THÈSE

présentée pour obtenir le grade de

Docteur de l'Université Paris Descartes

Spécialité : Mathématiques et informatique

Lionel PÉNARD

Reconstruction de façades de bâtiments par appariement dense d'images terrestres géoréférencées haute résolution

Soutenue le 31/03/2009 devant le jury composé de :

M. STAMON Georges – Président, CRIP5, Université Paris Descartes
M. STURM Peter – Rapporteur, INRIA, Grenoble
M. DHOME Michel – Rapporteur, LASMEA, Université Clermont-Ferrand
Mme BAILLARD Caroline – Examinatrice, SIRADEL, Rennes
M. GOULETTE François – Examinateur, CAOR, École des Mines de Paris
Mme VINCENT Nicole – Examinatrice, CRIP5, Université Paris Descartes
M. PAPARODITIS Nicolas – Encadrant, MATIS, IGN, Saint-Mandé
M. PIERROT-DESEILLIGNY Marc – Directeur de thèse, MATIS, IGN, Saint-Mandé

SOMMAIRE

Introduction 5

1 Contexte et état de l'art 9

2 Appariement dense d'images pour la reconstruction 3D 27

3 Reconstruction de surface 59

4 Vers une gestion des parties cachées pour l'appariement 81

Conclusion 97

Bibliographie 101

Table des figures 112

Liste des tableaux 113

Liste des algorithmes 115

INTRODUCTION

L'environnement urbain constitue un objet très complexe. Afin de l'appréhender, de l'organiser, de le développer, de l'aménager, d'archiver son état patrimonial, de s'y repérer ou de le visiter, sont apparues très tôt dans l'histoire humaine des *représentations* urbaines. Plans de ville, schémas d'aménagement, relevés patrimoniaux existent en effet depuis longtemps. La mise au point de générations d'ordinateurs puissants et grand public, capables de stocker et de manipuler de grands ensembles de données, ainsi que de les restituer en trois dimensions, a permis de franchir une nouvelle étape dans ces représentations : les modèles urbains tridimensionnels. La figure 1 montre un exemple d'image aérienne et de modèle urbain tridimensionnel.

Parallèlement au développement informatique ont émergé de nombreuses applications pour lesquelles ces modèles peuvent être utiles sinon indispensables. Sans prétendre à l'exhaustivité, en voici quelques exemples remarquables. Tout d'abord, l'ère de la mobilité (téléphone portable, WiFi, 3G,...) nécessite des simulations de propagation d'ondes électromagnétiques, pour lesquelles une bonne connaissance de la ville en trois dimensions est requise. Il en est de même pour la propagation des ondes acoustiques. En effet les nuisances sonores représentent désormais un enjeu majeur pour le citadin. Dans le domaine environnemental, les modèles tridimensionnels urbains peuvent également permettre, alliés à la dynamique des fluides, de modéliser la dispersion de polluants dans l'air ou le passage des eaux d'une rivière en crue. L'aménagement urbain peut aussi tirer partie de telles maquettes virtuelles, enrichies ou modifiées par les différents scénarios, en autorisant la prise de décision sur des bases plus objectives qu'auparavant. Les applications culturelles sont également nombreuses avec, par exemple, le souhait de parvenir à une conservation du patrimoine dématérialisée, ou encore de rendre possible un véritable tourisme virtuel, avec liberté de déplacement, sans risque de dégradation des monuments. Enfin, les puissantes industries des jeux vidéos et du cinéma mettent déjà depuis quelques années à profit les modèles tridimensionnels urbains, la première pour renforcer chez les joueurs la sensation d'immersion en proposant comme environnement de jeu de véritables reproductions de la réalité, la seconde préférant parfois pour des raisons de budget ou de simple faisabilité tourner dans des environnements virtuels co-

(a) Image aérienne du centre ville de Lyon (*source : Microsoft live maps*)

(b) Modèle 3D texturé du centre ville de Lyon (*source : PagesJaunes / InterAtlas / IGN / ArchiVideo*)

FIG. 1 – Image aérienne et modèle tridimensionnel d'une portion du centre ville de Lyon.

pies conformes de la réalité.

Face à ces nouveaux besoins et ces potentialités en termes de puissance des ordinateurs, il est nécessaire d'apporter des réponses en terme d'acquisition de données, puis de constitution des modèles tridimensionnels. Mais ces modèles sont-ils vraiment nécessaires ? La mise en ligne du service Street View de Google (service qui propose une navigation dans les rues des grandes métropoles des Etats-Unis, à l'aide d'une base de données très dense d'images panoramiques – voir figure 2) rend légitime cette question. Avec uniquement des panoramiques images, sans aucune données tridimensionnelles – hormis le géoréférencement des images, l'utilisateur est réellement immergé au coeur de la ville et peut en explorer toutes les rues et tourner la tête dans toutes les directions. Le réalisme est poussé au maximum puisque ce qui est vu n'est qu'un assemblage de photographies. Cependant ce système possède ses limitations. Pour l'application balade virtuelle il reste suffisant, même s'il est impossible de s'écarter de la trajectoire d'acquisition. En revanche, ce besoin de trois dimensions est impératif dans les applications, notamment de simulation, mentionnées plus haut. De plus, l'IGN – Institut Géographique National – en tant qu'agence cartographique, s'intéresse à la description sémantique mais surtout géométrique des objets existants. Ce sont en effet ces données qui répondent aux besoins de calculs et de simulations. Une base de données image n'est dans ce cadre qu'un complément d'un modèle géométrique en trois dimensions.

Les éléments parmi les plus présents et structurants des scènes urbaines vues du sol sont les bâtiments, et plus précisément les façades de bâtiments. Ce document présente des travaux permettant d'obtenir une description en trois dimensions des façades, de la manière la plus automatique possible à partir d'un ensemble d'images : c'est ce que l'on appelle la *reconstruction* de façades de bâtiments. Bien sûr, les façades ne constituent qu'un des éléments de l'environnement urbain et même s'il s'agit de du plus important, certaines applications en nécessitent d'autres : voirie, mobilier urbain, végétation. De nombreuses équipes investiguent ces thèmes de recherche.

Ce document suit le plan suivant :

FIG. 2 – Vue de rues de San Francisco dans Street View (*source : Google Maps*)

Le chapitre 1 est consacré à la présentation du contexte de reconstruction tridimensionnelle urbaine. Il fournit un panorama des systèmes d'acquisition terrestres mobiles et l'état de l'art associé, ainsi que les caractéristiques du système développé au sein du laboratoire MATIS et des images utilisées dans le cadre de cette thèse. Un état de l'art des techniques de reconstruction de bâtiments à partir d'images terrestres est également dressé. Le chapitre 2 présente l'appariement dense d'images pour la reconstruction en trois dimensions de façades de bâtiments. Les cas stéréoscopique d'une part et multi-stéréoscopique d'autre part sont successivement abordés. La méthode proposée associe mise en correspondance d'images par corrélation, régularisation par coupe minimale de graphes et processus multi-résolution. Dans le chapitre 3 est présentée la représentation par maillages triangulés des surfaces obtenues, et propose une méthode de fusion de telles surfaces en recouvrement partiel. Le chapitre 4 fournit des réponses au problème d'occultation partielle de la scène dans les images, dans le cadre de l'algorithme de reconstruction précédemment exposé. Enfin, le dernier chapitre (page 97) donne les conclusions de ces travaux et propose quelques perspectives.

CHAPITRE 1

CONTEXTE ET ÉTAT DE L'ART

1.1 Contexte général de la reconstruction de bâtiments

La reconstruction automatique de façades s'inscrit dans la thématique plus large de la reconstruction automatique ou semi-automatique de bâtiments, sujet de recherche difficile et abordé depuis plus de vingt ans. La figure 1.1 montre un exemple de reconstruction actuellement disponible pour le grand public.

FIG. 1.1 – Vue d'un modèle 3D de Paris (*source PagesJaunes/IGN/ArchiVidéo*)

Les approches envisagées se distinguent selon le niveau de détail recherché : pour la reconstruction d'édifices ou de monuments remarquables (par leur architecture, leur histoire, leur valeur de symbole dans une ville) pour lesquels le niveau de détail souhaité est important, ou des bâtiments standards pour lesquels une description basée sur des modèles plus simples est suffisante.

Dans ce dernier cas, les travaux se sont d'abord focalisés sur l'utilisation de données satellitaires

ou aériennes. En effet, l'acquisition de données aériennes ou satellitaires est automatique ou quasi-automatique. Ces données couvrent une large portion du sol contenant l'ensemble des bâtiments à reconstruire. Les données en question sont souvent des images, (voir par exemple FRÈRE et al. 1998, HUERTAS et al. 1999, basé sur une segmentation de la scène grâce à des données hyperspectrales, BRENNER et al. 2001, qui utilise une segmentation d'un modèle numérique de surface). Elles peuvent être enrichies de nuages de points laser (par exemple dans HAALA et BRENNER 1997 , MARTIN HUBER et BAUMGARTNER 2003 ou encore BRETAR et ROUX 2006).

Les techniques de reconstruction sont nombreuses. Il existe les modélisations par approches descendantes ("top-down"), qui partent d'une base de modèles et cherchent à les ajuster sur les données et les approches ascendantes ("bottom-up"), qui s'appuient sur des primitives (points, segments, éléments linéaires, facettes,...) pour construire des modèles.

En revanche, l'acquisition de données terrestres sur un territoire large comme une ville, représente un ensemble de données très volumineux, et tant le stockage que le traitement de tels volumes étaient rédhibitoires il y a encore quelques années, ce qui explique le choix premier de l'avion et du satellite.

Dans le cas de bâtiments ou de monuments remarquables, puisqu'il s'agit de zones d'étude beaucoup plus réduites, les données terrestres sont fréquemment utilisées, éventuellement de manière conjointe avec des données aériennes, afin de pouvoir en décrire toute la richesse architecturale . Là encore, images et laser peuvent être utilisés, de manière séparée ou, de plus en plus, de manière conjointe

Cependant, même dans le cas de bâtiments ordinaires, il peut être intéressant de disposer de données terrestres, pour améliorer la description géométrique mais surtout radiométrique au niveau des façades. De nombreux auteurs tentent de combiner modèles aériens et acquisition terrestre (par exemple BRENNER et al. 2001). L'acquisition manuelle de données terrestres interdit de pouvoir réaliser des chantiers étendus, en raison de son coût prohibitif. Si depuis quelques années plusieurs équipes utilisent massivement des données terrestres, c'est que d'une part le matériel informatique s'est amélioré en augmentant l'espace de stockage disponible et en accélérant fortement les traitements, et d'autre part que des techniques d'acquisition plus automatiques ont été mises au point.

L'automatisation de l'acquisition et du traitement des images terrestres est relativement récente. Les travaux de Teller et de son équipe du City Scanning Project du MIT (COORG et al. 1998) constituent

une des plus anciennes tentatives de collecte rapide d'images terrestres. L'équipe a développé une plateforme mobile constituée d'un appareil photo pouvant tourner autour d'un point central, d'une antenne GPS et d'un ordinateur. Cette plateforme est destinée à prendre des images dans toutes les directions (en pratique, une demi sphère) depuis un point de vue fixe, appelé "noeud". Ensuite, manuellement, la plateforme est déplacée pour réaliser une acquisition d'un autre point de vue, et ainsi de suite. Le positionnement et l'orientation absolus sont déterminés automatiquement, à partir d'une solution approchée donnée par le GPS, en exploitant les recouvrements d'images entre deux noeuds voisins. Ce mode d'acquisition a permis la reconstruction et la texturation de plusieurs bâtiments du campus du MIT (TELLER et al. 2003). Cependant, on reste dans le domaine du semi-automatique pour l'acquisition, ce qui limite la zone de travail visée.

