 Pro et sa caméra intégrée est utilisée pour ce jeu de données. Quelques images acquises sont présentées dans la figure 5.11.



FIGURE 5.11 – Quelques images du jeu de données *Viabon 1*.

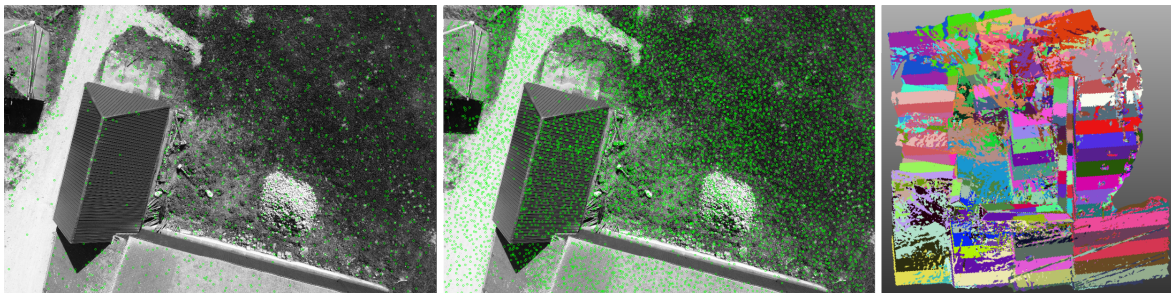


FIGURE 5.12 – Les points homologues *SIFT* (Gauche), *SIROP* (Milieu) et la partition de la scène selon les différentes images maîtresses (Droite).

Sur la figure 5.12, on peut constater que le nombre de points *SIROP* est plus grand et que leur distribution dans l'espace image est plus homogène. L'image à droite représente la partition de la scène entre les images maîtresses. Les figures 5.13 et 5.14 nous donnent une indication de l'erreur de re-projection des points en comparant entre la méthode *SIROP* et le processus classique. La carte des résidus en couleur montre une amélioration significative dans toute la scène de 0.58 pixel à 0.31 pixel en moyenne. La multiplicité des points est aussi améliorée, mais de manière moins sensible : les points *SIROP* ont une multiplicité moyenne de 5.6 comparée à 5.2 pour les points *SIFT*.

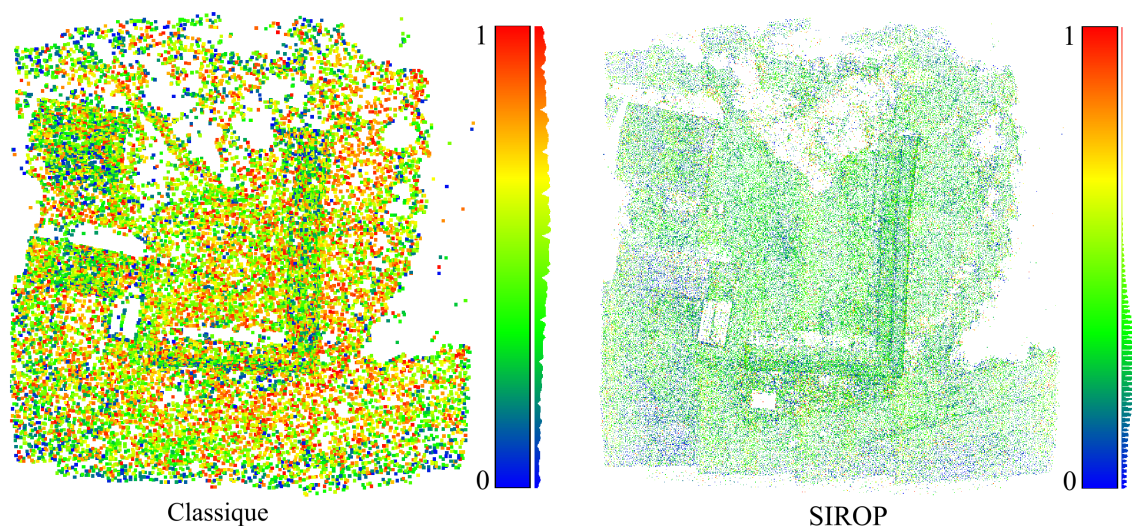


FIGURE 5.13 – Les points dans l'espace objet avec leurs résidus codés en couleur. *SIFT* (Gauche) et *SIROP* (Droite) (*Viabon 1*).

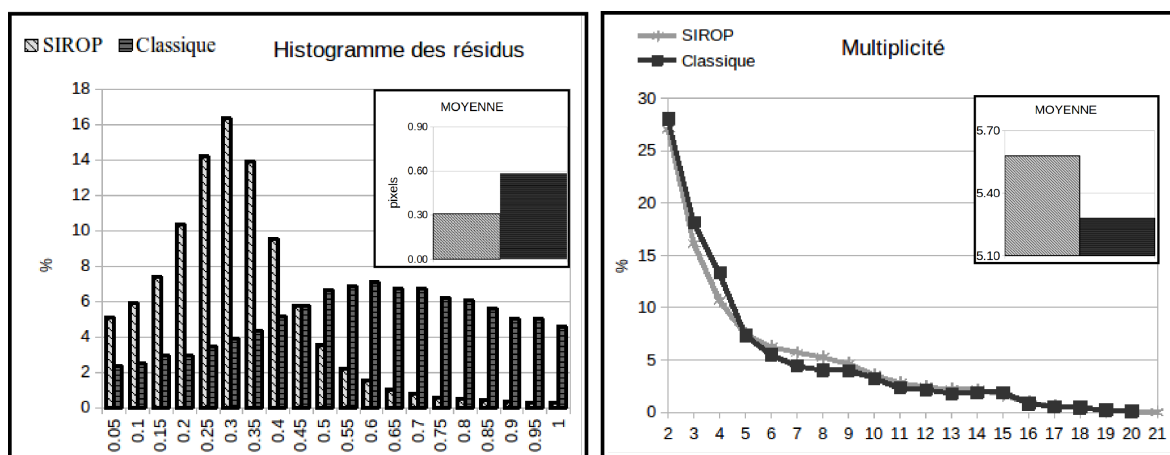


FIGURE 5.14 – Statistiques sur les points homologues entre les deux processus : *SIROP* et classique sur le jeu de données *Viabon 1*.

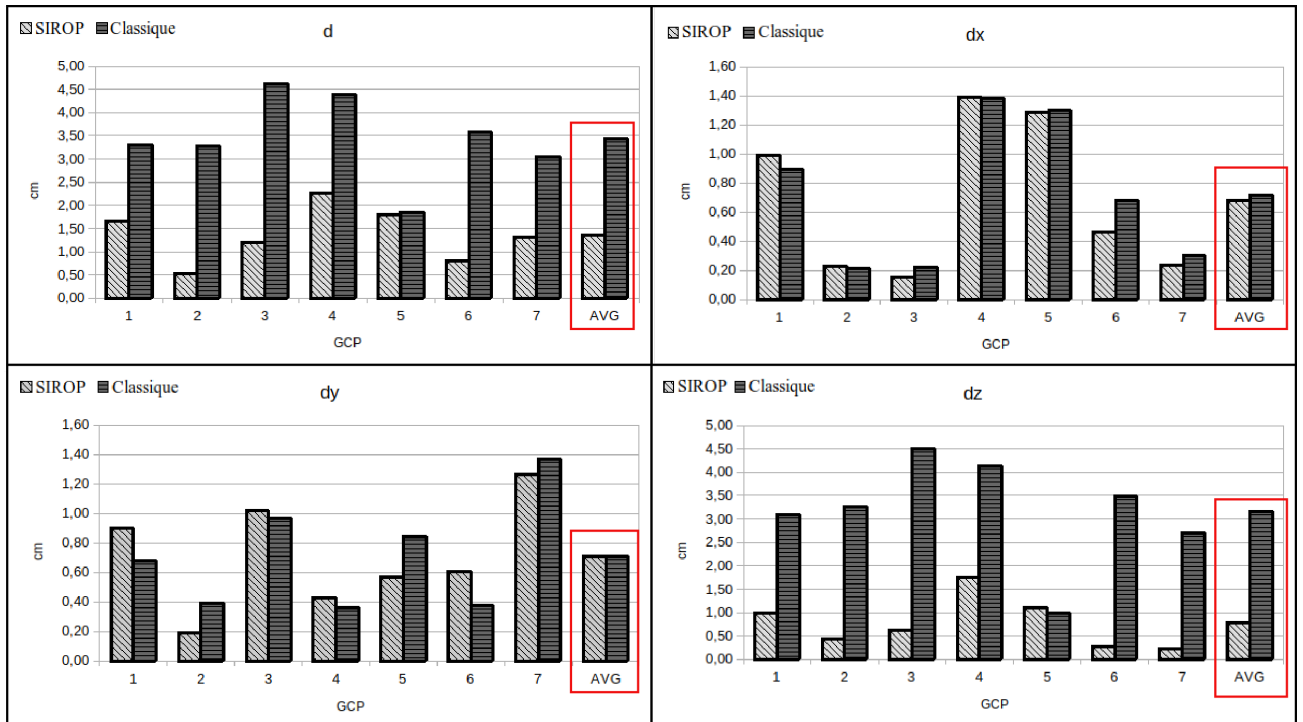


FIGURE 5.15 – Comparaison des résidus en 3D sur les points terrain entre les méthodes *SIROP* et classique.

Le contrôle avec les mesures terrain est montré sur la figure 5.15. L'orientation relative à l'issue du processus *SIROP* étant meilleure, l'orientation absolue s'en trouve améliorée aussi. Le résidu moyen des points terrain diminue de 3.43 cm à 1.37 cm.

5.4.4 Corridor

Le jeu de données *Corridor* est une acquisition drone en configuration linéaire longue, c'est le cas souvent rencontré dans les applications de surveillance de digues ou de chemin de fer par exemple. Ce type d'acquisition occasionne fréquemment le problème classique de courbure ou *dome effect* (voir Chapitre 3). En utilisant ce jeu de données, l'efficacité des points *SIROP* peut être évaluée par la correction de cet effet de distorsion.

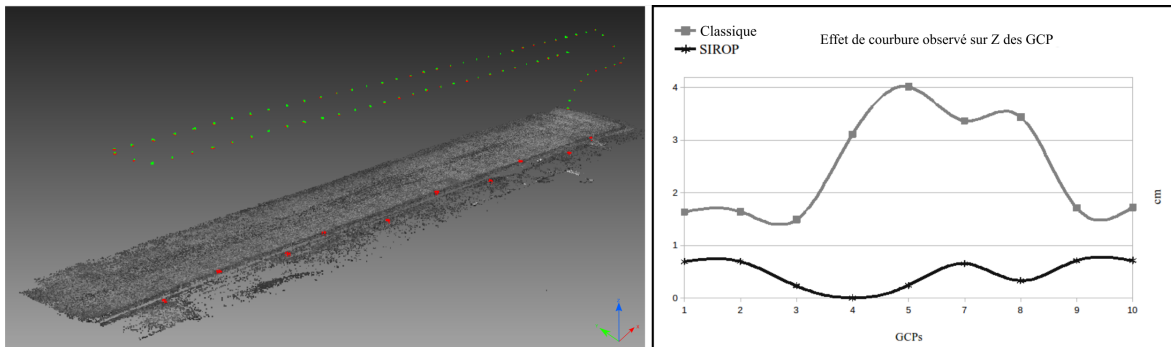


FIGURE 5.16 – **Gauche** : le jeu de données *Corridor* et ses points terrain (points rouges). **Droite** : l'effet de courbure observé sur l'axe *Z* à l'aide des points terrain le long de la scène.

La figure 5.16 montre la réduction sensible de l'effet de courbure par les points *SIROP*. Sur la figure de droite, l'effet de courbure est observé clairement sur les résidus

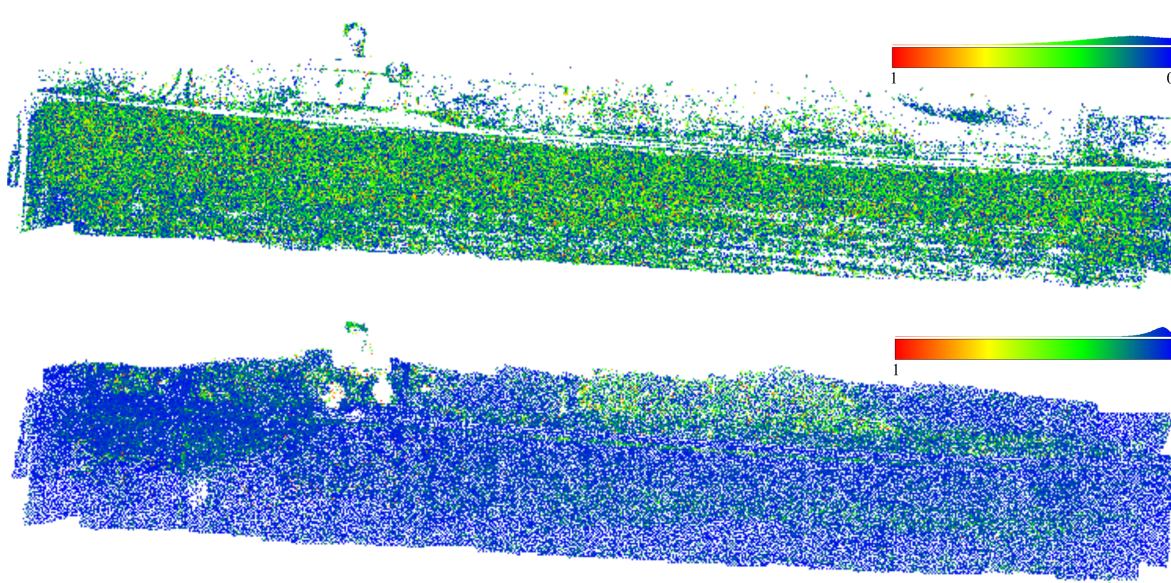


FIGURE 5.17 – Points dans l’espace objet avec leurs résidus images (pixels) codés en couleur avec *SIFT* (**Haut**) et avec *SIROP* (**Bas**) (*Corridor*).