1.2 Contexte particulier de la cartographie mobile

Les recherches plus récentes reposent toutes sur l'utilisation d'un véhicule pour accélérer et automatiser l'acquisition d'images et/ou de nuages de points laser. Ces techniques sont bien adaptées aux contextes de milieu urbain, où un réseau routier dense est présent. Bien sûr l'acquisition n'est possible que pour les façades donnant sur la rue, au détriment des cours intérieures ou des bâtiments non accessibles par voie carrossable. L'exemple du Mont Saint-Michel, fermé à toute circulation, le montre bien. Pour pallier ces limitations, des alternatives sont à l'étude : un complément manuel avec un piéton muni du système d'acquisition rendu portable et autonome du point de vue énergétique, ou un complément automatisé à l'aide de drones volant à hauteur de façade et embarquant les dispositifs d'acquisition. Cependant, pour la grande majorité de l'espace urbain, un véhicule est une solution très adaptée pour numériser rues et façades.

Les dispositifs s'appuyant sur de tels véhicules sont appelés systèmes de cartographie mobile (notés MMS pour "Mobile Mapping Systems" en anglais) et sont constitués de véhicules équipés de systèmes de navigation et de systèmes d'acquisition de données.

Les systèmes embarqués de navigation sont généralement les suivants : une ou plusieurs antennes

GPS, une centrale inertielle, un odomètre. Ce sont eux qui permettent le géoréférencement du véhicule, c'est à dire l'obtention de sa position et de son orientation dans un système de coordonnées cartographique global. Ceci permet de recaler des données acquises par ce système avec des données exogènes telles que des cartes, des bases de données préexistantes (qu'elles décrivent des routes, des bâtiments, le cadastre...), ou bien encore d'autres acquisitions de données image et/ou laser.

Les systèmes d'acquisition permettent de capturer sous forme numérique des données de la scène. Pour la partie géométrique, un ou plusieurs laser sont fréquemment utilisés. Les données radiométriques sont obtenues à partir d'une ou plusieurs caméras, vidéo ou photographique. L'avantage des caméras photographiques est qu'elles peuvent être synchronisées entre elles et ainsi être utilisées pour des algorithmes de vision stéréo pour scènes statiques. Ces données image sont précieuses à plus d'un titre. Elles peuvent en effet être utilisées pour texturer des modèles déjà reconstruits mais aussi pour extraire de l'information géométrique, ou encore sémantique. Enfin, les images peuvent servir à de la visualisation pure, pour des promenades virtuelles parmi les points de vue.

Le champ d'application de tels systèmes est très large : numérisation du patrimoine à des fins de conservation, production de bases de données images pour de la visualisation (par exemple Street View de Google), enrichissement de la géométrie tridimensionnelle des modèles urbains, état des lieux des infrastructures routières et de signalisation (observation de murs de soutènement, état de la chaussée, marquages au sol, panneaux et feux lumineux, ...), production de bases de données d'amers visuels destinés à permettre la navigation autonome d'autres véhicules,... Le choix des capteurs et des processus de traitement dépendra de l'application visée.

Parmi les équipes qui travaillent sur des données produites à partir de tels systèmes d'acquisition embarqués sur des véhicules, beaucoup utilisent comme source d'information géométrique sur les façades un scanner à balayage laser. L'image est utilisée uniquement pour texturer le modèle obtenu.

1.2.1 Etat de l'art : reconstruction par systèmes de cartographie mobiles

Dans cette partie, nous donnons un panorama des principales équipes travaillant sur la reconstruction 3D à partir de données acquises par des systèmes embarqués à bord de véhicules.

Equipe du laboratoire Vidéo et traitement d'image de UC Berkeley

Les travaux de Früh et Zahkor (FRÜH 2002 ; FRÜH et ZAKHOR 2004 ; FRÜH et al. 2005) portent sur la réalisation de toute une chaîne de production de modèles 3D photo-réalistes de villes, à partir de données aériennes et terrestres, image et laser. Le processus est entièrement automatique, depuis la collecte des données jusqu'à la réalisation des textures de façades. Le véhicule de ce projet est représentés figure 1.2(a).

(a) Véhicule équipé

(b) Nuage de points laser

(c) Image de texture redressée

(d) Modèle 3D avec données terrestres et aériennes fusionnées

FIG. 1.2 – Véhicule et reconstruction de l'équipe de Berkeley (*source EECS - UC Berkeley*)

A partir du nuage de points laser aérien, un modèle numérique de surface est créé. Une série de traitements aplatit les toits et supprime les petits éléments isolés, puis les toits sont découpés par des arêtes rectilignes. Le résultat est triangulé, simplifié puis texturé. La création du modèle 3D de façade est plus complexe. En effet il est nécessaire de réaliser le recalage de toutes les lignes verticales de scan 2D dans un unique référentiel géométrique. L'idée des auteurs est d'utiliser un deuxième scanner à balayage

horizontal cette fois. La mise en correspondance de deux de ces scans à deux instants consécutifs permet d'obtenir le mouvement en 2D du véhicule sur la route considérée comme plane. La trajectoire est ensuite recalée dans le système global de coordonnées pour corriger les dérives inhérentes au recalage relatif précédent. Ceci est fait par recalage des points laser terrestres (qui sont en majorité des points sur les façades) sur les discontinuités du MNS aérien, puis par l'utilisation d'une base de données routière en contraignant la trajectoire du véhicule à rester sur les tronçons routiers. La figure 1.2(b) montre un exemple de nuage de points remis en géométrie.

L'étape suivante est l'obtention du modèle 3D des façades. Une segmentation est faite pour discriminer l'avant plan (voitures, végétation, piétons,...) des façades proprement dites, puis les trous dus aux occultations du modèle de bâtiment sont rebouchés, et les points sont triangulés. Les textures sont ensuite calculées automatiquement. Une attention particulière a été portée à la réalisation de textures réalistes aux endroits où les occultations masquaient dans toutes les images l'apparence du bâtiment (FRÜH et al. 2005, et voir figure 1.2(c)). Les deux modèles, terrestre et aérien, sont alors fusionnés (figure 1.2(d)). Les résultats obtenus sont très bons : rappelons que tout ce processus est automatique et de plus les performances en termes de temps de calcul sont remarquables.

Equipe du 'Center for Spatial Information Science', université de Tokyo

Les travaux de Zhao et Shibasaki (ZHAO et SHIBASAKI 2001) reposent sur le même principe d'acquisition rapide et simple de données à l'aide d'une plateforme embarquée sur un véhicule. 3 scanners 2D ainsi que 6 caméras CCD ligne sont utilisés. Ici la localisation est assurée par GPS, INS et odomètres. Le géoréférencement de toutes les données est là encore un problème délicat. S'il est possible de calibrer le laser et l'image par rapport à la plateforme de navigation GPS/INS/odomètre, le géoréférencement absolu de l'ensemble, pour pallier les problèmes de dérive, est réalisée de manière interactive, en identifiant des points de liaisons (coins de bâtiments) entre les données terrestres et un MNS aérien. La qualité de géoréférencement dépend donc fortement de celle du MNS. La figure 1.3 montre les résultats obtenus. Une interface semi-automatique permet ensuite de modéliser les objets souhaités (principalement les bâtiments).

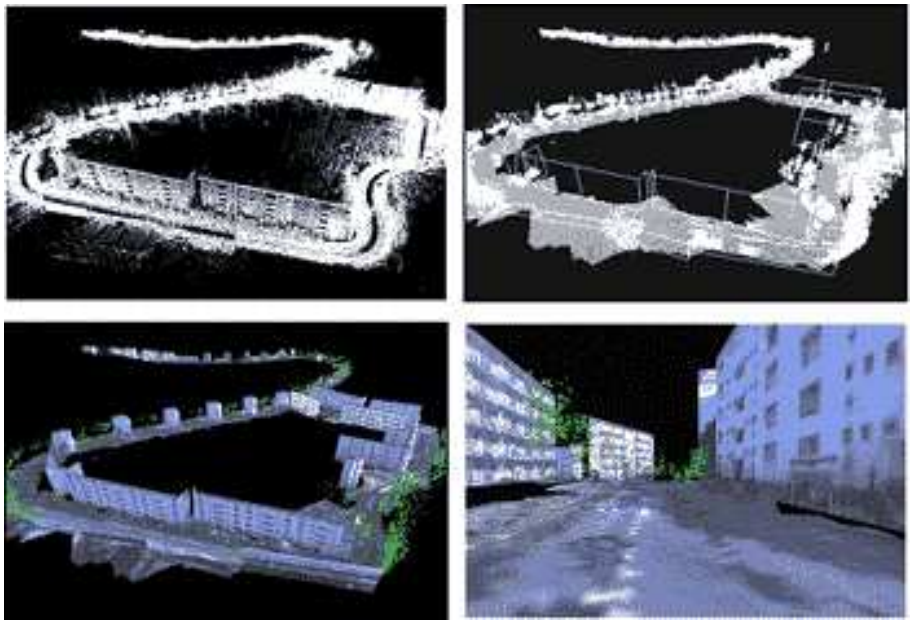


FIG. 1.3 – Résultats de la reconstruction automatique (*source : Center for Spatial Information Science, Tokyo University*)

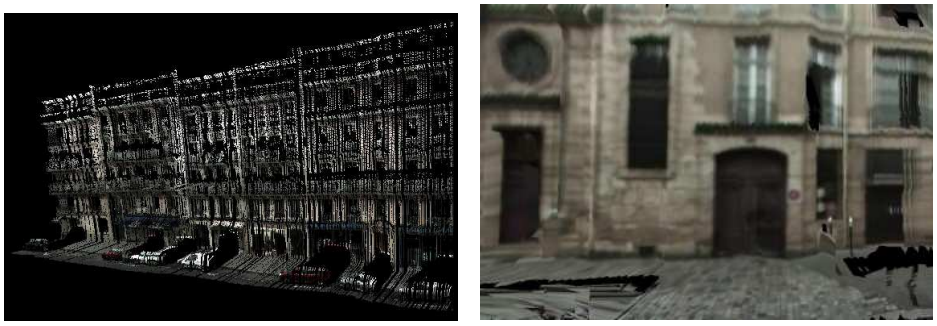
Equipe du département d'informatique, Université Columbia, New York

D'autres auteurs ont souhaité automatiser la collecte de données terrestres. On peut notamment citer les travaux de ALLEN et al. 2001, qui utilisent un robot équipé de capteurs de positionnement (GPS et odomètre), d'une caméra et d'un système laser. L'objectif est d'automatiser tant la partie acquisition que la partie modélisation. Là encore, les auteurs se servent de la complémentarité laser/images pour reconstruire la géométrie d'une part, les textures d'autre part.

Equipe du centre de robotique de l'ENSMP

En France, l'équipe du projet de route automatisée, au sein du centre de robotique de l'Ecole des Mines de Paris, a réalisé des travaux sur l'acquisition de données et la modélisation 3D (ABUHADROUS et al. 2004 ; BRUN et al. 2007). Un véhicule muni d'une antenne GPS et d'une centrale inertielle pour la partie navigation, ainsi que d'un scanner laser pour l'acquisition de la géométrie et une capteur CCD avec optique très grand angle pour les textures. Les auteurs proposent une stratégie pour géoréférencer toutes les données pour reconstruire l'environnement texturé en 3D. La figure 1.4 montre un nuage de

points laser et une reconstruction par facettes, triangulation directe du nuage précédent texturée avec l'image. Le résultat obtenu souligne la difficulté de recalage des points laser entre eux d'une part et avec l'image d'autre part.



(a) Nuage de points laser (b) Reconstruction par maillage triangulé

FIG. 1.4 – Reconstruction d'une portion de rue de Paris (*source : CAOR - ENSMP*)

Equipe GRAVIR du LASMEA

Au sein du LASMEA à Clermont-Ferrand, dans le cadre d'applications robotiques, l'équipe GRAVIR a conçu un système basé sur le "SLAM" (Simultaneous Localization And Mapping), c'est à dire que le véhicule doit réaliser dans le même temps la cartographie ou modélisation de l'environnement et s'y repérer (MOURAGNON et al. 2006). L'algorithme utilisé est incrémental : à partir des vues déjà mises en place, le géoréférencement d'une nouvelle image se fait par relèvement à partir de points de Harris dont des homologues dans les images précédentes sont déjà passés en 3D. Ils parviennent de cette manière à assurer une mise en place robuste et temps réel.

Projet Urbanscape

Ce projet est mené en collaboration entre l'université de Caroline du nord Chapel Hill et l'université du Kentucky (voir `http://cs.unc.edu/Research/urbanscape/`). Ici, il n'y a pas de scanner laser pour acquérir la géométrie de la scène ; les données sont obtenues par une base de quatre caméras vidéos synchronisées. Le géoréférencement des images est obtenu par fusion d'une part des données des capteurs de navigation (GPS et centrales inertielles) et d'autre part d'un suivi temporel de points

d'intérêt dans les images, au travers d'un filtre de Kalman. Ces points d'intérêts peuvent ensuite être reconstruits en 3D et fournissent une information d'orientation utilisée dans l'étape suivante. Il s'agit de faire calculer par la carte graphique (GPU) des cartes de profondeurs par une technique de balayage par des plans ("plane-sweep") dans de multiples directions. Ensuite est réalisée la fusion des cartes de profondeur et enfin la génération du maillage final. La figure 1.5 donne une illustration de ces travaux.

FIG. 1.5 – Reconstruction de façade par le projet Urbanscape (*source : Urbanscape project UNC / UK*)

1.2.2 Le système Stéréopolis

Si la plupart des équipes utilisent un ou plusieurs laser, la qualité des travaux du projet Urbanscape décrits ci-dessus montre qu'il est possible de ne travailler qu'à partir d'images. Les systèmes basés uniquement sur l'image présentent quelques avantages : diminution du coût, très fort débit d'acquisition des pixels et également une mise en place plus précise par rapport aux systèmes laser dont le recalage est délicat. Le laser apporte une robustesse certaine avec très peu de fautes. Dans tous les cas la présence de capteurs images est nécessaire (au moins pour la texturation).

L'expérience en matière de traitement d'images au sein du laboratoire MATIS a conduit, lors du lancement du projet Stéréopolis en 2002, à choisir la solution d'un véhicule muni de plusieurs caméras, ainsi que d'un système de géoréférencement direct (GPS), sans laser. C'est ce système qui a fourni les images utilisées pour ce travail.

Le système Stéréopolis a été rapidement décrit dans PAPARODITIS et al. 2005 ; BENTRAH et al. 2004, et il est représenté figure 1.6. En ce qui concerne la partie navigation, deux antennes GPS "shoke

FIG. 1.6 – Véhicule du projet Stéréopolis

ring" sont installées chacune sur un mât. Ces antennes sont étudiées pour limiter les multi-trajets en zone urbaine (c'est à dire les ondes des signaux GPS réfléchies par l'environnement - essentiellement les bâtiments - qui peuvent perturber la mesure de position puisqu'elle est réalisée à partir d'une mesure de temps de trajet). La présence de deux antennes permet de vérifier la cohérence entre les deux mesures. La partie acquisition de données est assurée par un ensemble de six caméras. Un signal électronique permet de les synchroniser parfaitement entre elles, ainsi qu'avec le GPS. Ces six caméras sont disposées de sorte d'imager les façades et le corps de rue en stéréo :

- une base stéréoscopique horizontale vise vers l'arrière du véhicule. Elle permet d'imager en stéréo le corps de rue, la voirie, les marquages routiers, les véhicules...
- une base verticale de chaque côté permet d'obtenir des images des façades des bâtiments.

Les caméras utilisées sont produites par le LOEMI (Laboratoire d'Opto-Electronique et de Micro-Informatique) de l'IGN pour l'imagerie aérienne. Elles ont été ici embarquées à bord de la première version du véhicule terrestre Stéréopolis, et sont celles qui ont produit les images sur lesquelles portent tous les travaux décrits dans ce document. Une description détaillée de ces caméras et de leurs principales

applications aériennes est donnée dans PAPARODITIS et al. 2006. Nous résumons ici leurs principales caractéristiques :

- capteur CCD, de taille 4096 x 4096 pixels, panchromatique ;
- dynamique de 2000 à 3000 niveaux de gris stockés sur 12 bits ;
- bruit très faible (rapport signal sur bruit de l'ordre de 300) ;
- vitesse d'acquisition : 3,2 s / image ;
- synchronisation assurée à moins de 10 μs.

Nous pouvons remarquer que ce prototype est complet et fournit de très bonnes images. Cependant il ne s'agit pas d'un système de cartographie mobile idéal pour les raisons qui suivent. Tout d'abord le débit d'images est faible, et il oblige à rouler à une vitesse de 10 $km.h^{-1}$ en centre ville pour assurer un recouvrement suffisant. Ensuite l'absence de couleur sur les capteurs images est regrettable pour toutes les applications de mise en valeur des données et de texturation des modèles. Enfin, le recouvrement vertical des bases stéréoscopiques latérales n'est pas suffisant, nous y revenons au paragraphe suivant.

Dans le cadre des travaux de cette thèse, nous ne nous intéressons qu'aux images de façades. La géométrie des bases verticales est étudiée pour que le champ de vue couvre la totalité du bâtiment en hauteur, en maximisant le recouvrement vertical. Avec une hypothèse d'une distance véhicule-façade de 10 m, en réglant la caméra du bas pour atteindre le pied du bâtiment, il est possible d'imager un bâtiment de 20 m, soit environ 6 étages. La figure 1.7 donne le schéma de prise de vue de la base verticale.

La plupart des bâtiments des centres urbains denses en France peuvent donc être imagés dans toute leur hauteur. Néanmoins nous verrons par la suite que l'hypothèse de distance véhicule-façade est souvent mise en défaut car les rues de ces centres urbains sont souvent étroites, et seul un trottoir étroit, avec parfois une file de stationnement, se situe entre la chaussée et les bâtiments. Il en résulte que seuls les premiers étages des bâtiments des rues étroite sont imagés par notre système. La figure 1.8 montre des images de façades prises par la base verticale du véhicule. Le recouvrement vertical des deux images est restreint à deux étages environ.

Pour ces raisons, un nouveau prototype de véhicule de cartographie mobile est en cours de développement, en tenant compte des limitations de la première version. Ce nouveau système intégre des capteurs images couleurs, à cadence d'acquisition plus grande pour permettre de rouler à la vitesse nor-

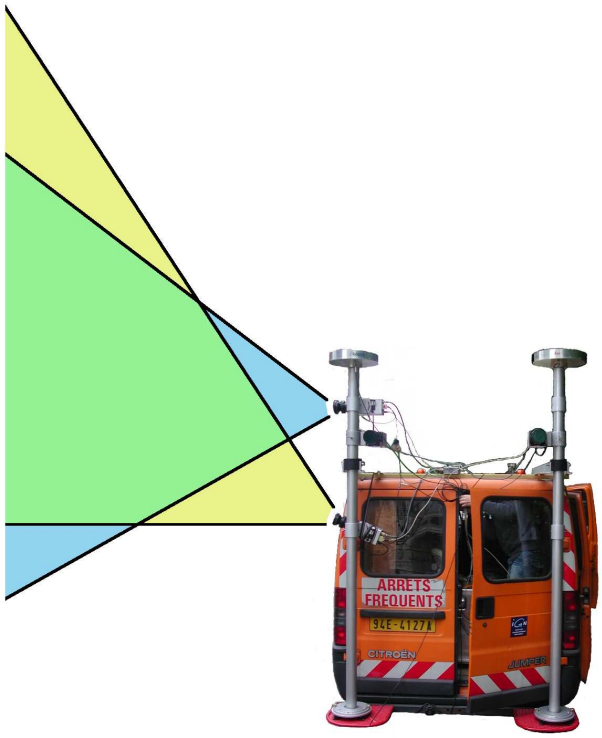


FIG. 1.7 – Schéma montrant la géométrie d'acquisition de la base verticale de Stéréopolis

male en ville, et avec un recouvrement vertical "total", c'est à dire du pied au sommet du bâtiment. Ce système a réalisé ses premières acquisitions au début de l'année 2008.

Lorsque l'estimation de pose (relative et absolue) est réalisée, il est possible de faire la reconstruction 3D à partir d'images. Dans notre cas de travail, plutôt que de tenter d'estimer les paramètres internes sur les images du jeu de données, ils sont estimés *a priori*. L'étape de calibration géométrique interne des caméras est donc réalisée au préalable. L'IGN dispose d'un polygone de calibration 3D, dont les cibles sont mesurées très précisément (les 50 cibles sont mesurées avec une précision de 0,5 mm). La focale, le point principal d'autocollimation ainsi que des paramètres de distorsion sont estimés à partir d'un ensemble de vues de ce polygone, par moindres carrés itérés. Des travaux précédents ont également développé une méthode d'auto-calibration, sans cibles, par calcul d'une carte dense de disparité sur des images d'une scène plane et texturée (voir PIERROT-DESEILLIGNY 2005). L'une et l'autre des deux méthodes précédentes (calibration avec ou sans cibles) peuvent être utilisées de manière conjointe. Les résultats obtenus fournissent des estimations précises des paramètres recherchés (les résidus image sont

FIG. 1.8 – Deux prises de vues consécutives de la base stéréoscopique verticale de Stéréopolis.

inférieurs au $\frac{1}{3}$ de pixel), qui ne seront pas remises en cause dans la suite.

La recherche automatique des paramètres extrinsèques absolus est plus complexe, et fait l'objet de recherches actives dans la communauté scientifique. En effet, les capteurs de navigation classiques (antennes GPS, centrales inertielles, odomètres) ne permettent pas d'obtenir des précisions suffisantes en milieu urbain. Les masques, multi-trajets et configuration satellitaires sont défavorables à un positionnement de qualité dans les canyons urbains - lorsque ceux-ci n'empêchent pas purement et simplement toute localisation. Les instruments de mesure de déplacement relatif dérivent, et sans mesure absolue (GPS) pour se recaler, cette dérive atteint rapidement (quelques secondes) une distance incompatible avec la précision recherchée.

Pour pallier ces problèmes de manière automatique, des auteurs proposent de se recaler sur une carte ou un modèle aérien de la zone, en appariant les points reconstruits avec les limites de bâtiments connues. Cette technique a l'inconvénient de fournir une précision de l'ordre de celle des bâtiments du modèle ou de la carte, généralement supérieure au mètre. Pour assurer une bien meilleure précision, la stratégie qui a été développée au sein du laboratoire repose sur deux idées : améliorer la précision du mouvement

relatif, et augmenter la quantité ainsi que la qualité des points de recalage en absolu. Ces deux méthodes s'appuient sur les images des caméras embarquées, et seront nécessairement complémentaire dans un système opérationnel.

La première stratégie (voir BENTRAH 2006) repose sur l'appariement 3D des scènes reconstruites à deux instants de prise de vue consécutifs pour estimer la transformation rigide qu'a subie le véhicule. Cela permet d'améliorer la connaissance du mouvement relatif entre deux instants de prise de vue.

La seconde stratégie repose sur la reconstruction et l'appariement d'objets de liaison entre les images acquises au sol et à bord de l'avion. L'hypothèse est que la pose des caméras aériennes est précisément déterminée par aérotriangulation. Les objets de liaison choisis sont les marquages au sol (sur la chaussée), car leurs spécifications sont contraintes et leur détection puis leur reconstruction sont possibles dans les deux systèmes (TOURNAIRE et al. 2006*a* ; SOHEILIAN et al. 2007). Ensuite, un ajustement par faisceaux est réalisée et permet de faire la compensation du bloc complet et d'obtenir ainsi les paramètres de prise de vue (voir TOURNAIRE et al. 2006*b*). Cel permet d'améliorer la pose absolue des images.