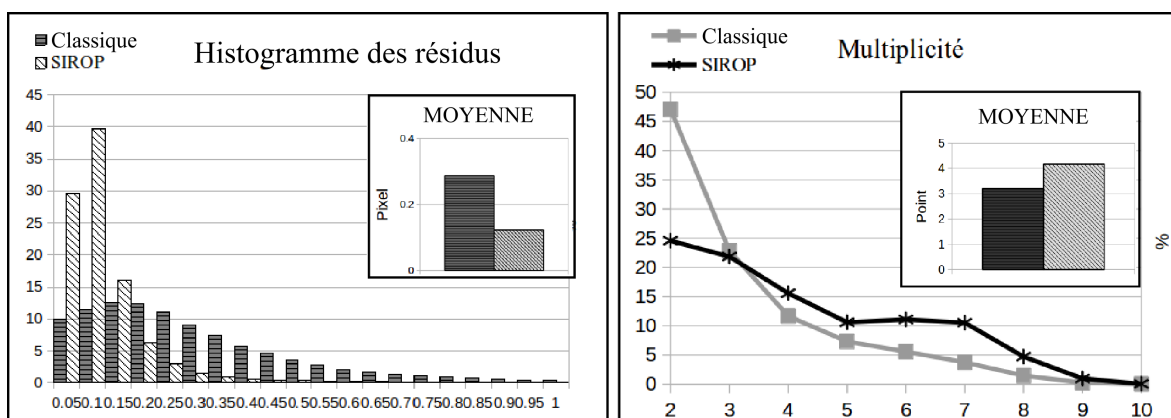


FIGURE 5.18 – Statistiques sur les points homologues entre les deux processus : *SIROP* et classique sur le jeu de données *Corridor*.

en Z des points terrain le long du profil de la scène, lorsque les données par la méthode classique. Les points *SIROP* permettent de limiter cet effet en diminuant les grands écarts en Z sur les points terrain 4,5,7,8. C’est un avantage de la méthode *SIROP* puisque nous utilisons ici que les contraintes internes aux images pour corriger cet effet.

La méthode *SIROP* diminue également l’erreur de re-projection des points en relatif comme on peut le constater sur la carte des résidus (la figure 5.17) et l’histogramme de re-projection (la figure 5.18). La valeur moyenne de l’erreur de re-projection est passée de 0.29 pixel en méthode classique à 0.12 pixel en méthode *SIROP*. La multiplicité des points homologues est augmentée, de 3.2 à 4.1 en moyenne sur tous les points.

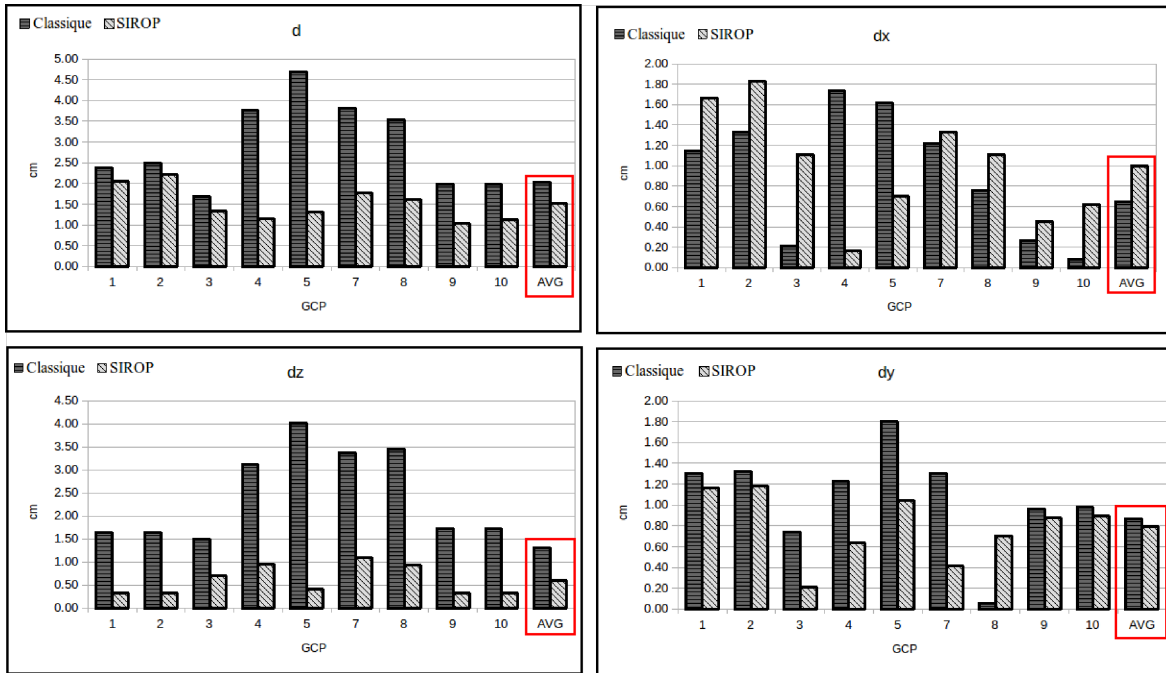


FIGURE 5.19 – Comparaison des résidus en 3D sur des points terrain entre les méthodes *SIROP* et Classical.

De la même façon, la présentation détaillée des résidus 3D sur les points terrain de la figure 5.19 révèle une amélioration notable sur la composante en Z , avec un ratio de diminution de 2.18.

5.4.5 LadyBug

Il s'agit d'une acquisition de type cartographie mobile. L'acquisition est réalisée avec le véhicule Stéréopolis de l'IGN équipé d'une caméra panoramique à 6 têtes (LadyBug 360). Chaque tête de caméra est équipée d'un objectif fisheye de focale équivalente 18.4mm, et les champs des différentes caméras se recouvrent entre eux.



FIGURE 5.20 – La caméra multi-têtes panoramique LadyBug.

Le parcours forme une boucle autour d'un ensemble de bâtiments. Le véhicule a acquis un bloc d'images (à chaque sommet de prise de vue, 6 images sont acquises par 6 têtes de caméras) en déclenchant automatiquement tous les 3m parcourus. Le parcours fait environ 230m de long (voir la figure 5.22).

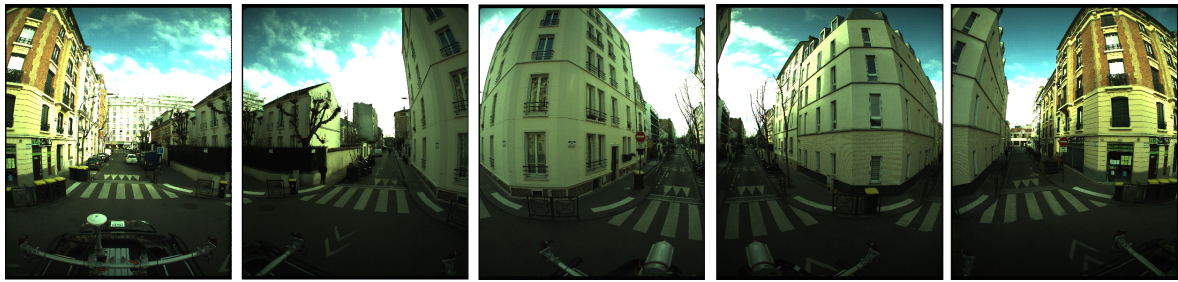


FIGURE 5.21 – 5 images du même bloc d’images de la caméra LadyBug.

Il y a seulement 5 images dans chaque bloc de 6 images qui sont utilisées (voir la figure 5.21). La 6^{ème} image est prise par la tête de caméra pointée vers le ciel. Il n’y a pas assez de points homologues pour utiliser cette 6^{ème} image dans le processus de traitement standard ⁴

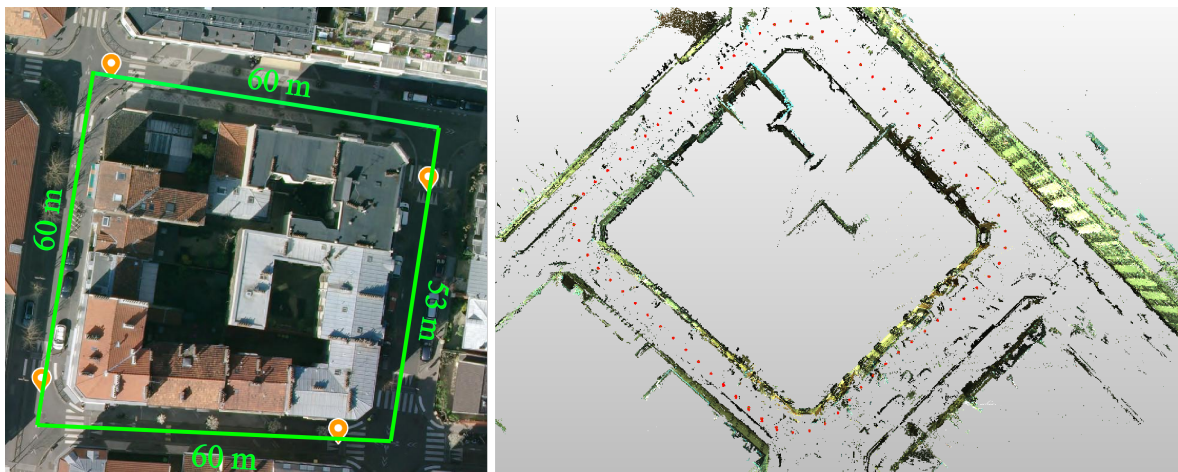


FIGURE 5.22 – **Gauche** : L’ensemble d’immeubles avec sa taille et l’indication des 4 points terrains utilisés pour la mise à l’échelle. **Droite** : Nuage de points et les blocs d’images (en rouge).

Il n’y a pas de vérité terrain pour ce jeu de données. Quelques coordonnées de points au sol sont prises sur la plateforme web Géoportail (voir la figure 5.22) pour pouvoir mettre le résultat à l’échelle métrique afin de mieux l’interpréter. Les données sont traitées sans la contrainte de fermeture de boucle pour être considérées comme issues d’une acquisition longue linéaire car c’est ce point qui servira de contrôle. Dans ce sens, les blocs d’images au début et à la fin de l’acquisition ne doivent pas avoir des liens homologues entre eux. Le stratégie suivante est donc utilisée dans la recherche de points homologues :

- la recherche de points homologues est réalisée morceaux par morceaux entre les blocs d’images ;
- pour chaque bloc d’images, la recherche de points homologues n’est effectuée que sur les 5 blocs précédents et les 5 blocs suivants.

La figure 5.23 représente les liens homologues entre les blocs d’images. Les champs de vue (représenté par les triangles verts) des blocs de fermeture se recouvrent, mais il n’existe pas de points de liaison entre eux.

⁴. Cependant, elle est probablement très utile pour rigidifier le bloc et sera intégrée dans les travaux ultérieurs

On peut donc estimer précisément l'erreur de fermeture de boucle grâce aux points communs au début et à la fin de la boucle, sans l'utilisation de points terrain (voir la figure 5.23, droite).

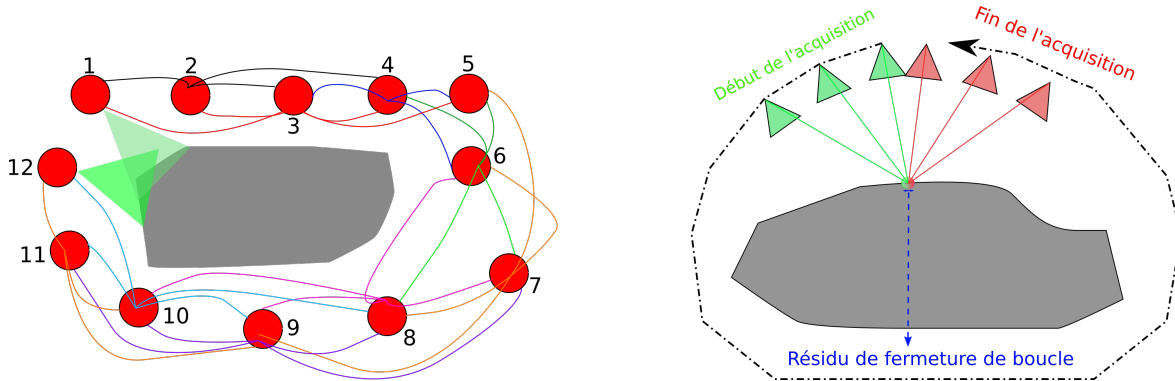


FIGURE 5.23 – L'erreur de fermeture de boucle : **Gauche** : les liens des points homologues pour pouvoir considérer une acquisition en boucle comme un cas linéaire ; **Droite** : l'estimation de l'erreur de fermeture de boucle avec deux triangulations différentes.

Sur ce jeu de données, en plus du processus de raffinement par les points *SIROP*, on évalue l'apport de la compensation avec la contrainte de caméra rigide (voir la section 5.22).

Les points homologues pour le contrôle de la fermeture de la boucle (points homologues de contrôle) sont détectés dans 6 images, 3 au début et 3 à la fin de l'acquisition, sur la même partie de la scène. Pour chaque point homologue de contrôle, on détermine deux positions 3D estimées par deux triangulations différentes entre le 1^{er} et le 2^{ème} passage, calculées par pseudo-intersection des faisceaux. L'écart entre deux points 3D triangulés est le résidu de fermeture de boucle. Les points homologues de contrôle sont mesurés par *SIFT* sur les images en pleine résolution. Au total 63 points homologues de contrôle ont été sélectionnés (les points manifestement faux ont été écartés).

Les quatre méthodes de traitements ci-dessous sont effectuées sur ce jeu de données, afin de comparer le résidu entre eux (voir la figure 5.24) :

- le traitement classique (Classique) ;
- la compensation avec la contrainte de caméra rigide (Rigide) ;
- la compensation avec les points *SIROP* (*SIROP*) ;
- la compensation avec la contrainte de caméra rigide et les points *SIROP* (*SIROP* + Rigide).