La chaîne décrite ci-dessus n'est pas encore opérationnelle dans l'enchaînement des étapes pour géoréférencer les données de cette thèse. Ainsi les images utilisées dans les travaux présentés ici ont été géoréférencées manuellement, à l'aide de points d'appui mesurés sur le terrain avec une grande précision (inférieure au millimètre). La figure 1.9 montre un ensemble de vues géoréférencées placées dans un modèle 3D des bâtiments de la zone.

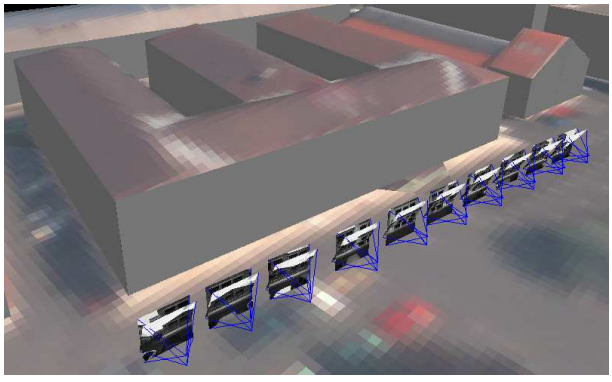


FIG. 1.9 – Vues géoréférencées acquises par la base verticale du véhicule Stéréopolis replacées dans le modèle 3D

Le géoréférencement une fois terminé, l'étape suivante est la reconstruction des façades de bâtiments

à partir des données collectées.

1.3 Etat de l'art en reconstruction de bâtiments à partir de données terrestres

Qu'elles disposent ou non d'un système de cartographie mobile pour automatiser la collecte de données, de nombreuses équipes se sont penchées sur le problème de reconstruction 3D de bâtiments à partir de données terrestres. En effet, le consensus a émergé très tôt sur la nécessité d'améliorer la description géométrique et les textures des modèles reconstruits à partir de données satellitaires ou aériennes. Nous avons déjà évoqué (paragraphe 1.2.1) la possibilité d'utiliser conjointement des données laser et images acquises depuis le sol. Ces techniques permettent d'obtenir d'excellents résultats, notament ceux obtenus par l'équipe de Berkeley (FRÜH et ZAKHOR 2004) ou encore STAMOS et ALLEN 2002.

D'autres auteurs se sont intéressés à la reconstruction de bâtiments uniquement à partir d'images, acquises au niveau du sol. Les applications architecturales ont guidé certains travaux. Dans DEBEVEC et al. 1996, les auteurs mettent en place un système quasi-automatique de modélisation d'éléments architecturaux. L'utilisateur n'a besoin de ne rentrer qu'un nombre réduit de correspondances entre primitives, ainsi que la structure de la scène, vue comme un assemblage de blocs élémentaires.

Une autre approche architecturale est proposée dans DICK et al. 2004. Il s'agit là encore d'un approche par modèles d'éléments architecturaux, cette fois entièrement automatique. A partir de primitives (points d'intérêt et segments) appariées automatiquement et de plans passant par ces données, il s'agit d'optimiser la reconaissance de modèles (facettes planes, colonnes, fenêtres,...) par une méthode de Monté-Carlo par chaînes de Markov. La figure 1.10 montre quelques résultats de ces travaux.

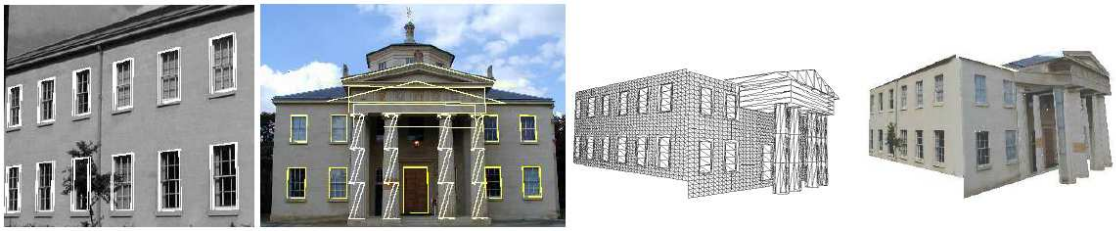


FIG. 1.10 – Résultats des algorithmes développés dans DICK et al. 2004

D'autres travaux sont guidés par la modélisation de scènes urbaines terrestres. Dans WERNER et ZISSERMAN 2002, les auteurs proposent de modéliser en premier lieu les façades planes, et ensuite d'enrichir le modèle avec divers éléments tels que les fenêtres, les mansardes... Pour trouver les plans présents dans la scène, les points de fuite sont utilisés pour trouver les directions principales de la scènes. Ensuite, un balayage de plan dans ces directions couplé avec un calcul de corrélation permettent de trouver la position des plans : elle correspond à un maximum de corrélation calculée sur un ensemble de points d'intérêts. Un exemple de résultat obtenu avant et après l'ajout d'élément sur les façades est montré figure 1.11.

FIG. 1.11 – Résultats des algorithmes développés dans WERNER et ZISSERMAN 2002

D'autres auteurs (par exemple BAUER et al. 2002) utilisent les mêmes techniques de reconstruction : balayage de plans dirigé par les directions des points de fuite pour obtenir les plans principaux de la scène, puis détections d'éléments plus petits (comme les fenêtres) pour enrichir la modélisation (SCHINDLER et BAUER 2003).

Les travaux menés au sein du projet Urbanscape (POLLEFEYS et al. 2008) se rapprochent davantage du but de nos propres recherches, tant du point de vue des données (seulement images) que de celui de l'objectif de fournir un modèle reconstruit des façades de bâtiments. Après une première étape de détection des plans principaux de la scène (façades, sol,...), un calcul rapide des maxima de corrélation est lancé par une méthode de balayage de plan pouvant être effectué par la carte graphique. Le balayage de plans est fait dans toutes les directions détectées. Enfin, une étape de fusion des cartes de profondeurs obtenue pour chaque image maîtresse est ajoutée, pour laquelle deux algorithmes sont proposés. Le

premier estime pour une vue référence la profondeur d'un pixel à l'aide de toutes les autres cartes de profondeurs issues des images en recouvrement. Le second prend en compte une mesure de confiance de la profondeur du pixel dans les cartes.

1.4 Bilan

En résumé, parmi les nombreux systèmes d'acquisition de données terrestres, deux grands groupes émergent : ceux qui disposent d'une plateforme couplée laser/image et ceux qui ne disposent que d'image. Le projet Stéréopolis a fait le choix pour l'équipement du véhicule d'utiliser uniquement deux caméras formant une base verticale et parfaitement synchronisées. Il s'agit d'un système multi-stéréoscopique ; pour chacune des bases stéréoscopiques, nous avons la certitude que la scène est identique dans chacune des images alors qu'elle peut changer (déplacements de véhicules ou de piétons) entre deux instants de prise de vue. La méthode de reconstruction que nous proposons tentera d'exploiter cette propriété en considérant dans un premier temps les bases prises indépendamment, et ensuite l'ensemble des images mises en jeu. Les directions des bases (verticales sur le véhicule et horizontales du fait du déplacement) sont très différentes, configuration favorable pour la reconstruction avec des images multiples.

D'autre part, plusieurs méthodes de reconstruction de bâtiments à partir d'images terrestres existent et ont été présentées dans la partie précédente. Elles sont soit basées sur une modélisation en éléments simples (facettes planes, fenêtres, autres éléments architecturaux), soit davantage sur une description topographique des scènes visualisées : c'est dans cette direction que s'inscrivent les objectifs de ces travaux. Le prochain chapitre expose notre solution de reconstruction, basée sur un appariement dense d'images.

CHAPITRE 2

APPARIEMENT DENSE D'IMAGES POUR LA RECONSTRUCTION 3D

L'objectif de ce travail est de proposer une solution à la problématique suivante : comment, à partir d'un ensemble d'images terrestres d'une scène, dont on connaît tous les paramètres de prise de vue – intrinsèques et extrinsèques – parvenir à reconstruire cette scène en trois dimensions. Dans ce chapitre, nous présentons notre méthode d'appariement dense d'images en stéréoscopie et en multi stéréoscopie.

La figure 2.1 montre un ensemble d'images en recouvrement multiple prises par le véhicule Stéréopolis.

De nombreux travaux ont déjà été menés sur cette question, qui est au centre de la vision par ordinateur. Nous investiguons dans un premier temps le cas stéréoscopique (partie 2.1), c'est à dire avec deux images. Ensuite, nous présentons dans la section 2.2 une technique d'optimisation de surface puis nous montrons comment les étapes précédentes se combinent dans un processus multirésolution (section 2.3). Enfin, nous étendrons ce travail au cas d'images multiples (partie 2.4).

2.1 Appariement dense de deux images

Comme nous l'avons déjà mentionné au chapitre précédent, dans le cas du projet Stéréopolis les images des façades de bâtiment proviennent de deux caméras, parfaitement synchronisées, formant une base stéréoscopique verticale. Cette base est relativement courte (environ 80 *cm*) ce qui est un critère très favorable pour la ressemblance des images, ainsi que la limitation des parties vues dans une image et cachées dans l'autre. Le recouvrement horizontal, obtenu entre un instant de prise de vue et le suivant, est caractérisé par une base plus longue, de l'ordre de 5 *m* pour permettre un recouvrement horizontal de

FIG. 2.1 – Images prises par le véhicule en multi-recouvrement sur une portion de façade. Photothèque de l'IGN, Saint-Mandé.

l'ordre de 60 %.

Dans un premier temps, nous exploitons les deux vues simultanées en stéréoscopie verticale, pour reconstruire une scène que l'on sait statique.

2.1.1 Etat de l'art

Indépendamment du contexte de reconstruction de bâtiments, la reconstruction tridimensionnelle de toute scène à partir d'images est un problème majeur de vision par ordinateur et déjà largement investigué. Diverses techniques ont été mises en oeuvre et peuvent dépendre du type de scène observée. Dans le cas de deux images, Scharstein et Szeliski (SCHARSTEIN et SZELISKI 2002) ont établi une étude très complète qui compare les algorithmes de stéréo dense à deux images.

2.1.2 Mesures de similarité

A la base de l'appariement dense de deux images se trouve la recherche d'homologues entre les images. Etant donné un pixel dans une image, on appelle homologue le pixel de l'autre image qui cor-

respond à l'image du même point de la scène. Pour résoudre ce problème d'appariement, il est utile de dégager les contraintes qu'il doit respecter.

La *contrainte d'unicité*, introduite par MARR et POGGIO 1977, rappelle qu'un point de la scène se projette dans au plus un point de l'image. Ainsi, l'homologue d'un point image dans une autre image est unique (mais l'existence n'est pas garantie). Les mêmes auteurs mettent également en avant une *contrainte de continuité* sur la surface de l'objet à reconstruire, qui implique une continuité de la carte de profondeur recherchée. Cette contrainte n'est cependant pas vérifiée de manière globale, mais locale.

La *contrainte d'ordre* indique que lorsque l'on se déplace dans une direction sur l'une des images, les objets sont rencontrés dans le même ordre lors du déplacement dans l'autre image. Il existe cependant des cas d'inversion.

D'autres contraintes sont utilisées pour la recherche d'appariement, il s'agit de la *contrainte épipolaire* et de la *contrainte de ressemblance*. L'utilisation de ces contraintes, notamment la ressemblance locale d'éléments homologues est détaillée dans la suite.

Localement, autour de ces deux pixels homologues, les images ont une similarité forte car elles correspondent à la même zone de la scène : c'est la contrainte de ressemblance. Afin de retrouver automatiquement des paires homologues, de nombreuses mesures de similarité ont été proposées. Une revue très détaillée de ces mesures a été réalisé dans la thèse de Chambon (CHAMBON 2005, chap. 3). Nous mentionnons ici les plus utilisées.

Corrélation

La corrélation est un indicateur statistique qui mesure la similarité de deux échantillons. Ainsi, deux vecteurs de paramètres identiques auront la similarité maximale, et donc le score maximal de corrélation, et deux vecteurs opposés auront à l'inverse le score minimal. En traitement d'image, le vecteur de paramètres utilisé en un pixel donné est le vecteur des radiométries de tous les pixels contenus dans un voisinage du pixel de référence. Pour un pixel de l'image i, ce vecteur est noté $\mathbf{V_i}$. $\overline{\mathbf{V_i}}$ désigne la moyenne des valeurs du vecteur $\mathbf{V_i}$. Plusieurs mesures de corrélation croisée ont été proposées.

Corrélation croisée normée

$$NCC(\mathbf{V_i}, \mathbf{V_j}) = \frac{\mathbf{V_i}.\mathbf{V_j}}{\|\mathbf{V_i}\|\|\mathbf{V_j}\|} \tag{2.1}$$

Les valeurs sont comprises dans l'intervalle $[0;1]$.

Corrélation croisée normée et centrée Elle est souvent notée *ZNCC* pour *Zero mean Normalized Cross-Correlation*.

$$ZNCC(\mathbf{V_i}, \mathbf{V_j}) = \frac{(\mathbf{V_i} - \overline{\mathbf{V_i}}).(\mathbf{V_j} - \overline{\mathbf{V_j}})}{\|\mathbf{V_i} - \overline{\mathbf{V_i}}\|\|\mathbf{V_j} - \overline{\mathbf{V_j}}\|} \tag{2.2}$$

Elle prend ses valeurs dans l'intervalle $[-1;1]$, et est une des mesures les plus utilisées. Elle présente en effet l'avantage d'être invariante par affinité sur les niveaux de gris de l'image.

2.1.3 Mise en place par corrélation

Le principe est le suivant : étant donné un pixel de référence dans une image, il s'agit de trouver dans une autre image, le pixel dont le voisinage est le plus semblable au sens du meilleur score de corrélation à celui du pixel de référence. Ceci fournit un appariement de deux pixels dans deux images, qui représentent le même détail de la scène. Cette mise en correspondance faite, la position 3D du détail de la scène est obtenue par l'intersection des deux rayons lumineux issus respectivement des deux pixels appariés. Cependant, il est inutile de parcourir toute l'image pour trouver le meilleur appariement, car l'homologue d'un pixel ne peut se trouver que sur la droite épipolaire correspondante, c'est à dire la projection du rayon lumineux issu du pixel sur l'autre image. L'espace de recherche est donc restreint à une dimension pour chaque paire d'image, comme le montre la figure 2.2.

La recherche de l'homologue d'un pixel de coordonnées (c,l) est pilotée par l'équation suivante :

$$\hat{p}(c,l) = \underset{p_{min}<p<p_{max}}{\text{argmax}} \; cor(c,l,p) \tag{2.3}$$

où p_{min} et p_{max} désignent respectivement la profondeur minimale et maximale, et *cor* la mesure de corrélation choisie. Nous prenons la mesure de corrélation normée centrée pour ses invariances, ce qui donne

$$cor(c,l,p) = ZNCC(\mathbf{V_i}, \mathbf{V_j}) \tag{2.4}$$

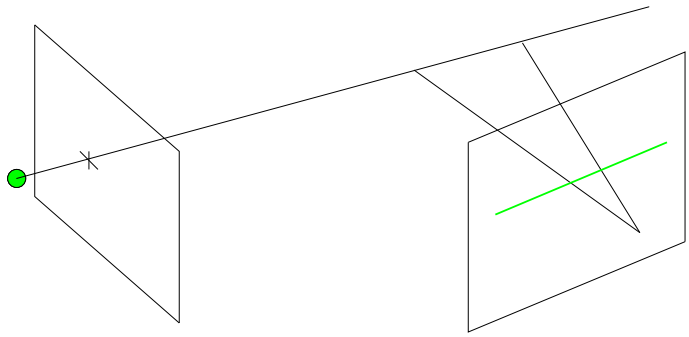


FIG. 2.2 – Géométrie épipolaire

avec $\mathbf{V_i}$ le vecteur des radiométries au voisinage du pixel de référence, $\mathbf{V_j}$ le vecteur des radiométries au voisinage du candidat homologue.

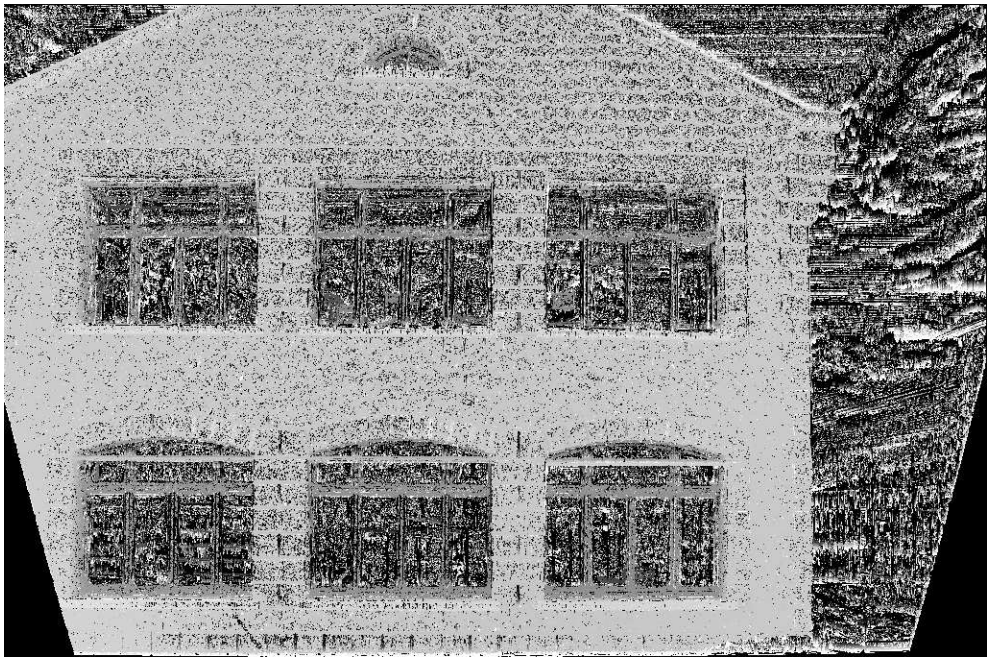
2.1.4 Mise en oeuvre et résultats

Nous disposons d'images avec un très faible bruit, et de plus nous cherchons à obtenir les détails les plus fins possibles. Nous faisons donc le choix d'utiliser comme voisinage pour le calcul de corrélation des fenêtres carrées, de petite taille (typiquement 3×3 ou 5×5), centrées sur le pixel. La délocalisation due aux déformations perspectives est ainsi minimisée.

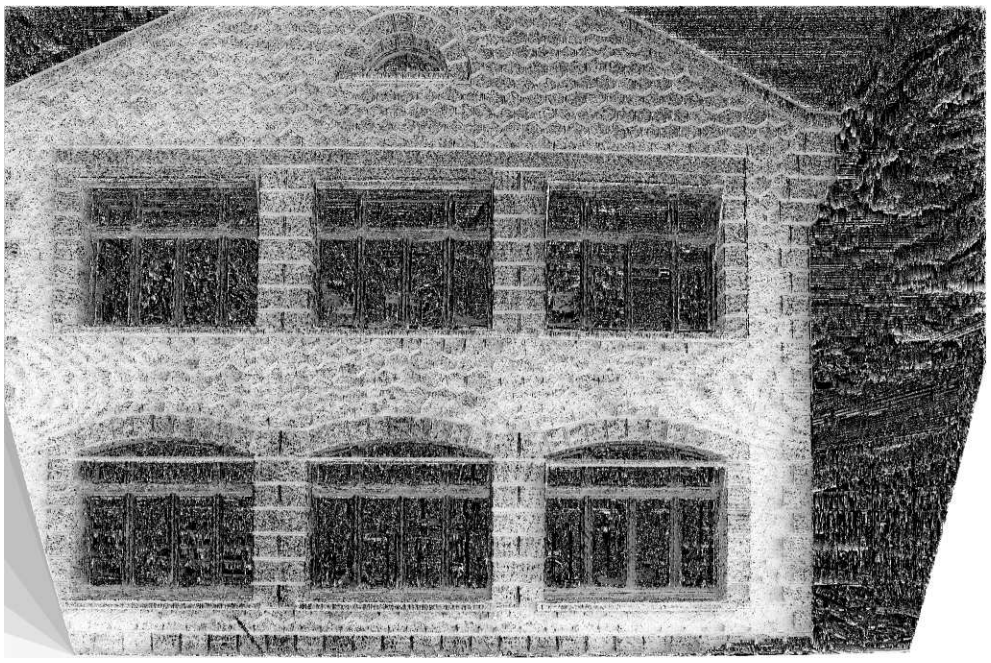
La figure 2.3 montre des résultats obtenus par l'algorithme, à partir de la paire stéréoscopique d'images présentée figure 2.4.

Dans la partie en recouvrement entre les deux images, ces premiers résultats mettent en évidence un fort bruit de reconstruction. En effet, si pour la majorité des pixels la profondeur obtenue est correcte, 10 à 25 % des pixels ont une profondeur erronée de plus de 10 *cm* (mesures effectuées sur les zones quasi-planes et texturées, donc plutôt favorables à une bonne reconstruction).

Nous allons donc dans la suite proposer une amélioration de ces résultats par régularisation de la surface obtenue.



(a) Carte de profondeur en niveau de gris : les faibles profondeurs sont représentées en clair, les grandes profondeurs en foncé.



(b) Même carte de profondeur en vue ombrée.

FIG. 2.3 – Résultats obtenus avec deux images

FIG. 2.4 – Couple stéréoscopique d'images

2.2 Régularisation de la surface

Jusqu'à présent, le choix a été fait de garder le meilleur score pour sélectionner la profondeur p à (c,l) fixés. Une des conséquences classiques de cette méthode est qu'elle génère souvent un résultat précis mais non fiable. Or, si la précision est un de nos objectifs, nous cherchons également un résultat robuste. Une régularisation des appariements est donc nécessaire. Pour ce faire, plusieurs approches de flot maximal ou de coupe de graphes ont été proposées (par exemple VOGIATZIS et al. 2007 ; KOLMOGOROV et ZABIH 2002 ; ROY et COX 1998). Le processus d'optimisation globale par flot maximal développé par Roy et Cox ROY et COX 1998 réalise le traditionnel compromis entre les termes d'attache aux données (ici les scores de corrélation calculés pour tous les pixels (c,l) et à toutes les profondeurs) et de régularisation. Le paramètre qui commande cette régularisation sera choisi suffisament faible, de manière à préserver les discontinuités réelles des façades.

Formulation du problème par flot maximal

Le problème de flot de graphe est un problème maintenant classique en algorithmique. Il peut être formulé comme un problème d'écoulement de fluide dans un réseau de canalisations. De manière intuitive, considérons un réseau de canalisations, alimenté par une source de fluide de débit infini, et se

déversant dans un puits de contenance également infinie. On cherche le flot maximal de fluide que l'on peut faire passer par le réseau pour aller de la source au puits. Comme ceux-ci ne sont pas des facteurs limitants, le flot maximal est uniquement déterminé par la capacité du réseau. En augmentant progressivement la quantité de fluide déversé par la source, on fini par atteindre le flot maximal. Dans cette situation, certaines canalisations sont saturées, et le flot maximal est alors égal à la somme de leurs capacités. Toute quantité supplémentaire de fluide déversé par la source ne peut être transporté par le réseau jusqu'au puits. L'ensemble des canalisations saturées est de capacité minimale, et leur somme renvoie la valeur du flot maximal.