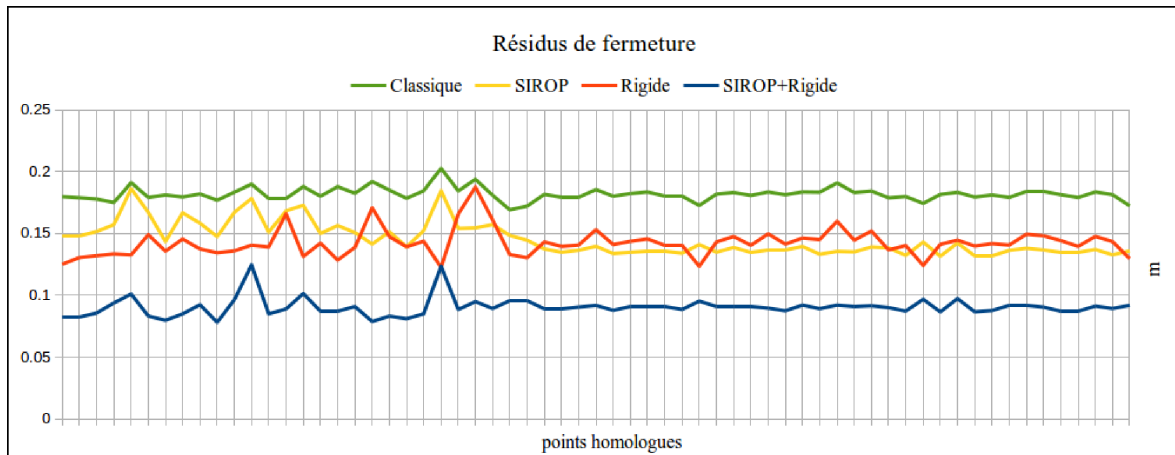


FIGURE 5.24 – Comparaison des résidus de fermeture de boucle issus des différents tests.

La figure 5.24 montre la comparaison des résidus de fermeture de boucle entre les 4 méthodes de traitement. L'axe horizontal est les points homologues de contrôle et l'axe vertical représente leur résidu de fermeture. Le résidu moyen et son écart-type sont présentés dans le tableau 5.2.

Méthode	Moyenne (cm)	Écart Type (cm)	Max (cm)	Min (cm)
Classique	18.19	0.53	20.26	16.92
SIROP	14.54	1.35	18.64	13.15
Rigide	14.24	1.13	18.71	12.24
SIROP + Rigide	9.05	0.77	12.46	7.81

TABLEAU 5.2 – Tableau de comparaison des résidus de fermeture de boucle entre les 4 méthodes.

Le résidu moyen de fermeture de boucle passe de 18.19cm en méthode classique à 14.54cm en méthode *SIROP* et à 14.24cm en contrainte de caméra rigide. Ces deux méthodes permettent de limiter la dérive à la fin de cette acquisition longue et linéaire. Une combinaison des deux méthodes *SIROP* et la contrainte caméra rigide réduit nettement la dérive. Le résidu moyen passe de 18.19cm en méthode classique à 9.05cm en combinant les deux méthodes.

Chapitre 6

Conclusions et perspectives

Sommaire

6.1 Conclusions	79
6.2 Discussion	80
6.2.1 Plusieurs itérations	81
6.2.2 Temps de calcul	82
6.2.3 Accélération du processus classique	82
6.3 Perspectives	83
6.3.1 Images satellites	83
6.3.2 Optimisation des paramètres de la traitement classique	83
6.3.3 Mixage SIFT/SIROP	84
6.3.4 Modèle 3D théorique	84

6.1 Conclusions

Cette thèse aborde le problème de l'amélioration de la précision de la mise en place des images en utilisant uniquement des contraintes internes aux images. Ainsi, la méthode proposée peut être utilisée comme un post-traitement compatible avec tous les processus de traitements photogrammétriques classiques. Ce processus se révèle utile dans de nombreux cas, particulièrement ceux où des points d'appui ou des observations faites sur le terrain sont coûteuse à acquérir.

La méthode consiste principalement à extraire de nouveaux points homologues avec la meilleure qualité photogrammétrique possible. La qualité de l'extraction des points homologues influe fortement sur les résultats des étapes suivantes. Notre méthode utilise l'information 3D de la scène sous la forme d'un maillage 3D, combinée avec l'orientation interne et l'orientation externe qui résultent du traitement classique pour guider l'extraction des points homologues *SIROP*. Elle permet d'extraire des points en optimisant leur distribution spatiale, leur précision ainsi que leur multiplicité.

Ce manuscrit décrit l'idée principale, la formation de l'algorithme et l'ensemble des tests réalisés pour évaluer son efficacité à travers différents scénarios, y compris une scène aérienne, une prise de vue en corridor, une longue acquisition en sous-terrain et une de cartographie mobile. Les acquisitions sont aussi réalisées avec différents types de caméras comme la caméra du drone DJI, la caméra photogrammétrique CamLight de l'IGN ou encore la caméra panoramique LadyBug. Cette méthode a prouvé son efficacité, sa robustesse ainsi que son application sur différents types de chantiers.

Plus en détail, le résidu 3D calculé sur les points terrain est réduit d'un facteur de 2.1, 2.83, 1.33 en moyenne, respectivement sur les jeux de données *Couloir*, *Viabon 1* et *Corridor*, en comparaison entre le traitement classique et la méthode *SIROP*. En coordonnées relatives, l'évaluation des points *SIROP* est effectuée sur l'erreur de reprojection, avec un facteur d'amélioration moyen de 1.9 à 2.4. La multiplicité des points est également améliorée sur deux jeux de données, *Viabon 1* et *Corridor*, respectivement, de 5.2 à 5.6 et de 3.2 à 4.17. Sur le jeu de données *Couloir*, en raison de la stratégie du choix de l'image maîtresse, qui favorise la partie résolue de la scène, la multiplicité est dans ce cas légèrement diminuée (2.9 à 2.5).

Un récapitulatif des résultats issus des tests de ces différents jeux de données est présenté dans le tableau 6.1.

Jeu DD	Méthode	Valeur					
		d [cm]	d_x [cm]	d_y [cm]	d_z [cm]	Rep_Err [pxl]	Multiplicity [Points]
Couloir	Classique	1.32	0.50	0.68	0.87	0.45	2.92
	<i>SIROP</i>	0.63	0.39	0.32	0.23	0.24	2.50
Viabon 1	Classique	3.43	0.71	0.71	3.17	0.59	5.2
	<i>SIROP</i>	1.21	0.64	0.69	0.52	0.31	5.6
Corridor	Classique	2.03	0.65	0.87	1.31	0.29	3.20
	<i>SIROP</i>	1.52	1.00	0.79	0.60	0.12	4.17

TABLEAU 6.1 – Comparaison entre le traitement classique et le processus *SIROP* sur les résidus 3D, erreur de reprojection et multiplicité des points.

Le jeu de données *LadyBug* montre une autre utilisation de la méthode dans le cas de la cartographie urbaine, correspondant à une acquisition longue linéaire avec un système de caméras rigide. La combinaison entre les points *SIROP* et la contrainte de rigidité inter-caméras dans le dernier calcul d'ajustement de faisceaux diminue nettement la dérive de l'acquisition. Sur un parcours de 230m, une dérive de 18.19cm en moyenne est apparue en utilisant le traitement classique. L'utilisation de la contrainte de rigidité inter-caméras diminue la dérive à 14.24cm, soit 21.7%. De plus, les points *SIROP* réduisent le résidu de fermeture de boucle à 14.54cm, soit 20% d'erreur. La combinaison entre *SIROP* et la contrainte de rigidité donne une dérive moyenne de 9.05cm sur cette acquisition, soit 50.24% de diminution de dérive par rapport à la méthode classique.

Le problème d'estimation de poses des images de systèmes de numérisation mobile est profondément étudié dans [Cannelle, 2013]. L'expérimentation avec ce jeu de données montre également l'efficacité de la méthode *SIROP*, combinée avec la contrainte de rigidité dans l'ajustement de faisceaux, pour améliorer la précision de l'orientation des images. Dans ce cas l'effet de dérive est grandement corrigé en utilisant uniquement les contraintes internes au système d'acquisition.

6.2 Discussion

Dans cette partie, quelques points divers autour de l'algorithme sont discutés afin d'avoir un regard plus clair. Les points principaux sont l'influence du nombre d'itérations du processus *SIROP* et les temps de traitement.

6.2.1 Plusieurs itérations

Le processus *SIROP* utilise l'orientation comme entrée et l'améliore, on pourrait espérer qu'il y ait un gain à itérer plusieurs fois ce processus. *À priori*, notre intuition est que ce gain est plutôt marginal, car les orientations relatives entre images voisines sont très peu modifiées; il nous a cependant semblé intéressant de le tester du point de vue expérimental.

Afin de prouver cela, une évaluation de la convergence des paramètres internes de la caméra après des itérations *SIROP* est examinée sur le jeu de données *Corridor*. Le traitement classique est effectué au début, suivi du maillage 3D, de l'orientation externe et de la calibration de la caméra qui sont utilisés pour extraire les points *SIROP*. Après l'extraction de ces nouveaux points, un ajustement de faisceaux est effectué en utilisant tous ces points comme seules observations pour raffiner les paramètres internes et externes de toutes les images. Les itérations *SIROP* suivantes utilisent l'orientation externe, la calibration et le mesh en sortie de l'itération précédente.

La figure 6.1 montre que l'amélioration la plus remarquable est entre le traitement classique et le processus *SIROP* qui suit (*2^{ème} itération*). Les itérations de *SIROP* suivantes ne font pas évoluer les résidus terrain et image. L'évolution de la focale et du point principal sont présentées sur la figure 6.2. On peut constater que la focale converge après la *2^{ème}* itération et que le point principal ne varie pas significativement.

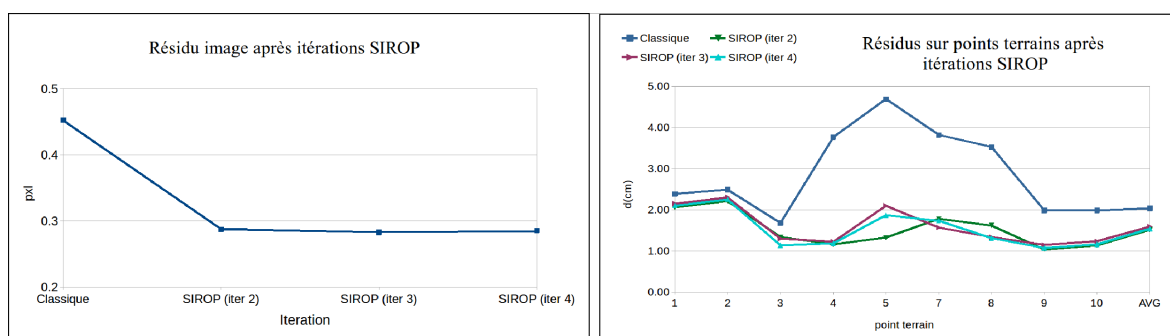


FIGURE 6.1 – Résidus images et terrains après l'exécution successive du processus *SIROP*.

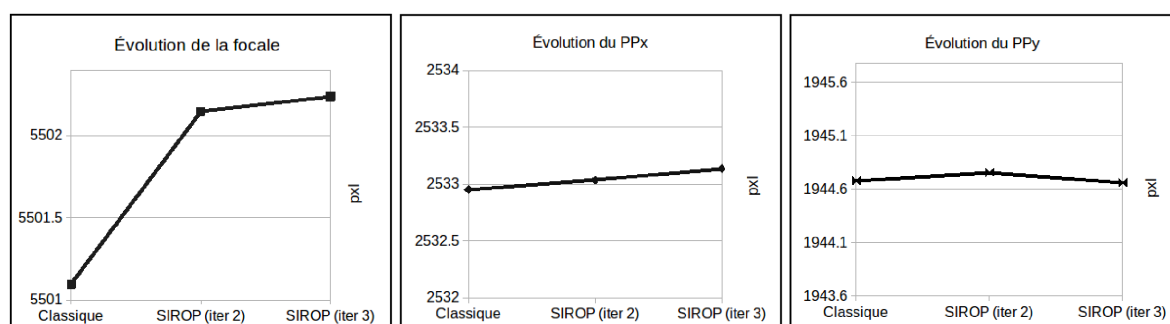


FIGURE 6.2 – Évolution des paramètres internes de la caméra après les itérations *SIROP*.

Le test de plusieurs itérations montre que l'amélioration de la *2^{ème}* itération est la plus sensible. Il est possible d'exécuter le processus *SIROP* plusieurs fois à la suite, mais l'amélioration sur les résidus terrain et image ainsi que l'évolution de la calibration de la caméra sont minimes.

6.2.2 Temps de calcul

L'aspect du temps de calcul est aussi à discuter. Celui-ci dépend principalement des facteurs suivants :

- nombre d'image ;
- nombre de triangle dans la maillage 3D ;
- niveau de réduction du nombre de points homologues (voir section 4.2.3).

Le processus d'extraction des points *SIROP* dans notre implémentation dans *MicMac* est parallélisée. L'algorithme est simple à paralléliser à plusieurs niveaux :

- le calcul des occlusions image par image (4.2.1) ;
- le traitement principal est réalisée par triangle indépendamment.

Malgré la complexité de l'algorithme, le temps de calcul ajouté par ce post-traitement est acceptable. L'évaluation du temps de calcul est fait sur le jeu de données *Corridor* avec 77 images de 20 mégapixels, sur le PC précédemment décrit 5.3.

L'extraction des points *SIROP* prend 12,5 minutes avec un maillage 3D de 281000 triangles. L'ajustement de faisceaux prend 1 minute et 30 secondes de plus. Un temps total de 14 minutes est donc à ajouter pour le post-traitement *SIROP* sur ce jeu de données. Le processus classique prend 96 minutes pour l'extraction des points homologues *SIFT*, 20 minutes pour l'estimation de pose et 53 minutes pour l'appariement dense et la génération du maillage 3D. Au total, 169 minutes sont nécessaires pour le traitement classique et 14 minutes pour le raffinement avec les points *SIROP*.

Ainsi *SIROP* ne demande pas un temps de calcul significatif pour obtenir une amélioration notable.