De manière plus théorique, le réseau est formellement représenté par un graphe pondéré

$$\mathscr{G} =< \nu, \mathscr{E} >$$

où ν représente l'ensemble des noeuds et $\mathscr{E} \subset \nu^2$ l'ensemble des arêtes orientées, et deux noeuds terminaux, s et t, représentant la source et le puits. Pour un noeud donné x, on définit $\mathscr{E}_e(x)$ l'ensemble des arêtes entrantes et $\mathscr{E}_s(x)$ l'ensemble des arêtes sortantes. On définit une fonction capacité qui à une arête ε associe un réel positif noté $|\varepsilon|$ et une fonction flot qui à une arête associe un réel $Flot(\varepsilon)$. La fonction flot est dite valide si

$$\forall \varepsilon \in \mathscr{E}, Flot(\varepsilon) \leqslant |\varepsilon| \tag{2.5}$$

$$\forall x \in \nu \backslash \{s,t\}, \sum_{\varepsilon_e \in \mathscr{E}_e(x)} Flot(\varepsilon_e) = \sum_{\varepsilon_s \in \mathscr{E}_s(x)} Flot(\varepsilon_s) \tag{2.6}$$

Le flot du graphe est alors :

$$Flot(\mathscr{G}) = \sum_{\varepsilon \in \mathscr{E}_s(s)} Flot(\varepsilon) \tag{2.7}$$

On peut également définir une coupe de ce graphe, $\mathscr{C}$, qui partitionne le graphe en deux sous-ensembles, l'un contenant le puits, l'autre la source. La capacité de la coupe est définie comme la somme des capacités des arrêtes coupées :

$$|\mathscr{C}| = \sum_{(x,y) \in \mathscr{C}} |\varepsilon(x,y)| \tag{2.8}$$

Pour un graphe donné, la recherche du flot maximal est équivalente à la recherche d'une coupe

minimale (FORD et FULKERSON 1962) et on a la relation suivante :

$$\max_{Flot} Flot(\mathscr{G}) = \min_{\mathscr{C}} |\mathscr{C}| \tag{2.9}$$

Cette relation, connue sous le nom de théorème mincut-maxflow, permet de résoudre un problème de minimisation d'énergie en obtenant le minimum global de la fonction. Pour cela, il faut que la fonction d'énergie puisse être écrite sous forme d'un graphe. Le problème de minimisation est équivalent à un problème de recherche de flot maximal pour lequel des algorithmes ont été proposés.

2.2.1 Etat de l'art

La méthode développée par Roy et Cox (ROY et COX 1998) propose également de calculer une carte dense de profondeur, avec une régularisation paramétrable, par une technique de coupe minimale de graphe. Le paragraphe suivant développe cet algorithme.

2.2.2 Algorithme de Roy et Cox

L'algorithme de flot maximal introduit par Roy et Cox (ROY et COX 1998) propose de construire le graphe de manière classique pour un problème de flot maximal en image. Le graphe est un maillage en trois dimensions, et les sommets ont pour coordonnées (c,l,d) où c et l sont les coordonnées (colonne,ligne) de l'image, et d la disparité. On ajoute la source, reliée à tous les sommets de disparité minimale, et le puits relié à tous les sommets de disparité maximale. Les autres sommets sont reliés en 6-connexité (voir figure 2.5).

La capacité d'une arrête liant deux noeuds u et v est définie par :

$$|\varepsilon_{(u,v)}| = \begin{cases} 0 & \text{si les noeuds } u \text{ et } v \text{ ne sont pas voisins} \\ \infty & \text{si } u = s \text{ ou } v = t \\ |\varepsilon_{(u,v)}|_{disp} & \text{si } d_u = d_v \\ |\varepsilon_{(u,v)}|_{occ} & \text{si } d_u \neq d_v \end{cases} \tag{2.10}$$

avec $|\varepsilon_{(u,v)}|_{disp} = \frac{cout(u)+cout(v)}{2}$, capacité d'une arrête de disparité (joignant deux noeud de disparité

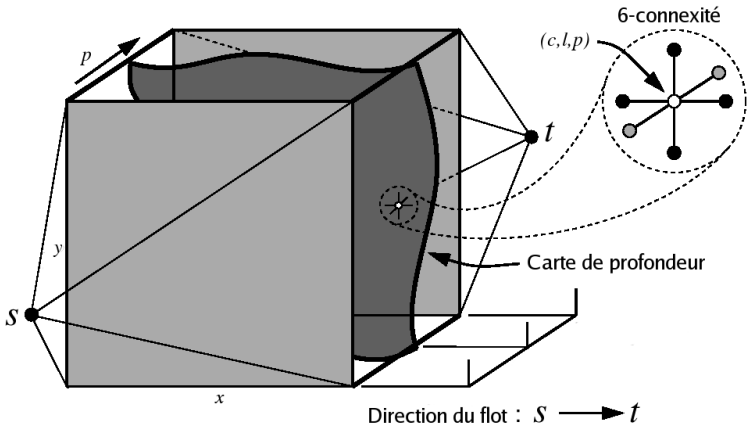


FIG. 2.5 – Carte de profondeur obtenue par l'algorithme de Roy et Cox. D'après ROY et COX 1998

différente) et pour une arrête d'occultation : $|\varepsilon_{(u,v)}|_{occ} = k|\varepsilon_{(u,v)}|_{disp}$, $k \in [0,\infty]$. Le coût d'un noeud $u = (c,l,d)$ est défini comme la variance des intensités des pixels reprojetés sur les N images :

$$cout(u) = \frac{1}{N}\sum_{i=0} NI_i(u) \tag{2.11}$$

Le paramètre k règle le lissage de la solution obtenue. Plus k sera grand, plus la solution sera lisse car la pénalisation des changements de disparité augmente avec k.

La carte de disparité s'obtient directement à partir de la coupe minimale, elle même calculée par un algorithme de flot maximal (dans ROY et COX 1998, l'algorithme utilisé est "*preflow-push lift-to-front*"). L'adaptation à notre contexte se réalise en construisant le graphe de Roy et Cox à partir de l'espace en trois dimensions discrétisé (celui de la géométrie image), chaque voxel devenant sommet du graphe, avec ses coordonnées (c,l,p) et sa pondération dérive directement du score de corrélation précédemment obtenu :

$$cout(u) = 1 - max(0, cor_i(c,l,p)$$

2.2.3 Mise en oeuvre et résultats

L'étape de régularisation nécessite de connaitre l'ensemble des pondérations, une forte charge mémoire est donc nécessaire. Par exemple, pour une image de taille 4096×4096, en recouvrement à 80%,

avec 100 niveaux de discrétisation en profondeur, avec des pondérations codées sur 4 octets, il serait nécessaire de disposer d'une place mémoire de $5Go$. Nous procédons à une résolution du problème par morceau, avec un découpage en dalles de l'image initiale. La taille des dalles est choisie de manière adaptative en fonction de la taille de l'image et de l'occupation mémoire maximale acceptée.

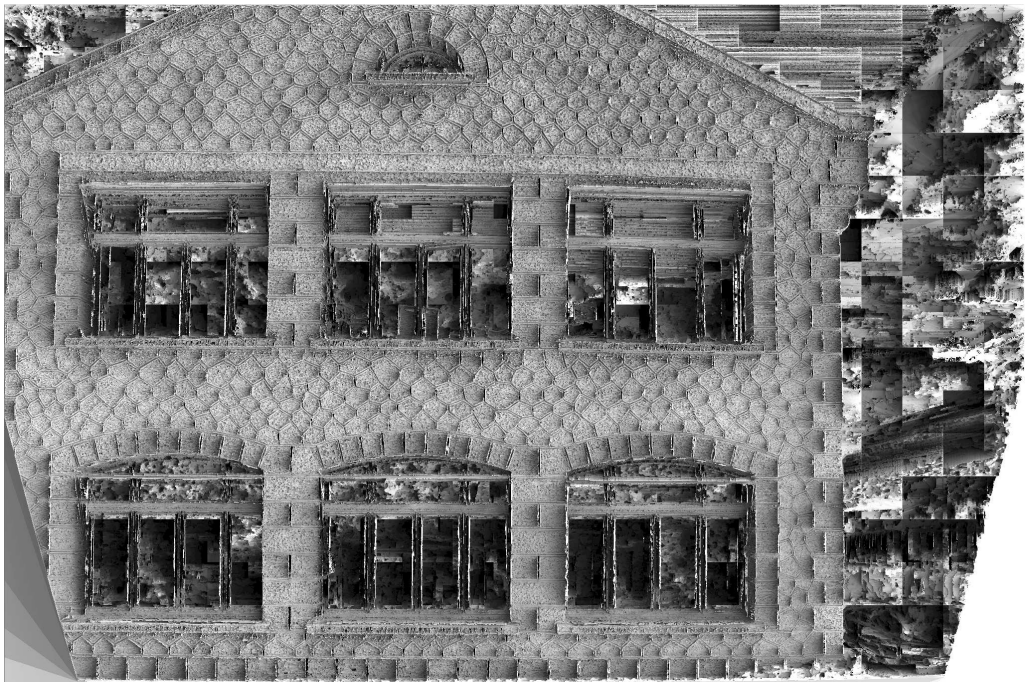
La figure 2.6 montre l'effet obtenu par la régularisation pour différentes valeurs du paramètre k. Le découpage en dalles est visible sur les résultats du fait de la régularisation. En effet, cette dernière est indépendante sur chacune des dalles. Ainsi, pour les zones peu texturées ou qui sortent de la zone de recherche en profondeur les termes d'attache aux données deviennent négligeables devant les termes de régularisation. En revanche, il est important de noter que dans les autres cas, les raccords entre dalles ne présentent aucune discontinuité.

2.3 Multi-résolution

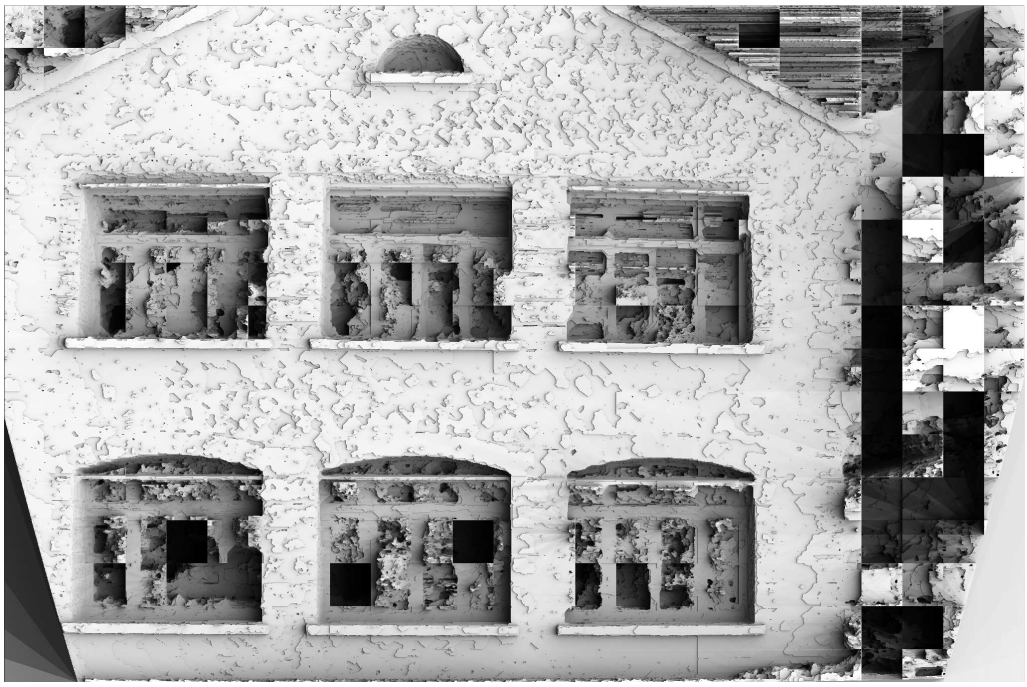
L'ensemble des opérations décrites nécessite un temps de calcul important : calcul des scores de corrélation multi-images pour tous les voxels d'une part et optimisation par l'algorithme de Roy et Cox d'autre part représentent la quasi totalité du temps machine nécessaire, soit plusieurs heures pour une image de taille 4096×4096 pixels et un pas de discrétisation cohérent dans la direction de profondeur de la scène. Il est exclu d'utiliser cette méthode avec de telles performances sur un nombre important d'images. En pratique, le temps passé est inacceptable si l'on cherche à reconstruire plus d'une façade. Afin de réduire le temps de calcul et également la mémoire utilisée, une stratégie en multi-résolution a été implémentée. L'idée est de partir d'une discrétisation très grossière de la scène (sous-échantillonnage de l'image, pas de discrétisation grossiers), de reconstruire la surface à cette résolution, puis de se servir de cette surface grossière pour réduire l'espace de recherche de la résolution supérieure.