6.2.3 Accélération du processus classique

Le processus *SIROP* peut également être utilisé pour accélérer la chaîne de traitement classique pour atteindre une bonne précision. Les études de [Dwarakanath et al., 2012; Nikolova, 2014] montrent un ralentissement en temps de calcul de *SIFT* avec l'augmentation de la résolution de l'image. Au lieu d'espérer un résultat final précis en utilisant les points *SIFT* extraits à pleine résolution à la première étape, il est possible d'effectuer tout d'abord un traitement classique avec *SIFT* à basse résolution puis un post-traitement par *SIROP* pour améliorer la précision avec un temps supplémentaire d'environ 10% du temps d'exécution total de la chaîne classique.

Pour mieux analyser l'efficace de l'accélération du traitement, prenons l'exemple des jeu de données *Corridor*. Le temps de traitement dans les différents cas suivants sont :

- chaîne classique + *SIFT* avec images en pleine résolution : 169 minutes ;
- chaîne classique + *SIFT* avec images sous-échantillonnées d'un facteur 2 : 121 minutes ;
- chaîne classique + *SIFT* avec images sous-échantillonnées d'un facteur 2 + raffinement par *SIROP* : 135 minutes.

Le traitement avec une seconde itération par *SIROP* et l'initialisation par la chaîne classique avec *SIFT* en basse résolution dans ce cas diminue de 20% le temps de calcul total.

6.3 Perspectives

6.3.1 Images satellites

Dans le cadre de cette thèse, le processus *SIROP* a été développé et testé pour de la photogrammétrie acquise par des appareils photos (caméras coniques). Cependant, cette méthode aurait un intérêt potentiel dans le cadre des images satellites pour les raisons suivantes :

- avec ces images on dispose déjà d’une orientation approchée et d’un modèle 3D approché (type SRTM), on peut donc directement aborder la deuxième itération ;
- ces images sont souvent de très grande taille, donc un processus d’appariement qui utilise les informations de localisation approchée y est particulièrement adapté ;
- enfin ces images sont souvent diachroniques¹, ce qui rend l’appariement difficile et rend à nouveau intéressant un processus qui utilise les informations de localisation *a priori*.

On a donc testé l’utilisation des points *SIROP* pour raffiner l’orientation d’images satellite de type RPCs avec des points homologues ([Rupnik et al., 2016]) ou l’extraction des points homologues dans les images satellites diachroniques. Une comparaison est effectuée sur un extrait d’une paire d’images diachroniques entre *SIFT* et *SIROP* montre un nombre de points beaucoup plus élevé extrait par *SIROP* (voir la figure 6.3).

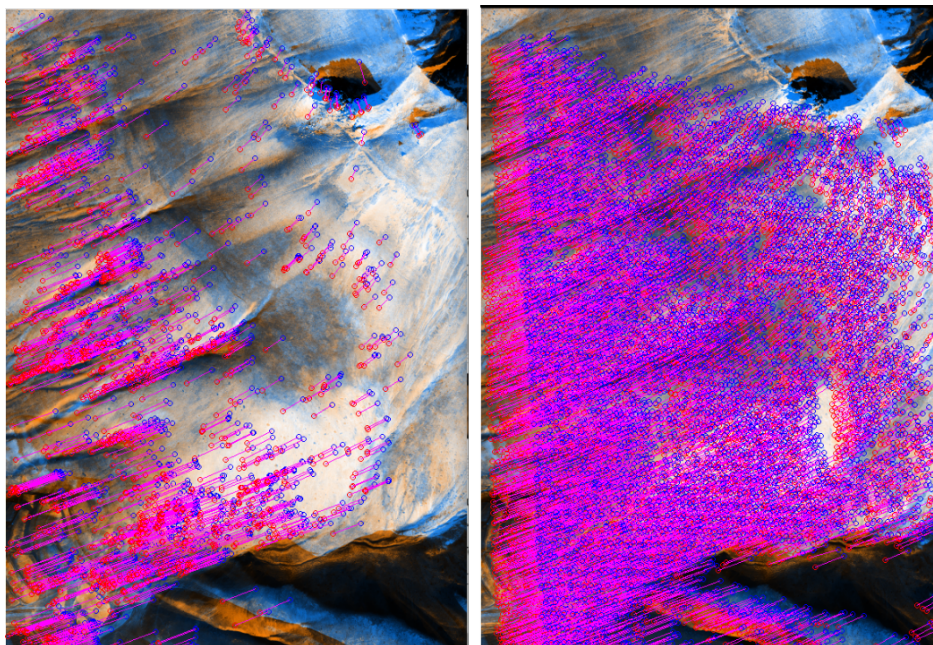


FIGURE 6.3 – L’extraction des points homologues sur les images diachroniques : **Gauche** : *SIFT* **Droite** : *SIROP*

Ces tests ont aussi permis de montrer que l’intégration logicielle de *SIROP* fonctionnait avec les différents types de capteurs intégrés dans la chaîne MicMac.

1. acquises à des dates différentes

6.3.2 Optimisation des paramètres de la traitement classique

La méthode *SIROP* utilise les résultats de la traitement classique comme donnés d'entrée. La paramétrisation des étapes de traitement classique influence en conséquence sur la résultat de la deuxième itération. Une étude peuvent mener à l'optimisation des paramètres de traitement classique. Elle permet de conserver la précision de la résultat obtenu avec le processus de deuxième itération en minimisant le temps de traitement classique.

Comme présenté précédemment au chapitre 4, les points *SIROP* sont extraits à partir d'une modélisation géométrique de la scène donné par la maillage 3D (mesh) et les poses des images. La méthode utilise ces informations pour prédire la localisation de correspondant et corriger la déformation perspective entre des zones d'appariement. L'algorithme implémenté accepte l'erreur de prédiction vient de l'imprécision de la résultat de traitement classique. Ceci est réglé par les seuils géométrique présenté dans la tableau 4.1. Un traitement classique avec des paramètres tolérants peut donc effectué en conservant la bon fonctionnalité et la précision de la résultat de la deuxième itération. Ces paramètres sont :

- la résolution d'extraction des points homologues,
- la résolution d'appariement dense,
- la résolution de génération de la maillage 3D,
- la type de méthode de génération de la maillage 3D.

Une autre approche potentielle est d'utilisation un modèle de calibration basique (peu de paramètres) au cours de traitement classique, puis exécuter le deuxième processus avec les points *SIROP* et un modèle de calibration plus sophistiqué. Cette approche peut accélérer la traitement.

6.3.3 Mixage SIFT/SIROP

Comme présenté dans la section 5.4.2, une limitation de la méthode d'extraction des points *SIROP* est le choix de la partie de la scène de chaque image maîtresse pour extraire des points (voir la section 4.2.1). L'algorithme va choisir les parties le plus résolues de la scène par rapport à l'orientation de chaque image maîtresse. Donc, avec des scènes longues et des images de GSD variable, la partie de la scène choisie pour extraire des points homologues ne couvre pas tout l'espace image (voir la figure 5.5).

Une combinaison entre les points *SIFT* extraits dans un traitement classique et les points *SIROP* dans la deuxième itération peut être utilisée afin de garder les points homologues dans les zones lointaines de la scène.

6.3.4 Modèle 3D théorique

La méthode *SIROP* peut également être utilisée pour extraire des points homologues entre des images acquises dans une scène dont le modèle théorique *a priori* existe. C'est généralement le cas pour des mesures de déformation dans l'industrie, avec un modèle théorique de type CAO², ou pour le suivi régulier en s'appuyant sur le maillage 3D de la campagne précédente.

2. Conception assistée par ordinateur

Annexe A

Modèle de caméra *Pinhole*/Sténopé

Le modèle de caméra *Pinhole* ou sténopé est un modèle de projection en perspective qui permet de calculer la projection d'un point dans l'espace 3D sur un plan image ($\mathbb{R}^3 \rightarrow \mathbb{R}^2$). Un ensemble de rayons lumineux traverse un petit trou sur la paroi de la chambre noire et crée une image inverse sur la surface sensible¹ (voir la figure A.1). La version sans distorsion de ce modèle suppose que le système optique² de la caméra respecte les conditions de Gauss³. L'objectif d'une caméra réelle contient un ensemble de lentilles complexes équivalent à une seule lentille dans le modèle de la caméra sténopé. Le centre de projection correspond au point *pinhole* où tous les rayons passent.

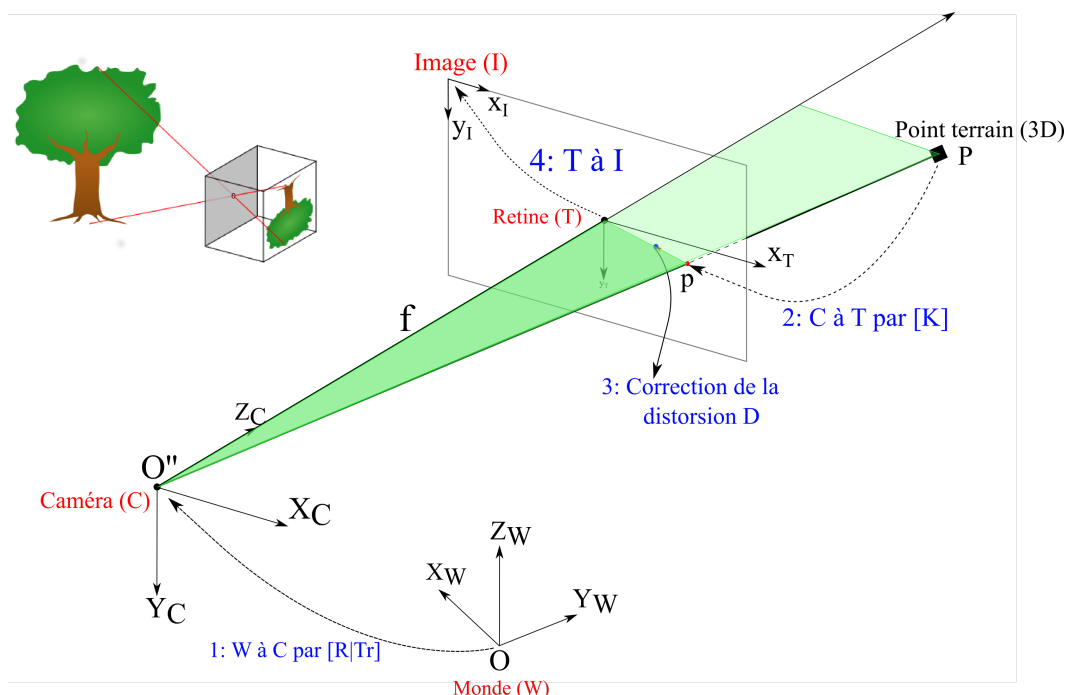


FIGURE A.1 – Gauche : le caméra sténopé. (Source : *Wikipedia*) – Droite : Modèle mathématique de la caméra sténopé.

Le modèle mathématique de la caméra sténopé et les termes utilisés sont présentés dans la figure A.1 :

1. plan image, plan focal, plan capteur
2. lentille, objectif
3. les rayons lumineux sont proches et peu inclinés par rapport à l'axe optique

- centre de projection : point où passent tous les rayons lumineux ;
- plan image : plan où l'image se forme dans la chambre noire ;
- axe optique : ligne droite perpendiculaire au plan image et passant par le centre de projection ;
- focale : distance entre le plan image et le centre de projection (f) ;
- point principal : point d'intersection entre l'axe optique et le plan image (PP).

Mathématiquement, on définit les repères suivants (voir la figure A.1) :

1. Repère du monde⁴ W : repère 3D métrique. On peut décrire les coordonnées d'un point 3D d'un objet indépendamment de la position de la caméra.
2. Repère caméra⁵ C : repère 3D métrique, relié à la caméra. L'origine du repère correspond au centre de projection de la caméra. L'orientation du repère suis celle de caméra aussi.
3. Repère rétine T : repère 2D métrique, relié au plan image. L'origine du repère correspond au point principal.
4. Repère image I : repère 2D en pixel, relié au plan image. L'origine conventionnelle du repère se situe en haut à gauche du plan image.

Pour passer d'un point du repère du monde W ($OXYZ$ dans la figure A.1) au repère image 2D I ($O''UV$ dans la figure A.1), on effectue trois passages :

- du repère du monde W au repère caméra C ;
- du repère caméra C au repère rétine T ;
- du repère rétine T au repère image I .

Ces trois passages sont présentés dans les parties suivantes.

A.1 Du repère du monde au repère caméra

Il s'agit ici d'une rotation, notée R , et d'une translation, notée Tr . La rotation R est représenté par une matrice 3×3 . La translation Tr est représentée par un vecteur 3D qui déplace l'origine du repère du monde W vers celle de la caméra C (les deux origines des repères notés O et O'' sont présentes dans la figure A.1).

Notons P_W est un point 3D avec les coordonnées (P_{xW}, P_{yW}, P_{zW}) dans le repère du monde W et (P_{xC}, P_{yC}, P_{zC}) dans le repère caméra C . On peut regrouper R et Tr en une seule matrice $[R|Tr]$, en utilisant les coordonnées homogènes :

$$P_C = [R|Tr]P_W = \begin{pmatrix} R_{00} & R_{01} & R_{02} & Tr_x \\ R_{10} & R_{11} & R_{12} & Tr_y \\ R_{20} & R_{21} & R_{22} & Tr_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} P_{xW} \\ P_{yW} \\ P_{zW} \\ 1 \end{pmatrix} = \begin{pmatrix} P_{xC} \\ P_{yC} \\ P_{zC} \end{pmatrix} \quad (\text{A.1})$$

La matrice $[R|Tr]$ encode les paramètres extrinsèques de la caméra.

4. espace objet, *world coordinates*

5. camera coordinates

A.2 Du repère caméra au repère image

Cette transformation transfère un point P_C du repère caméra 3D C au point p_T dans le repère image I . On peut exprimer cette transformation en deux parties :

- une transformation qui transfère un point du repère caméra C au repère rétine T ;
- une transformation qui transfère un point du repère rétine T au repère image I (unité en pixel).

Un point 3D P_C se transfère en point 2D p_T sur le plan à une distance f de l'origine du repère par des triangles similaires (voir les triangles dessinés en vert dans la figure A.1) :

$$(P_{xC}, P_{yC}, P_{zC}) \rightarrow \left(\frac{fP_{xC}}{P_{zC}}, \frac{fP_{yC}}{P_{zC}}, \frac{fP_{zC}}{P_{zC}} \right) = \left(\frac{fP_{xC}}{P_{zC}}, \frac{fP_{yC}}{P_{zC}}, f \right) \quad (\text{A.2})$$

La troisième coordonnée prend toujours la valeur f . On peut donc le négliger sans perdre de l'information :

$$(P_{xC}, P_{yC}, P_{zC}) \rightarrow \left(\frac{fP_{xC}}{P_{zC}}, \frac{fP_{yC}}{P_{zC}} \right) = (p_{xT}, p_{yT}) \quad (\text{A.3})$$

La transformation A.3 peut être représentée sous forme de matrices :

$$\begin{pmatrix} p_{xT} \\ p_{yT} \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} P_{xC} \\ P_{yC} \\ P_{zC} \\ 1 \end{pmatrix} \quad (\text{A.4})$$

Dans la plupart des algorithmes et des logiciels photogrammétriques, les coordonnées des images conventionnelles sont décrites en pixels. Une transformation, de mise à l'échelle est donc effectuée pour passer des coordonnées métriques p_T aux coordonnées pixels p_I . L'échelle peut se calculer à partir de la taille du photo-site ou de la taille du capteur (i.e nombre de pixels par millimètre).

Un autre problème est la convention de l'origine du plan image, souvent située en haut à gauche de l'image (repère image I). Le point x_T est actuellement exprimé dans le repère rétine T avec l'origine située au niveau du point principal (PP_x, PP_y). D'où une translation est ajoutée pour passer du repère T au repère I .

Notons k_u et k_v des facteurs d'échelle respectivement en ligne et colonne du capteur (i.e nombre de pixels par millimètre) et K la matrice de transformation. On obtient alors :

$$K = \begin{pmatrix} fk_u & 0 & PP_x & 0 \\ 0 & fk_v & PP_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (\text{A.5})$$

Un paramètre de *skew* s est ajouté pour modéliser le cas où le plan image n'est pas perpendiculaire à l'axe optique :

$$K = \begin{pmatrix} fk_u & s & PP_x & 0 \\ 0 & fk_v & PP_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (\text{A.6})$$

La matrice K encode les paramètres intrinsèques de la caméra. K s'appelle la matrice de calibration sans distorsion. Ces paramètres dépendent seulement de la géométrie interne de la caméra. Ils sont indépendants de l'orientation ou de la position de la caméra dans l'espace.

A.3 La transformation finale

Une combinaison entre deux transformations $[R|Tr]$ (voir l'équation A.1) et K (voir l'équation A.6) nous donne la matrice de projection π . Cette transformation de projection transfère un point 3D du repère du monde W en coordonnées pixels sur l'image :

$$\pi = K[R|Tr] \quad (\text{A.7})$$

Annexe B

Modèle de distorsion

Un petit trou (*pinhole*) sur la paroi de la chambre noire de la caméra sténopé produit une image nette de la scène sur les plans parallèles à la paroi à n'importe quelle distance (focale). Par contre, ce petit trou ne laisse pas passer une grande quantité de lumière. Plus le *pinhole* est agrandi, plus l'image rendue est floue. Une lentille convexe est alors placée devant le *pinhole* afin de rendre l'image plus nette sur un plan focal fixé. Les objectifs modernes contiennent plusieurs éléments de lentille afin de faire la mise au point, de corriger les effets d'aberrations etc. Par contre, l'objectif cause des effets de distorsions sur l'image rendue.

Le modèle de la caméra sténopé avec sa matrice de projection A.7 transforme un point 3D du repère monde W en point 2D du repère image I . Pour pouvoir appliquer le modèle de la caméra sténopé avec un système optique réel, il faut prendre en compte et corriger la distorsion. Ce phénomène est souvent modélisé par un modèle radial afin de corriger la partie la plus importante de la distorsion. Le modèle de distorsion radiale [McGlone, 2013] est justifié par la symétrie cylindrique [Weng et al., 1992] du système de lentilles (surfaces sphériques dont les centres sont alignés sur un même axe). La distorsion tangentielle [McGlone, 2013] est causée par le décentrement des lentilles par rapport à l'axe optique [Drap and Lefèvre, 2016; Salvi et al., 2002; Swaminathan and Nayar, 2000; Weng et al., 1992]. Dans les objectifs récents, cette distorsion est négligable.

Cette partie décrit le processus de calibration de la caméra utilisé dans nos travaux expérimentaux. On se limite ici à présenter les modèles de distorsions utilisés dans nos expérimentations : le modèle de distorsion radiale [Brown, 1966; Duane, 1971], le modèle Fraser [Fraser, 1997] et le modèle fisheye.

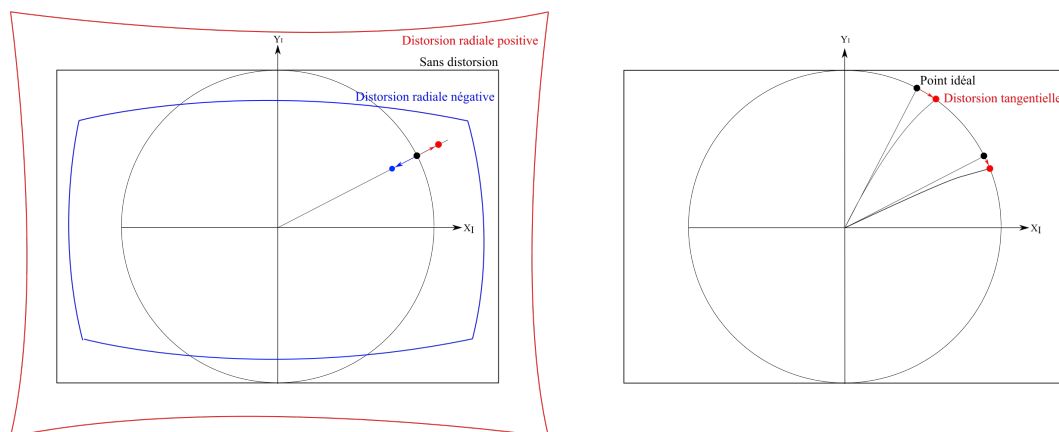


FIGURE B.1 – La distorsion radiale et distorsion tangentielle.

B.1 Distorsion radiale

Cette distorsion cause un déplacement radial des images des objets par rapport au centre de distorsion. La distance de déplacement varie en fonction de la distance radiale du pixel. Empiriquement, plus l'angle d'ouverture est grand, plus la distorsion est importante, car il devient de plus en plus complexe de concevoir des objectifs sans distorsion. Il existe deux types de distorsions radiales (voir la figure B.1) : la distorsion radiale négative¹ et la distorsion radiale positive².

Si les lentilles dans un objectif sont alignées et centrées sur l'axe optique, le centre de distorsion (C_x, C_y) peut être considéré comme identique au point principal (PP_x, PP_y) . Cette distorsion est modélisée par un polynôme de degré 2,4,6 ou plus. La correction d'effet radial est apportée principalement par les premiers termes du polynôme.

Mathématiquement, dans la repère image I (voir la figure A.1), notons les coordonnées pixel (x_{u_I}, y_{u_I}) du point corrigé (sans distorsion) et (x_{d_I}, y_{d_I}) ceux obtenu initialement sur image (point avec la distorsion). (δ_x, δ_y) sont les corrections de distorsion selon x et y . On a :

$$\begin{cases} x_{d_I} = x_{u_I} + \delta_x \\ y_{d_I} = y_{u_I} + \delta_y \end{cases} \quad (\text{B.1})$$

où

$$\begin{cases} \delta_x(r) = x_{u_I}(k_1 r^2 + k_2 r^4 + k_3 r^6) \\ \delta_y(r) = y_{u_I}(k_1 r^2 + k_2 r^4 + k_3 r^6) \end{cases} \quad (\text{B.2})$$

$$\text{avec } r = \sqrt{(x_{u_I} - C_x)^2 + (y_{u_I} - C_y)^2} \quad (\text{B.3})$$

En suite, la distorsion radiale en repère image I avec l'origine à la centre de distorsion (C_x, C_y) :

$$\begin{cases} x_{d_I} - C_x = (x_{u_I} - C_x) + (x_{u_I} - C_x)(k_1 r^2 + k_2 r^4 + k_3 r^6) \\ y_{d_I} - C_y = (y_{u_I} - C_y) + (y_{u_I} - C_y)(k_1 r^2 + k_2 r^4 + k_3 r^6) \end{cases} \quad (\text{B.4})$$

$$\Rightarrow \begin{cases} x_{d_I} = C_x + d_x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ y_{d_I} = C_y + d_y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \end{cases} \quad (\text{B.5})$$

$$\text{avec } d_x = x_{u_I} - C_x \quad ; \quad d_y = y_{u_I} - C_y \quad (\text{B.6})$$

$$r = \sqrt{d_x^2 + d_y^2} \quad (\text{B.7})$$

Le fonction B.5 nous donne les coordonnées d'un point 3D projeté sur l'image en pris en compte la distorsion. Il est le fonction directe dans le sens du repère du monde au repère image. Les inverses de ces fonctions sont les fonctions de correction de la distorsion. Ces fonctions inverses peuvent être utilisées pour corriger une image distordu.

1. *pincushion*
2. *barrel*

B.2 Distorsion Fraser

Cette partie présente le modèle de distorsion du Fraser [Fraser, 1997]. Il compose la correction :

- radiale ;
- tangentielle ;
- affine³.

En expression mathématique, on a :

$$\begin{cases} \delta_x = \delta_{xr} + \delta_{xt} + \delta_{xa} \\ \delta_y = \delta_{yr} + \delta_{yt} + \delta_{ya} \end{cases} \quad (\text{B.8})$$

avec $(\delta_{xr}, \delta_{xt}, \delta_{xa})$ sont les corrections de distorsion radiale, tangentielle et affine, respectivement. Idem pour y . La distorsion tangentielle est modélisée comme ci-dessous :

$$\begin{cases} \delta_{xt} = p_1(3x_{u_I}^2 + y_{u_I}^2) + 2p_2x_{u_I}y_{u_I} \\ \delta_{yt} = p_2(3y_{u_I}^2 + x_{u_I}^2) + 2p_1x_{u_I}y_{u_I} \end{cases} \quad (\text{B.9})$$

La correction affine est modélisée par :

$$\begin{cases} \delta_{xa} = b_1x_{u_I} + b_2y_{u_I} \\ \delta_{ya} = 0 \end{cases} \quad (\text{B.10})$$

Finalement, en groupant tout les trois termes de correction de distorsions décrit par les équations B.5, B.9, B.10, on a le modèle de distorsion Fraser final :

$$\begin{cases} x_{d_I} = x_{u_I} + x_{u_I}(k_1r^2 + k_2r^4 + k_3r^6) + p_1(3x_{u_I}^2 + y_{u_I}^2) + 2p_2x_{u_I}y_{u_I} + b_1x_{u_I} + b_2y_{u_I} \\ y_{d_I} = y_{u_I} + y_{u_I}(k_1r^2 + k_2r^4 + k_3r^6) + p_2(3y_{u_I}^2 + x_{u_I}^2) + 2p_1x_{u_I}y_{u_I} \end{cases} \quad (\text{B.11})$$

$$r = \sqrt{(x_{u_I} - C_x)^2 + (y_{u_I} - C_y)^2} \quad (\text{B.12})$$

B.3 L'objectif Fisheye

L'objectif Fisheye accepte les rayons de lumière avec un angle d'entrée (angle entre le rayon et l'axe optique/angle d'incidence) θ de 90° (angle de champ de 180°) ou même plus. Le modèle de caméra sténopé avec la fonction de représentation classique (rectilinéaire) $r = f \tan(\theta)$ va élargir le plan image à l'infini ou réduire la focale à 0. (r est la distance radiale d'un point d'image (x_{d_I}, y_{d_I}) par rapport au centre de distorsion (C_x, C_y)).

Pour pouvoir modéliser une caméra avec l'objectif Fisheye, une fonction de représentation adaptée avec la construction de l'objectif doit être utilisée. La fonction de représentation est une fonction qui associe à un angle d'entrée d'un point objet sa position sur le

3. *in-plane distorsion*

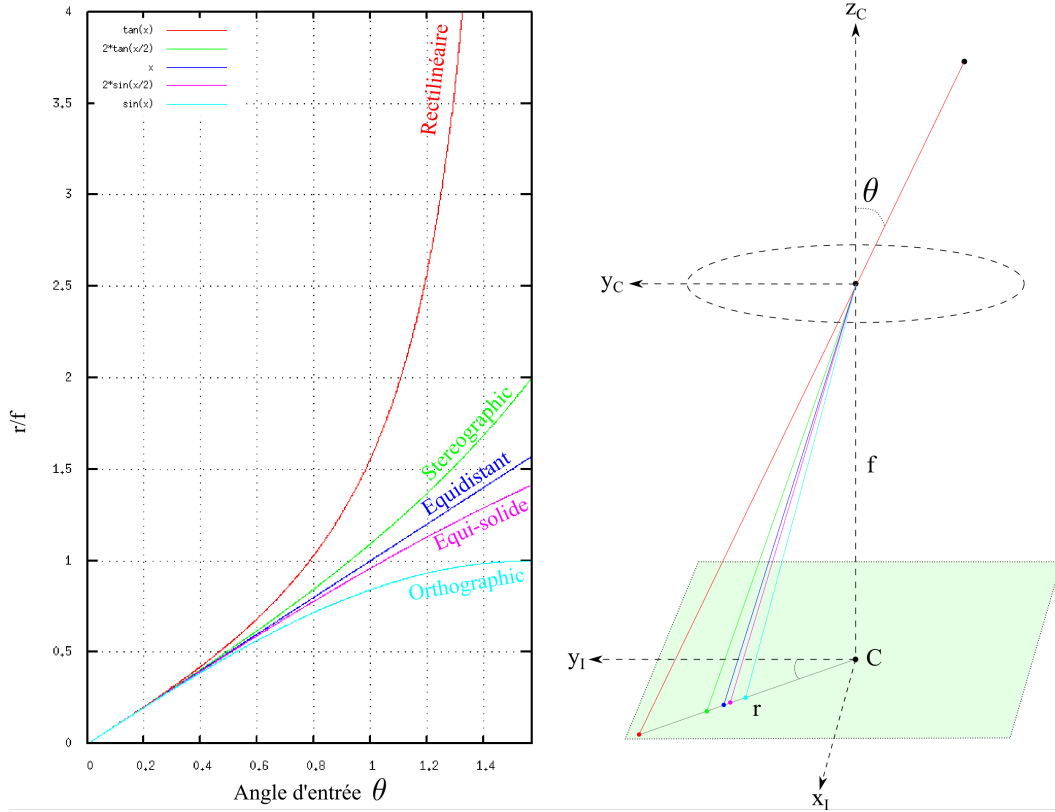


FIGURE B.2 – Différentes fonctions radiales de correspondance entre l'angle d'entrée θ et la position sur le plan image r

plan image (voir la figure B.2). Il existe plusieurs fonctions de représentation pour le Fisheye comme équidistante (équilinéaire), équisolide [Abraham and Förstner, 2005; Schneider et al., 2009] ou stéréographique pour quelques objectifs comme le Samyang 8mm [Marcato Junior et al., 2015]. Parmi eux, les deux plus populaires sont :

- équidistante : $r = f\theta$
- équisolide : $r = f \sin(\theta/2)$.