2.3.1 Utilisation de la multi-résolution

L'algorithme proposé dans les paragraphes précédents a été implémenté pour traiter une pyramide d'images de la moins résolue à l'image originale. Après génération de la pyramide d'images, la première passe de l'algorithme se fait à la résolution la plus basse. Le résultat est une surface très approximative

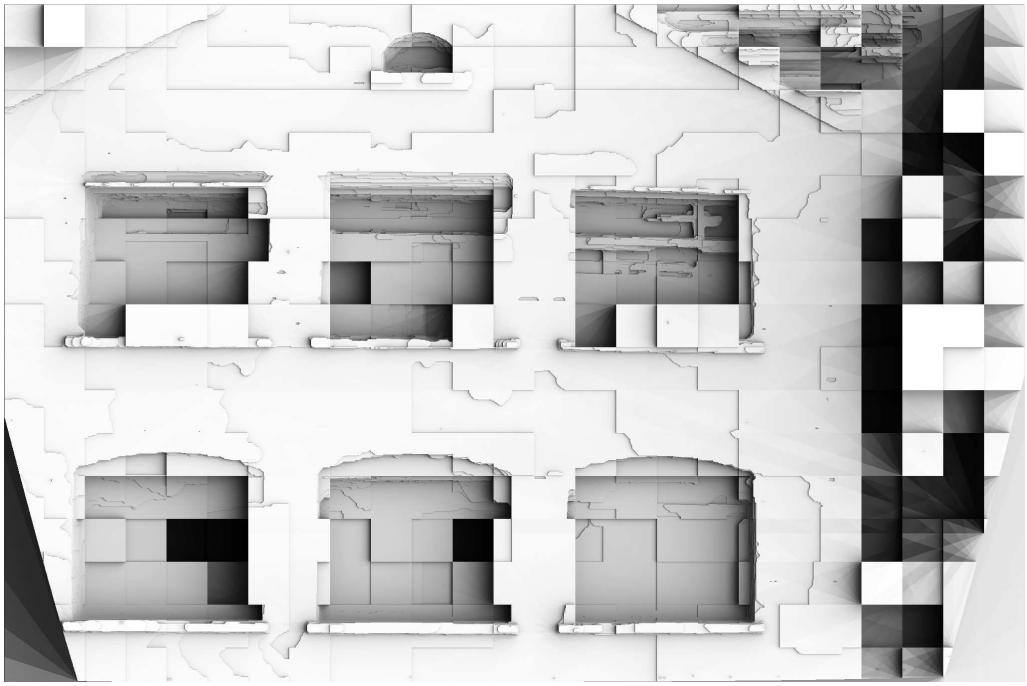


(a) Coefficient $k = 0,02$



(b) Coefficient $k = 0,04$

FIG. 2.6 – Résultats de régularisation de surface pour différentes valeurs du coefficient k.



(c) Coefficient $k = 0,2$

FIG. 2.6 – Résultats de régularisation de surface pour différentes valeurs du coefficient k.

et grossière. Elle sert de surface d'initialisation pour la recherche d'une meilleure surface à la résolution supérieure. L'espace de recherche est généré par dilatation dans le plan de la façade (directions colonne et ligne), et dans la direction perpendiculaire (profondeur). L'enveloppe de la surface délimite alors l'espace de recherche à la résolution supérieure (voir figure 2.7).

Le pas de discrétisation en profondeur est divisé par le facteur de sous-échantillonnage à chaque changement de résolution afin d'améliorer la précision de la recherche en profondeur.

Le choix des paramètres de dilatation, que ce soit dans le plan de la façade ou dans la direction orthogonale, est réalisé pour tenir compte des deux impératifs contradictoires suivants : il convient d'une part de dilater suffisamment afin de pallier une éventuelle déficience de l'algorithme à la résolution inférieure (à faible résolution, il est possible que des reliefs d'emprise peu étendue soit ignorés par l'algorithme), d'autre part de conserver un espace de recherche de petite taille pour ne pas augmenter les temps de traitement. Le compromis que nous faisons est de prendre une dilatation de l'ordre de 5 fois le pas local en profondeur.

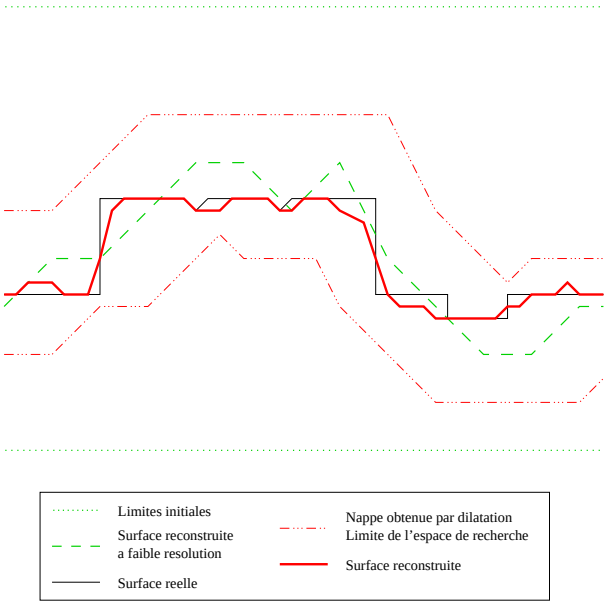


FIG. 2.7 – Dilatation à partir d'une surface pour délimiter l'espace de recherche à la résolution supérieure

En pratique, le processus devient donc tel que décrit par l'algorithme 2.1.

2.3.2 Apports de la multi-résolution

En ce qui concerne le temps de calcul, en fixant tous les autres paramètres, nous observons en moyenne un gain de temps de l'ordre d'un facteur 7 sur la même machine, ainsi qu'une division par 2 de la place mémoire requise lors de l'étape d'optimisation. A titre d'exemple, le tableau 2.1 donne les performances de l'algorithme pour un point de vue. Cette amélioration a pour conséquence directe que le traitement de l'optimisation, qui nécessitait auparavant un découpage de l'image en de multiples sous-images peut maintenant se faire avec un nombre réduit de sous-images. La cohérence globale de la surface en est améliorée. Le gain en temps de calcul procure l'avantage immédiat de pouvoir traiter plus rapidement nos données. Il permet également de trouver un meilleur compromis entre temps de calcul et précision géométrique du résultat, puisque celle-ci dépend directement du pas de discrétisation choisi en profondeur.

L'autre limitation supprimée par la multi-résolution est l'obligation que nous avions d'avoir un espace de recherche initial limité. Celui-ci pouvait provenir d'un modèle 3D grossier du bâtiment ou encore

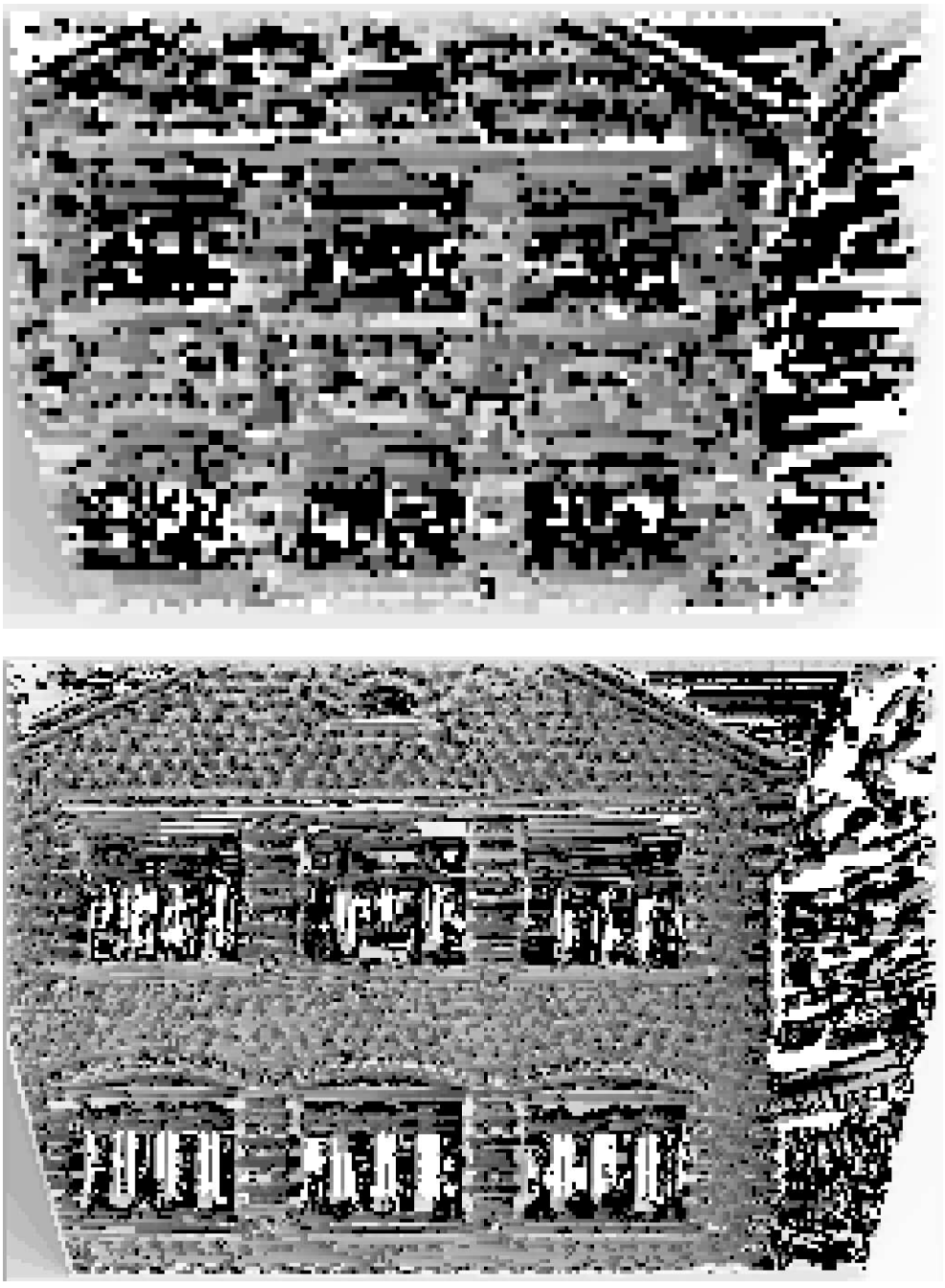


FIG. 2.8 – Surface à plusieurs résolutions successives en vue ombrée. La résolution augmente de haut en bas. L'image maîtresse se trouve figure 2.1, en bas à droite.