Dans le cas présenté ci-dessous, le modèle équidistant $r = f\theta$ va être utilisé. Un point 3D (X_C, Y_C, Z_C) dans le repère de caméra C peut s'exprimer en coordonnées sphériques : $(R \sin \theta \cos \phi, R \sin \theta \sin \phi, R \cos \theta)$. Un point 2D (x_{d_I}, y_{d_I}) dans la repère d'image I peu s'exprimer en coordonnées polaires : $(R \cos \omega, R \sin \omega)$ (voir la figure B.3 pour les notions mathématiques). f est la focale d'objectif. On a :

$$r(\theta) = f \cdot \theta = f \cdot \Phi(\theta) \quad (\text{B.13})$$

$$\Rightarrow R \begin{pmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{pmatrix} \rightarrow f \cdot \Phi(\theta) \begin{pmatrix} \sin \omega \\ \cos \omega \end{pmatrix} \quad (\text{B.14})$$

En développant la partie droite de l'équation B.14 en coordonnées cartésiennes

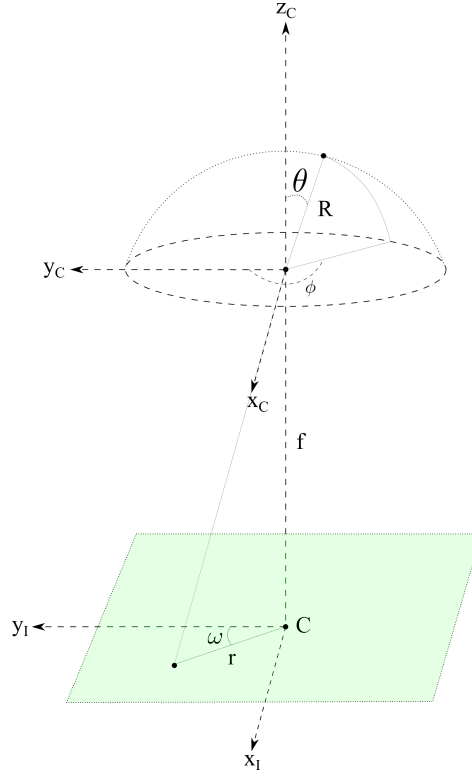


FIGURE B.3 – Modèle de caméra avec l'objectif Fisheye

(repère image I), on a :

$$f \cdot \Phi(\theta) \begin{pmatrix} \sin \omega \\ \cos \omega \end{pmatrix} = \frac{f}{r} \cdot \Phi(\theta) \begin{pmatrix} x_{d_I} - C_x \\ y_{d_I} - C_y \end{pmatrix} = \frac{f}{r} \cdot \theta \begin{pmatrix} x_{d_I} - C_x \\ y_{d_I} - C_y \end{pmatrix} \quad (\text{B.15})$$

$$= \frac{f}{r} \cdot \arctan\left(\frac{r}{f}\right) \begin{pmatrix} x_{d_I} - C_x \\ y_{d_I} - C_y \end{pmatrix} = \frac{f}{r} \cdot \Phi\left(\arctan\left(\frac{r}{f}\right)\right) \begin{pmatrix} x_{d_I} - C_x \\ y_{d_I} - C_y \end{pmatrix} \quad (\text{B.16})$$

$$r = \sqrt{(x_{d_I} - C_x)^2 + (y_{d_I} - C_y)^2}$$

Le modèle d'approximation de l'objectif Fisheye peut s'écrire de la manière suivante :

$$\begin{pmatrix} x_{d_I} \\ y_{d_I} \end{pmatrix} = \frac{f}{r} \cdot \Phi\left(\arctan\left(\frac{r}{f}\right)\right) \begin{pmatrix} x_{d_I} - C_x \\ y_{d_I} - C_y \end{pmatrix} \quad (\text{B.17})$$

On ajoute la partie de modélisation de la distorsion. Notons (C_x, C_y) sont les coordonnées de la centre de distorsion, on pose :

$$A = \tan(\phi_x) = \frac{x_{d_I} - C_x}{f} \quad B = \tan(\phi_y) = \frac{y_{d_I} - C_y}{f} \quad (\text{B.18})$$

$$r_N = \sqrt{A^2 + B^2} \quad (\text{B.19})$$

Notons :

$$\phi = \lambda r_N \Rightarrow \lambda = \frac{\phi}{r_N} = \frac{\arctan\left(\frac{R}{f}\right)}{r_N} = \frac{\arctan(r_N)}{r_N} \quad (\text{B.20})$$

$$\Rightarrow \begin{cases} a = \phi_x = \lambda A \\ b = \phi_y = \lambda B \\ \rho = \sqrt{a^2 + b^2} \\ \lambda = \frac{\arctan(r_N)}{r_N} \end{cases} \quad (\text{B.21})$$

Le modèle de distorsion D_n selon les coordonnées normalisées (a, b) , y compris la partie radiale, décentralisée et affine peut s'écrire :

$$D_n \begin{pmatrix} a \\ b \end{pmatrix} = (1 + k_1 \rho^2 + k_2 \rho^4 \dots) \begin{pmatrix} a \\ b \end{pmatrix} + p_1 \begin{pmatrix} 2a^2 + \rho^2 \\ 2ab \end{pmatrix} + p_2 \begin{pmatrix} 2ab \\ 2b^2 \rho^2 \end{pmatrix} + l_1 \begin{pmatrix} a \\ 0 \end{pmatrix} + l_2 \begin{pmatrix} b \\ a \end{pmatrix} \quad (\text{B.22})$$

Après dénormalisation de l'équation B.22, on obtient le modèle de distorsion dans le repère image I :

$$\begin{pmatrix} x_{d_I} \\ y_{d_I} \end{pmatrix} = f \cdot D_n \begin{pmatrix} a \\ b \end{pmatrix} \quad (\text{B.23})$$

avec (a, b) calculé par l'équation B.21.

Le détail sur les autres modèles mathématiques de distorsion est présenté dans la thèse de Vincent Tournard [Tournadre, 2015] et dans la documentation du logiciel MicMac [Deseilligny et al., 2018].

Annexe C

Auto-calibration

L'auto-calibration est un processus qui permet de déterminer les paramètres internes de la caméra directement à partir de plusieurs images non calibrées prises dans une scène non structurée. Contrairement à la méthode classique de calibration, l'auto-calibration n'a pas besoin de connaître *a priori* la structure de la scène (i.e un objet calibré ou des coordonnées connues). Les paramètres internes sont résolus à partir des points homologues mesurés sur des images. L'image est la projection obtenue par le modèle de caméra sténopé (A.7) avec un modèle de distorsion non-linéaire (B). Le principe du processus d'auto-calibration est présenté dans [Faugeras et al., 1992].

Dans nos travaux expérimentaux, l'auto-calibration est effectuée par le processus d'optimisation non-linéaire. Les paramètres internes sont obtenus par les itérations avec la contrainte de minimisation d'une fonction déterminée. La fonction de minimisation correspond à l'erreur de projection des points homologues (voir la section D.1). Cette stratégie demande une première initialisation assez proche de la solution finale pour éviter le risque de divergence.

Le processus d'auto-calibration que nous avons utilisé peut se décrire en trois étapes :

- initialisation des paramètres internes :
 - focale (à partir de méta-données) ;
 - position du point principal (au milieu d'image) ;
 - coefficients de distorsion (nuls, sauf pour l'objectif Fisheye où un modèle simple est utilisé) ;
 - centre de distorsion (identique à la position du point principal) ;
- Initialisation des paramètres externes :
 - par les méthodes d'estimation de pose ;
 - en fixant les paramètres internes ;
 - effectuer l'ajustement des faisceaux régulièrement.
- Après initialisation des poses des images, faire l'ajustement des faisceaux en libérant les paramètres internes dans l'ordre :
 - deux premiers coefficients de distorsion ;
 - focale ;
 - point principal ;
 - troisième coefficient de distorsion ;
 - centre de distorsion ;

- coefficients décentralisés ;
- coefficients de distorsion de degré le plus haut un par un ;
- polynômes supplémentaires.

Le choix de cet ordre est à la fois arbitraire et sensible ; il faut éviter de libérer trop de paramètres simultanément (risque de divergence) ou dans le "mauvais" ordre (par exemple libérer le centre de distorsion quand la distorsion radiale n'est pas encore assez bien définie) ; bien sûr, beaucoup d'autres ordres sont possibles.

Le processus d'auto-calibration utilisé dans nos travaux expérimentaux est effectué avec l'outil *Tapas* de la chaîne MicMac (voir la documentation [Deseilligny et al., 2018]). L'ordre de libération des paramètres internes peut être modifiée suivant les besoins de l'utilisateur.



FIGURE C.1 – La correction de distorsion d'une image prise par l'objectif Fisheye. **Gauche** : les vecteurs du résidu de calibration, **Milieu** : image originale, **Droite** : image corrigée.

Annexe D

Contraintes photogrammétriques

Cette partie décrit deux contraintes qui sont utilisées dans l’ajustement des faisceaux de nos travaux expérimentaux. Il s’agit de :

- la contrainte de projection des points homologues, cette contrainte est toujours présente et c’est souvent la seule ;
- la contrainte de rigidité inter-caméras.

La contrainte de projection des points homologues est une contrainte interne des images classiques dans la photogrammétrie. Elle utilise des points homologues mesurés sur des images comme observations. Cette contrainte est utilisée dans l’ajustement des faisceaux de tous les jeux de données présentés dans le chapitre 5.

La contrainte de rigidité inter-caméras est une contrainte interne au système de prise de vue. Elle est utilisée dans le cas d’acquisitions faites avec un système de caméra multi-têtes, avec une structure connue *a priori*. La structure du système de caméra est l’orientation relative entre les têtes de la caméra dans le système.

D.1 Contrainte de projection des points homologues

Un point homologue i mesuré sur l’ensemble des images $(I_1, \dots, I_k, \dots, I_n)$, a pour coordonnées $p_{(i,k)}$ sur l’image k . Les coordonnées 3D (monde) de ce point sont P_i . La matrice de projection π_k (voir l’équation A.7, section A) projette un point 3D, noté P_i , sur le plan image k , sans prendre en compte la distorsion :

$$p_{(i,k)}^0 = \pi_k(P_i) \tag{D.1}$$

En prenant en compte la distorsion d’optique D (voir les modèles de distorsions B.5, B.11, B.22 dans la section B), on a les coordonnées de la projection d’un point 3D P_i sur plan image k :

$$\hat{p}_{(i,k)} = D(\pi_k(P_i)) \tag{D.2}$$

On cherche à minimiser l’écart entre les coordonnées de la projection $\hat{p}_{(i,k)}$ et les coordonnées mesurées $p_{(i,k)}$:

$$E_{i,k}(\sigma_{pix}) = \sum_{i,k} \left(\frac{P(i,k) - \hat{P}(i,k)}{\sigma_{pix}} \right)^2 \quad (D.3)$$

$$= \sum_{i,k} \left(\frac{P(i,k) - D(\pi_k(P_i))}{\sigma_{pix}} \right)^2 \quad (D.4)$$

L'équation D.4 est la contrainte de projection des points homologues.

D.2 Contrainte de rigidité inter-caméras

Le système de caméra multi-têtes contient les caméras nommées (a, b, c, \dots) . L'acquisition est faite avec ce système qui contient plusieurs blocs d'images. Un bloc d'images est un ensemble d'images pris par un déclenchement de toutes les caméras synchronisées à certain moment k .

On note $P_{a,k}$ la pose d'images prises par la caméra a au moment k (ou bloc d'images k). Avec l'hypothèse que la structure du système est rigide mécaniquement, entre deux déclenchements, les positions relatives entre ces caméras (a, b, \dots) sont fixées :

$$\begin{aligned} P_{a,k}^{-1} P_{b,k} &= P_{a,k'}^{-1} P_{b,k'} \\ \text{avec } P_{a,k} &= (C_{a,k}, R_{a,k}) \end{aligned} \quad (D.5)$$

où $C_{a,k}$ et $R_{a,k}$ est le centre et la rotation de la pose $P_{a,k}$. $(P_{a,k}, P_{b,k'})$ s'expriment dans un même système de coordonnées arbitraire.

À partir de l'équation D.5, on peut exprimer la différences entre deux caméras (a, b) dans deux blocs d'images (k, k') en :

— rotation :

$$R_{a,k}^{-1} R_{b,k} - R_{a,k'}^{-1} R_{b,k'} = \Delta_{\omega}(a, b, k, k') \quad (D.6)$$

— translation :

$$(-C_{a,k} + R_{a,k}^{-1} C_{b,k}) - (-C_{a,k'} + R_{a,k'}^{-1} C_{b,k'}) = \Delta_{Tr}(a, b, k, k') \quad (D.7)$$

où Δ_{ω} et Δ_{Tr} sont les différences de rotation et de translation.

Dans le cas où le système de caméra est calibré, notons (Q_a, Q_b, \dots) les poses calibrées de la caméra (a, b, \dots) , qui s'expriment dans un même système de coordonnées arbitraire. Ce système de coordonnées a la même échelle que le système de coordonnées $(P_{a,k}, P_{b,k'})$.

Les équations D.6, D.7 peuvent s'écrire sous la forme :

$$\begin{aligned} R_{a,k}^{-1} R_{b,k} - r_a^{-1} r_b &= \Delta_{\omega}^g(a, b, k) \\ (-C_{a,k} + R_{a,k}^{-1} C_{b,k}) - (-c_a + r_a^{-1} c_b) &= \Delta_{Tr}^g(a, b, k) \end{aligned} \quad (D.8)$$

où $Q_a = (c_a, r_a)$ et $(\Delta_{\omega}^g(a, b, k), \Delta_{Tr}^g(a, b, k))$ sont les différences de rotation et de translation entre deux caméra (a, b) dans un bloc d'images k et dans la calibration.

Pour chaque bloc d'image k , la contrainte de rigidité inter-caméras est exprimée sous forme de minimisation de deux termes $(\Delta_{\omega}^g(a, b, k), \Delta_{Tr}^g(a, b, k))$ dans l'ajustement des faisceaux. L'erreur à minimiser est exprimée dans l'équation D.9 :

$$E^g(\sigma_{Tr}^g, \sigma_\omega^g) = \sum_{k,b} \left(\frac{\Delta_{Tr}^g(a, b, k)}{\sigma_{Tr}^g} \right)^2 + \left(\frac{\Delta_\omega^g(a, b, k)}{\sigma_\omega^g} \right)^2 \quad (\text{D.9})$$

où $(\sigma_{Tr}^g, \sigma_\omega^g)$ sont des variances.

On peut ajouter une contrainte qui exprime l'attachement de la calibration du système $(Q_a, Q_b \dots)$ avec des paramètres à estimer :

$$E^0(\sigma_{Tr}^0, \sigma_\omega^0) = \sum_k \left(\frac{c_b - C_{b,k}}{\sigma_{Tr}^0} \right)^2 + \left(\frac{r_b - R_{b,k}}{\sigma_\omega^0} \right)^2 \quad (\text{D.10})$$

L'équation [D.9](#) ou [D.10](#) présentent la contrainte de rigidité inter-caméras.

D.3 Références

- Abdel-Hakim, A. E., Farag, A. A., 2006. Csift : A sift descriptor with color invariant characteristics. In : *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. Vol. 2. IEEE, pp. 1978–1983.
- Abraham, S., Förstner, W., 2005. Fish-eye-stereo calibration and epipolar rectification. *ISPRS Journal of photogrammetry and remote sensing* 59 (5), 278–288.
- Ackermann, F., 1984. Digital image correlation : performance and potential application in photogrammetry. *The Photogrammetric Record* 11 (64), 429–439.
- Afsharnia, H., Azizi, A., Arefi, H., 2015. Accuracy improvement by the least squares image matching evaluated on the cartosat-1. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40 (1), 11.
- Aguilar, M., Pozo, J., Aguilar, F., García, A., Fernández, I., Negreiros, J., Sánchez-Hermosilla, J., 2010. Application of close-range photogrammetry and digital photography analysis for the estimation of leaf area index in a greenhouse tomato culture. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Part 5).
- Alcantarilla, P. F., Bartoli, A., Davison, A. J., 2012. Kaze features. In : *European Conference on Computer Vision*. Springer, pp. 214–227.
- Aron, M., Simon, G., Berger, M.-O., 2007. Use of inertial sensors to support video tracking. *Computer Animation and Virtual Worlds* 18 (1), 57–68.
- Arya, S., Mount, D. M., Netanyahu, N. S., Silverman, R., Wu, A. Y., 1998. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM (JACM)* 45 (6), 891–923.
- Aschwanden, P., Guggenbuhl, W., 1992. Experimental results from a comparative study on correlation-type registration algorithms. *Robust computer vision*, 268–289.
- Atkinson, K. B., 1996. *Close range photogrammetry and machine vision*. Whittles Publ.
- Audi, A., Pierrot-Deseilligny, M., Meynard, C., Thom, C., 2017. Implementation of an imu aided image stacking algorithm in a digital camera for unmanned aerial vehicles. *Sensors* 17 (7), 1646.
- Baltsavias, E. P., Stallmann, D., 1993. Spot stereo matching for digital terrain model generation. In : *2nd Swiss Symposium” Pattern Recognition and Computer Vision”*. Swiss Federal Institute of Technology, Institute of Geodesy and Photogrammetry.
- Barazzetti, L., Remondino, F., Scaioni, M., 2010. Extraction of accurate tie points for automated pose estimation of close-range blocks. In : *ISPRS Technical Commission III Symposium on Photogrammetric Computer Vision and Image Analysis*.
- Battiato, S., Gallo, G., Puglisi, G., Scellato, S., 2007. Sift features tracking for video stabilization. In : *Image Analysis and Processing, 2007. ICIAP 2007. 14th International Conference on*. IEEE, pp. 825–830.

- Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-up robust features (surf). *Computer vision and image understanding* 110 (3), 346–359.
- Beis, J. S., Lowe, D. G., 1997. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In : *cvpr. IEEE*, p. 1000.
- Bentley, J. L., 1975. Multidimensional binary search trees used for associative searching. *Communications of the ACM* 18 (9), 509–517.
- Bentrah, O., 2006. Reconstruction de la géométrie d’acquisition de séquences d’images acquises par un véhicule en milieu urbain. Ph.D. thesis, thèse de doctorat dirigée par Horaud, Radu et Pierrot-Deseilligny, Marc Imagerie, vision, robotique Grenoble INPG 2006.
URL <http://www.theses.fr/2006INPG0166>
- Berrani, S.-A., Feb. 2004. Approximate search of nearest neighbors with probabilistic control of the accuracy; application to content-based image retrieval. Theses, Université Rennes 1.
URL <https://tel.archives-ouvertes.fr/tel-00532854> :tel-00532854
- Bethmann, F., Luhmann, T., 2011. Least-squares matching with advanced geometric transformation models. *Photogrammetrie-Fernerkundung-Geoinformation* 2011 (2), 57–69.
- Briechele, K., Hanebeck, U. D., 2001. Template matching using fast normalized cross correlation. In : *Optical Pattern Recognition XII. Vol. 4387. International Society for Optics and Photonics*, pp. 95–103.
- Brown, D. C., 1966. Decentering distortion of lenses. *Photogrammetric Engineering and Remote Sensing*.
- Burgos, S., Mota, M., Noll, D., Cannelle, B., 2015. Use of very high-resolution airborne images to analyse 3d canopy architecture of a vineyard. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40 (3), 399.
- Calantropio, A., Deseilligny, M. P., Rinaudo, F., Rupnik, E., 2018. Evaluation of photogrammetric block orientation using quality descriptors from statistically filtered tie points. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 42 (2).
- Cannelle, B., 2013. Estimation de pose de grands blocs d’images panoramiques issues de systèmes de numérisation mobile. Ph.D. thesis, Université Paris-Est.
- Canny, J., 1986. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence* (6), 679–698.
- Chambon, S., Crouzil, A., 2003. Dense matching using correlation : new measures that are robust near occlusions. In : *BMVC. Vol. 1. pp. 143–152*.
- Cheng-Yi, C., Shih-Hong, C., 2012. Improved sift algorithm to match points in the texture region image. *The 33rd Asian Conference On Remote Sensing ACRS*.

- Daakir, M., Pierrot-Deseilligny, M., Bossler, P., Pichard, F., Thom, C., Rabot, Y., Martin, O., 2017. Lightweight uav with on-board photogrammetry and single-frequency gps positioning for metrology applications. *ISPRS Journal of Photogrammetry and Remote Sensing* 127, 115–126.
- Delevoie, C., Dutailly, B., Mora, P., Vergnieux, R., 2012. Un point sur la photogrammétrie. *Archéopages. Archéologie et société* (34), 86–89.
URL <https://journals.openedition.org/archeopages/410>
- Deseilligny, M. P., Clery, I., 2011. Apero, an open source bundle adjustment software for automatic calibration and orientation of set of images 269277.
- Deseilligny, M. P., Rupnik, E., M, D., Nguyen, T., et al., 2018. Micmac, apero, pastis and other beverages in a nutshell. *MicMac Documentation* 4.
- Drap, P., Lefèvre, J., 2016. An exact formula for calculating inverse radial lens distortions. *Sensors* 16 (6), 807.
- Duane, C. B., 1971. Close-range camera calibration. *Photogramm. Eng* 37 (8), 855–866.
- Dwarakanath, D., Eichhorn, A., Halvorsen, P., Griwodz, C., 2012. Evaluating performance of feature extraction methods for practical 3d imaging systems. In : *Proceedings of the 27th Conference on Image and Vision Computing New Zealand*. ACM, pp. 250–255.
- Ehlers, M., 1985. The effects of image noise on digital correlation probability. *Photogrammetric Engineering and Remote Sensing* 51, 357–365.
- Eltner, A., Kaiser, A., Castillo, C., Rock, G., Neugirg, F., Abellán, A., 2016. Image-based surface reconstruction in geomorphometry—merits, limits and developments. *Earth Surface Dynamics* 4 (2), 359–389.
- Eltner, A., Schneider, D., 2015. Analysis of different methods for 3d reconstruction of natural surfaces from parallel-axes uav images. *The Photogrammetric Record* 30 (151), 279–299.
- Faugeras, O. D., Luong, Q.-T., Maybank, S. J., 1992. Camera self-calibration : Theory and experiments. In : *European conference on computer vision*. Springer, pp. 321–334.
- Fayyad, T. M., Lees, J. M., 2014. Application of digital image correlation to reinforced concrete fracture. *Procedia Materials Science* 3, 1585–1590.
- Förstner, W., 1982. On the geometric precision of digital correlation. *Int. Arch. Photogrammetry & Remote Sensing* 24 (3), 176–189.
- Förstner, W., 1986. A feature based correspondence algorithm for image matching. *ISPRS ComIII, Rovaniemi*, 150–166.
- Fraser, C. S., 1997. Digital camera self-calibration. *ISPRS Journal of Photogrammetry and Remote sensing* 52 (4), 149–159.
- Friedman, J. H., Bentley, J. L., Finkel, R. A., 1977. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software (TOMS)* 3 (3), 209–226.

- Furukawa, Y., Ponce, J., 2010. Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence* 32 (8), 1362–1376.
- Gaiani, M., Remondino, F., Apollonio, F. I., Ballabeni, A., 2016. An advanced pre-processing pipeline to improve automated photogrammetric reconstructions of architectural scenes. *Remote Sensing* 8 (3).
URL <http://www.mdpi.com/2072-4292/8/3/178>
- Gervais, Y. G. F., 2017. From off-the-shelf to off-the-charts : shifting the camera paradigm in commercial fixed wing uas to increase accuracy. *UAV Metric Camera Special Sessions DGPF Annual Meeting 2017 Würzburg, Germany*.
- Górski, F., Kuczko, W., Wichniarek, R., Zawadzki, P., 2010. Application of close range photogrammetry in reverse engineering. In : *Proceedings of the 7th International Conference of DAAAM Baltic, Industrial Engineering*.
- Gruen, A., 1985. Adaptive least squares correlation : a powerful image matching technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography* 14 (3), 175–187.
- Gruen, A., 1996. Least squares matching : a fundamental measurement algorithm. *Close Range Photogrammetry and Machine Vision*.
- Gruen, A., 2012. Development and status of image matching in photogrammetry. *The Photogrammetric Record* 27 (137), 36–57.
- Gruen, A., Baltsavias, E. P., 1988. Geometrically constrained multiphoto matching. *Photogrammetric engineering and remote sensing* 54 (5), 633–641.
- Gruen, A., Stallmann, D., 1991. High-accuracy edge-matching with an extension of the mpgc-matching algorithm. In : *Industrial Vision Metrology*. Vol. 1526. International Society for Optics and Photonics, pp. 42–56.
- Gruen, A. W., Baltsavias, E. P., 1986. Adaptive least squares correlation with geometrical constraints. In : *Computer Vision for Robots*. Vol. 595. International Society for Optics and Photonics, pp. 72–83.
- Grün, A., 1978. Progress in photogrammetric point determination by compensation of systematic errors and detection of gross errors. *Inform. Relative to Cartography and Geodesy. Ser. 2 : Transl.* p 113-140(SEE N 80-12455 03-43).
- Haala, N., 2014. Dense image matching final report. *EuroSDR Publication Series, Official Publication* 64, 115–145.
- Hanke, K., Grussenmeyer, P., 2002. Architectural photogrammetry : basic theory, procedures, tools. In : *ISPRS Commission*. Vol. 5. pp. 1–2.
- Harris, C., Stephens, M., 1988. A combined corner and edge detector. In : *Alvey vision conference*. Vol. 15. Citeseer, pp. 10–5244.
- Heitger, F., Von Der Heydt, R., Peterhans, E., Rosenthaler, L., Kübler, O., 1998. Simulation of neural contour mechanisms : representing anomalous contours. *Image and Vision Computing* 16 (6-7), 407–421.

- Helava, U., 1978. Digital correlation in photogrammetric instruments. *Photogrammetria* 34 (1), 19–41.
- Hobrough, G., 1959. Automatic stereo plotting. *Photogrammetric Engineering* 25 (5), 763–769.
- Honkavaara, E., Jaakkola, J., 1998. On properties of automatically measured tie point observations. *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY AND REMOTE SENSING* 32, 42–51.
- James, M., Robson, S., 2012. Straightforward reconstruction of 3d surfaces and topography with a camera : Accuracy and geoscience application. *Journal of Geophysical Research : Earth Surface* 117 (F3).
- James, M. R., Robson, S., 2014. Mitigating systematic error in topographic models derived from uav and ground-based image networks. *Earth Surface Processes and Landforms* 39 (10), 1413–1420.
- Jazayeri, I., Fraser, C., 2008. Interest operators in close-range object reconstruction. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 37, B5.
- Kaiser, A., Neugirg, F., Rock, G., Müller, C., Haas, F., Ries, J., Schmidt, J., 2014. Small-scale surface reconstruction and volume calculation of soil erosion in complex moroccan gully morphology using structure from motion. *Remote Sensing* 6 (8), 7050–7080.
- Kazhdan, M., Hoppe, H., 2013. Screened poisson surface reconstruction. *ACM Transactions on Graphics (ToG)* 32 (3), 29.
- Ke, Y., Sukthankar, R., 2004. Pca-sift : A more distinctive representation for local image descriptors. In : *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. Vol. 2. IEEE, pp. II–II.
- Kennie, T. J. M., McKay, W. M., 1987. Monitoring of geotechnical processes by close range photogrammetry. *Geological Society, London, Engineering Geology Special Publications* 4 (1), 163–170.
URL <http://egsp.lyellcollection.org/content/4/1/163>
- Khoshelham, K., Díaz-Vilariño, L., 2014. 3d modelling of interior spaces : Learning the language of indoor architecture. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40 (5), 321.
- Koci, J., Leon, J. X., Jarihani, B., Sidle, R. C., Wilkinson, S. N., Bartley, R., 2017. Strengths and limitations of uav and ground-based structure from motion photogrammetry in a gullied savanna catchment. Preprints.
- Koltsida, I., Tomor, A., Booth, C., 2013. The use of digital image correlation technique for monitoring masonry arch bridges.
- Li, Z., Wang, J., 2014. Least squares image matching : A comparison of the performance of robust estimators. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2 (1), 37.

- Lindeberg, T., 1998. Feature detection with automatic scale selection. *International journal of computer vision* 30 (2), 79–116.
- Lingua, A., Marenchino, D., Nex, F., 2009. Performance analysis of the sift operator for automatic feature extraction and matching in photogrammetric applications. *Sensors* 9 (5), 3745–3766.
- Lisein, J., 2016. Application des techniques de photogrammétrie par drone à la caractérisation des ressources forestières. Ph.D. thesis, thèse de doctorat dirigée par Pierrot-Deseilligny, Marc et Lejeune, Philippe Sciences et Technologies de l'Information Géographique Paris Est 2016.
URL <http://www.theses.fr/2016PESC1049>
- Lisein, J., Pierrot-Deseilligny, M., Bonnet, S., Lejeune, P., 2013. A photogrammetric workflow for the creation of a forest canopy height model from small unmanned aerial system imagery. *Forests* 4 (4), 922–944.
- Lowe, D. G., 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60 (2), 91–110.
- Luhmann, T., 2010. Close range photogrammetry for industrial applications. *ISPRS journal of photogrammetry and remote sensing* 65 (6), 558–569.
- Marcato Junior, J., Moraes, M. V. A. d., Tommaselli, A. M. G., 2015. Experimental assessment of techniques for fisheye camera calibration. *Boletim de Ciências Geodésicas* 21 (3), 637–651.
- Martin, O., Meynard, C., Pierrot-Deseilligny, M., Souchon, J., Thom, C., 2014. Réalisation d'une caméra photogrammétrique ultralégère et de haute résolution.
- Matas, J., Chum, O., Urban, M., Pajdla, T., 2004. Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing* 22 (10), 761–767.
- McGlone, J. C., 2013. *Manual of Photogrammetry - Sixth Edition*. ASPRS.
- Micheletti, N., Chandler, J. H., Lane, S. N., 2015. Investigating the geomorphological potential of freely available and accessible structure-from-motion photogrammetry using a smartphone. *Earth Surface Processes and Landforms* 40 (4), 473–486.
- Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence* 27 (10), 1615–1630.
- Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L., 2005. A comparison of affine region detectors. *International journal of computer vision* 65 (1-2), 43–72.
- Morel, J.-M., Yu, G., 2009. Asift : A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences* 2 (2), 438–469.
- Mortensen, E. N., Deng, H., Shapiro, L., 2005. A sift descriptor with global context. In : *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE computer society conference on*. Vol. 1. IEEE, pp. 184–190.

- Moulon, P., 2014. Positionnement robuste et précis de réseaux d'images. Ph.D. thesis, thèse de doctorat dirigée par Marlet, Renaud Signal, Image, Automatique Paris Est 2014.
 URL <http://www.theses.fr/2014PEST1033>
- Moutinho, O. F. G., 2015. Evaluation of photogrammetric solutions for rpas : Commercial vs open source.
- Muja, M., 2013. Scalable nearest neighbour methods for high dimensional data. Ph.D. thesis, University of British Columbia.
- Muja, M., Lowe, D. G., 2014. Scalable nearest neighbor algorithms for high dimensional data. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (11), 2227–2240.
- MUJA, M. A., LOWE, D. G., 2009. Fast approximate nearest neighbors with automatic algorithm configuration. *VISAPP*, 331–340.
- Murtiyoso, A., Grussenmeyer, P., Börlin, N., Vandermeerschen, J., Freville, T., 2018. Open source and independent methods for bundle adjustment assessment in close-range uav photogrammetry. *Drones* 2 (1), 3.
- Nikolova, I., 2014. Performance analysis of robust image features detection algorithms. *Information Technologies and Control* 11 (3), 2–15.
- Nocerino, E., Menna, F., Remondino, F., 2014. Accuracy of typical photogrammetric networks in cultural heritage 3d modeling projects. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40 (5), 465.
- Nocerino, E., Menna, F., Remondino, F., Saleri, R., 2013. Accuracy and block deformation analysis in automatic uav and terrestrial photogrammetry-lesson learnt. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 2, 5.
- Noury, N., Sur, F., Berger, M.-O., 2010. How to overcome perceptual aliasing in asift ? In : *International Symposium on Visual Computing*. Springer, pp. 231–242.
- Ordóñez, C., Riveiro, B., Arias, P., Armesto, J., 2009. Application of close range photogrammetry to deck measurement in recreational ships. *Sensors* 9 (9), 6991–7002.
- Paparoditis, N., Papelard, J.-P., Cannelle, B., Devaux, A., Soheilian, B., David, N., Houzay, E., 2012. Stereopolis ii : A multi-purpose and multi-sensor 3d mobile mapping system for street visualisation and 3d metrology. *Revue française de photogrammétrie et de télédétection* 200 (1), 69–79.
- Paul-Henri FAURE, V. T., Jacques BEILIN, M. P.-D., 2014. Métrologie par photogrammétrie aéroportée légère appliquée aux digues - optimiser l'auscultation des ouvrages. In : *Revue XYZ*. Vol. 141. pp. 47–54.
- Pierrot-Deseilligny, M., Paparoditis, N., 2006. A multiresolution and optimization-based image matching approach : An application to surface reconstruction from spot5-hrs stereo imagery. *Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (1/W41).

- Pons, J.-P., Keriven, R., Faugeras, O., 2007. Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *International Journal of Computer Vision* 72 (2), 179–193.
 URL <http://www.di.ens.fr/sierra/pdfs/07ijcv.pdf>
- Ravanelli, R., Nascetti, A., Di Rita, M., Belloni, V., Mattei, D., Nisticò, N., Crespi, M., 2017. A new digital image correlation software for displacements field measurement in structural applications. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 42, 139.
- Remondino, F., 2006. Detectors and descriptors for photogrammetric applications. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (3), 49–54.
- Remondino, F., El-Hakim, S. F., Gruen, A., Zhang, L., 2008. Turning images into 3-d models. *IEEE Signal Processing Magazine* 25 (4).
- Remondino, F., Spera, M. G., Nocerino, E., Menna, F., Nex, F., 2014. State of the art in high density image matching. *The Photogrammetric Record* 29 (146), 144–166.
- Remondino, F., Spera, M. G., Nocerino, E., Menna, F., Nex, F., Gonizzi-Barsanti, S., 2013. Dense image matching : comparisons and analyses. In : *Digital Heritage International Congress (DigitalHeritage)*, 2013. Vol. 1. IEEE, pp. 47–54.
- Renaudin, V., Yalak, O., Tomé, P., 2007a. Hybridization of mems and assisted gps for pedestrian navigation. Tech. rep.
- Renaudin, V., Yalak, O., Tomé, P., Merminod, B., 2007b. Indoor navigation of emergency agents. *European Journal of Navigation* 5 (3), 36–45.
- Rieke-Zapp, D. H., Nearing, M. A., 2005. Digital close range photogrammetry for measurement of soil erosion. *The Photogrammetric Record* 20 (109), 69–87.
- Rolin, P., Berger, M.-O., Sur, F., 2015. Viewpoint simulation for camera pose estimation from an unstructured scene model. In : *International Conference on Robotics and Automation*.
- Roma, N., Santos-Victor, J., Tomé, J., 2002. A comparative analysis of cross-correlation matching algorithms using a pyramidal resolution approach. In : *Empirical Evaluation Methods in Computer Vision*. World Scientific, pp. 117–142.
- Rosenholm, D., 1987. Least squares matching method : some experimental results. *The Photogrammetric Record* 12 (70), 493–512.
- Rosten, E., 2006. Fast corner detection. Engineering Department, Machine Intelligence Laboratory, University of Cambridge. Available from :< <http://mi.eng.cam.ac.uk/er258/index.html>.
- Rosu, A.-M., Assenbaum, M., De la Torre, Y., Pierrot-Deseilligny, M., 2015a. Coastal digital surface model on low contrast images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40 (3), 307.

- Rosu, A.-M., Pierrot-Deseilligny, M., Delorme, A., Binet, R., Klinger, Y., 2015b. Measurement of ground displacement from optical satellite image correlation using the free open-source software micmac. *ISPRS Journal of Photogrammetry and Remote Sensing* 100, 48–59.
- Rupnik, E., Daakir, M., Deseilligny, M. P., 2017. Micmac—a free, open-source solution for photogrammetry. *Open Geospatial Data, Software and Standards* 2 (1), 14.
- Rupnik, E., Deseilligny, M. P., Delorme, A., Klinger, Y., 2016. Refined satellite image orientation in the free open-source photogrammetric tools apero/micmac. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 3, 83.
- Salvi, J., Armangué, X., Batlle, J., 2002. A comparative review of camera calibrating methods with accuracy evaluation. *Pattern recognition* 35 (7), 1617–1635.
- Samaan, M., 2016. La photogrammétrie rapprochée au service de l’archéologie préventive. Ph.D. thesis, Paris Est.
- Schneider, D., Schwalbe, E., Maas, H.-G., 2009. Validation of geometric models for fisheye lenses. *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (3), 259–266.
- Siebert, T., Becker, T., Spiltthof, K., Neumann, I., Krupka, R., 2007. Error estimations in digital image correlation technique. In : *Applied Mechanics and Materials*. Vol. 7. Trans Tech Publ, pp. 265–270.
- Silpa-Anan, C., Hartley, R., 2008. Optimised kd-trees for fast image descriptor matching. In : *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, pp. 1–8.
- Smith, S. M., Brady, J. M., 1997. Susan—a new approach to low level image processing. *International journal of computer vision* 23 (1), 45–78.
- Soontranon, N., Pierrot-Deseilligny, M., Mouaddib, E. M., 2013. Mise en correspondance de points 3d obtenus avec une grande” base-line”. In : *Orasis, Congrès des jeunes chercheurs en vision par ordinateur*.
- Sun, Y., Zhao, L., Huang, S., Yan, L., Dissanayake, G., 2014. L2-sift : Sift feature extraction and matching for large images in large-scale aerial photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing* 91, 1–16.
- Sur, F., Noury, N., Berger, M.-O., 2011. Image point correspondences and repeated patterns. Ph.D. thesis, INRIA.
- Swaminathan, R., Nayar, S. K., 2000. Nonmetric calibration of wide-angle lenses and polycameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (10), 1172–1178.
- Tang, Z., Monasse, P., Morel, J.-M., 2014. Improving the matching precision of sift. In : *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, pp. 5756–5760.

- Thomas Kraft, H. M., Ralf Berger, K. S., 2017. Preliminary results of developing a metric aerial camera for uav applications. UAV Metric Camera Special Sessions DGPF Annual Meeting 2017 Würzburg, Germany.
- Toth, C., Paska, E., 2007. Mobile mapping and autonomous vehicle navigation. *Revue Francaise Photogramm. Teledetection* 185, 57–61.
- Tournadre, V., Dec. 2015. Metrology with light aerial photogrammetry : application to dykes monitoring. Theses, Université Paris-Est.
URL <https://tel.archives-ouvertes.fr/tel-01373128> :tel-01373128
- Tournadre, V., Pierrot-Deseilligny, M., Faure, P., 2015. Uav linear photogrammetry. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40 (3), 327.
- Tournadre, V., Pierrot-Deseilligny, M., Faure, P.-H., 2014. Uav photogrammetry to monitor dykes-calibration and comparison to terrestrial lidar. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40 (3), 143.
- Triggs, B., McLauchlan, P. F., Hartley, R. I., Fitzgibbon, A. W., 1999. Bundle adjustment—a modern synthesis. In : *International workshop on vision algorithms*. Springer, pp. 298–372.
- Vedaldi, A., 2007. An open implementation of the SIFT detector and descriptor. Tech. Rep. 070012, UCLA CSD.
- Von Gioi, R. G., Jakubowicz, J., Morel, J.-M., Randall, G., 2012. Lsd : a line segment detector. *Image Processing On Line* 2, 35–55.
- Wackrow, R., Chandler, J., Gardner, T., 2008. Minimising systematic errors in dems caused by an inaccurate lens model. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37 (B5), 1–6.
- Weng, J., Cohen, P., Herniou, M., et al., 1992. Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on pattern analysis and machine intelligence* 14 (10), 965–980.
- Werner, T., Schaffalitzky, F., Zisserman, A., 2002. Automated architecture reconstruction from close-range photogrammetry. *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY REMOTE SENSING AND SPATIAL INFORMATION SCIENCES* 34 (5/C7), 352–359.
- Wu, C., 2014. Critical configurations for radial distortion self-calibration. In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 25–32.
- Wu, J., Cui, Z., Sheng, V. S., Zhao, P., Su, D., Gong, S., 2013. A comparative study of sift and its variants. *Measurement science review* 13 (3), 122–131.
- Yang, M. Y., Cao, Y., Förstner, W., McDonald, J., 2010. Robust wide baseline scene alignment based on 3d viewpoint normalization. In : *International Symposium on Visual Computing*. Springer, pp. 654–665.

Zheltoy, S., Sibiryakov, A., 1997. Adaptive subpixel cross-correlation in a point correspondence problem. Proceedings of Optical 3D Measurement Techniques, 86–95.