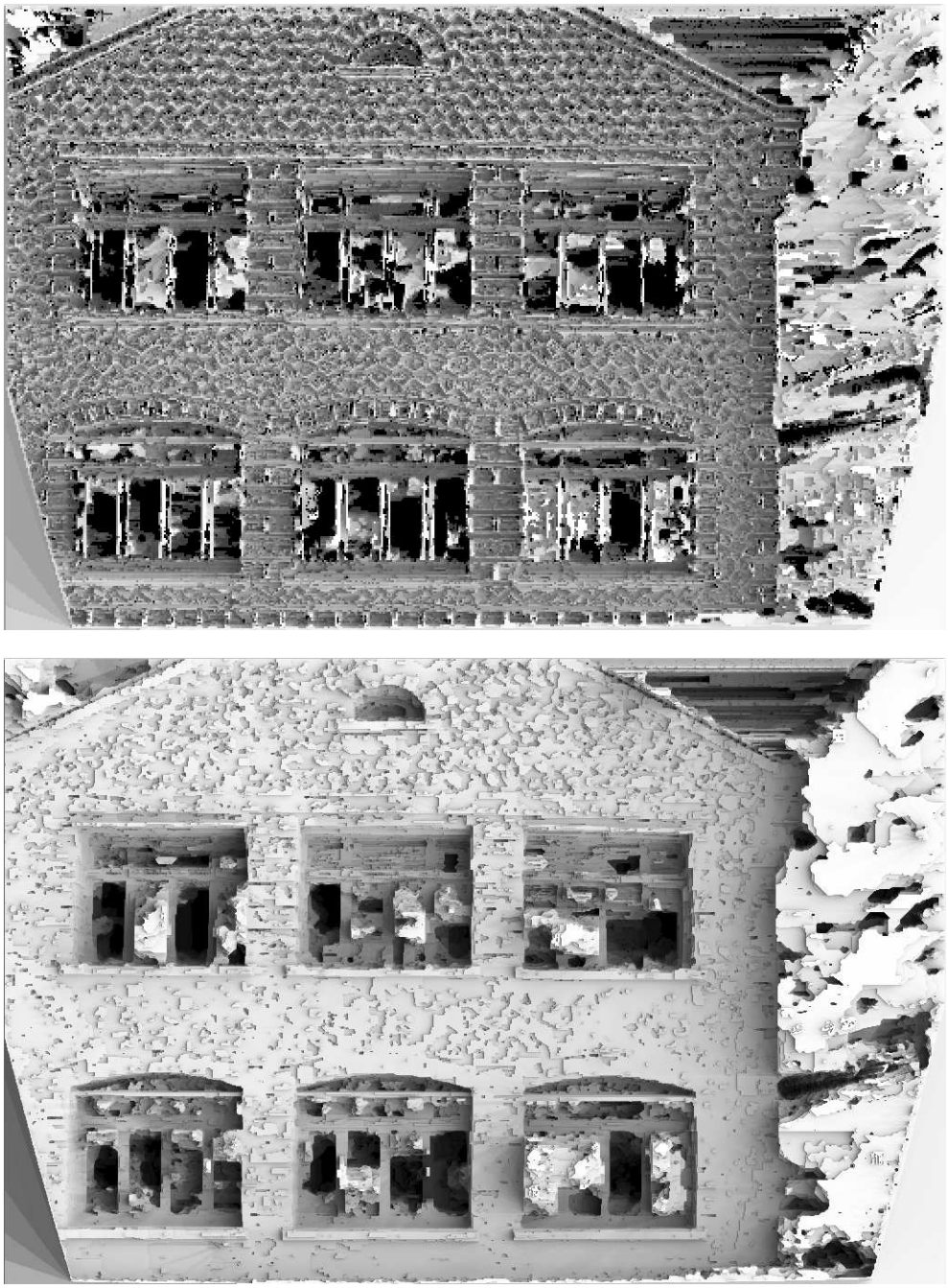


FIG. 2.8 – Surface à plusieurs résolutions successives. La résolution augmente de haut en bas. Colonne de gauche : profondeur en niveaux de gris. Colonne de droite : même surface en vue ombrée. Dernière ligne : résultat après filtrage. L'image de départ se trouve figure 2.1, en bas à droite.

Algorithme 2.1 : Stratégie d'appariement dense en multi-résolution à deux images

Données : I_1, I_2,deux images d'une base verticale
Sorties : Carte de profondeur pour I_1

pour tous les $I \in \{I_1, I_2\}$ **faire**
  Création de la pyramide d'images sous-échantillonnées jusqu'à la résolution resol_min
fin
pour resol = resol_min **à** resol_max **faire**
  si resol $\neq$ resol_min **alors**
    Calcul de la zone de recherche par dilatation de nappe_courante
  fin
  Découpage de l'espace image en dalles;
  pour tous les *dalles* **faire**
    pour tous les *voxels* $v \in$ *zone de recherche* **faire**
      Calcul du score de corrélation $cor_{I_1,I_2}(v)$;
    fin
    Optimisation;
  fin
  nappe_courante $\leftarrow$ `assemblage_dalles()` ;
  resol = resol + 1 ;
fin

Multi-résolution	Temps de calcul	Mémoire
Sans	1h29'51"	160 Mo
Avec	11'28"	76 Mo
Facteur de gain	7,8	2,1

TAB. 2.1 – Performances de l'algorithme avec ou sans l'utilisation de la multi-résolution pour un point de vue (figure 2.1, en bas à droite). Machine utilisée : P4 3GHz

d'un plan 2D de type cadastre. Puisque la première étape de recherche d'une surface possible se fait à très basse résolution, il ne coûte quasiment rien en temps de parcourir une vaste profondeur à la recherche d'un bâtiment.

En revanche, passer en multi-résolution implique de choisir le paramètre de dilatation des nappes en profondeur. Ce paramètre est sensible pour le comportement de l'algorithme puisque il conditionne les parties de l'espace de la scène explorées ou non à la résolution suivante.

2.4 Appariement dense d'images multiples

2.4.1 Etat de l'art : reconstruction à partir d'images multiples

Le passage de deux images à un nombre supérieur peut permettre d'améliorer les résultats. En effet, à deux images, la contrainte épipolaire est à une dimension. Si les centres optiques des caméras ne sont pas alignés, on ajoute alors une contrainte dans une autre direction, ce qui peut résoudre les ambiguïtés. Par ailleurs, l'augmentation du nombre d'observations relève le rapport signal à bruit, ce qui permet d'améliorer la robustesse des résultats. Une étude (SEITZ et al. 2006) propose une comparaison pour les algorithmes de reconstruction multi-images. Les auteurs réalisent une classification des méthodes selon six critères : le mode de représentation de la scène, le type de mesure de similarité inter-images, le modèle de visibilité, l'*a priori* sur la forme, l'algorithme de reconstruction proprement dit, et enfin l'initialisation requise. Ils proposent une évaluation basée sur deux éléments : la précision (mesure de l'écart entre les points de la surface reconstruite et la vérité terrain) et la complétude de la reconstruction (de manière duale, mesure de l'écart entre les points de la vérité terrain et la surface reconstruite).

Space Carving

La méthode de "space carving" a été développée par KUTULAKOS et SEITZ 1999. Le principe de la méthode est de partir d'un volume englobant de l'objet à reconstruire, puis de supprimer les voxels n'appartenant pas à l'objet, de manière à obtenir le volume maximal (ainsi que les radiométries de voxels de surface) qui soit compatible avec l'ensemble des observations image. Un voxel de la surface du volume compatible avec l'ensemble des observations est appelé photo-consistant. L'algorithme de "space carving" est le suivant :

- initialisation du volume de départ, contenant l'objet ;
- pour chaque voxel v de la surface du volume, détermination de sa photo-consistance par reprojection dans toutes les images voyant de point jusqu'à ce qu'un voxel non photo-consistant soit trouvé
- si aucun voxel non photo-consistant n'est trouvé, fin de l'algorithme, sinon retour à l'étape précédente en enlevant v du volume.

De bons résultats sont obtenus par les auteurs, mais des limites existent : bruit important sur la reconstruction et caractère irréversible de la suppression de voxel pour les plus notables. La figure 2.9 donne un exemple de résutat obtenu.

FIG. 2.9 – Résultat de l'algorithme de space carving sur une statuette en bois obtenu avec 21 images en entrée

Voxel Coloring

Un autre algorithme, baptisé voxel coloring (SEITZ et DYER 1999) procède de la façon suivante : on colore itérativement les voxels d'un volume, de manière compatible avec les observations image. En détail, les voxels sont classés en "couches". L'ensemble des couches forme une partition de la scène, et les voxels appartenant à une couche n'occultent aucun voxel des couches inférieures pour toutes les caméras. De cette manière, l'algorithme fait une passe unique sur chacun des voxels, couche après couche, pour en déterminer la couleur. Certaines limitations sont soulevées par les auteurs : l'objet reconstruit n'est qu'une des possibilités parmi les formes compatibles. De plus, de nombreux artefacts apparaissent dan les régions hautes fréquences des images. Le résultat final est relativement bruité (voir figure 2.10).

Level Sets

La technique des ensembles de niveaux, plus connue sous son nom anglais de "Level Sets", a été popularisée ces dernières années, et s'est imposée comme une technique majeure de reconstruction 3D. Ces méthodes reposent sur l'écriture d'un principe variationnel qui doit être satisfait par les surfaces que l'on cherche à reconstruire, et qui peut intégrer des hypothèses faites sur la scène. Du principe variationnel sont tirées des équations aux dérivées partielles utilisées pour déformer les surfaces initiales jusqu'à

convergence. L'implémentation sous la forme level set donne de très bons résultats (voir KERIVEN et FAUGERAS 1998 ; PONS et al. 2005). La technique des level sets présente néanmoins l'inconvénient de ne pas bien se prêter à la reconstruction de forme polyédriques, mais est plus adaptée à des surface plus molles, partout dérivable, comme le buste sculpté reconstruit figure 2.11.

Reconstruction par estimation de cartes de profondeurs

De nombreux auteurs proposent de résoudre le problème de reconstruction à partir d'images multiples par le calcul des cartes de profondeur des images.

On peut citer par exemple GOESELE et al. 2006. Les auteurs adoptent une stratégie en deux étapes : la première assure le calcul des cartes de profondeur, en utilisant la mesure de corrélation normée centrée, la seconde reprend la méthode de fusion de maillage développée par CURLESS et LEVOY 1996.

KOLMOGOROV et ZABIH 2002 proposent également de fournir la carte de profondeur pour toutes les images. Ils formulent le problème comme la minimisation d'une énergie composée d'un terme d'attache aux données, d'un terme de régularisation et d'un terme de visibilité et réalisent la minimisation par coupe minimale de graphe.

Il est également possible de calculer les cartes de profondeur de l'ensemble des images par une formulation bayésienne. Dans GARGALLO et STURM 2005 et STRECHA et al. 2006, les auteurs proposent une optimisation par l'algorithme EM pour trouver les cartes de profondeur et de visibilité qui expliquent au mieux les images observées.

Dans NISTÉR 2004, l'auteur utilise une formulation bayésienne pour calculer les cartes de profondeur. L'optimisation est réalisée par coupe minimale de graphe. Les cartes de profondeurs issues des différents points de vue sont ensuite fusionnées par une méthode appelée "fusion médiane".

HERNÁNDEZ et SCHMITT 2004 utilisent une approche par contours déformables ("snakes") en 3D. Ils utilisent la mesure de corrélation comme terme d'attache aux données de leur snake, initialisé à l'enveloppe visuelle extraite dans un premier temps. Ils obtiennent les cartes de profondeur à l'intérieur des silhouettes de l'objet, avant de réaliser le maillage et la texturation de la forme reconstruite.

Les travaux de l'équipe du projet Urbanscape (voir section 1.2.1) portent également sur le calcul de cartes de profondeurs pour un ensemble d'images de références (AKBARZADEH et al. 2006). Les cartes

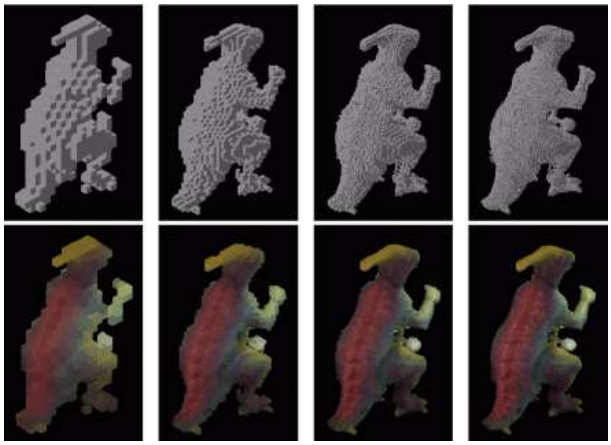


FIG. 2.10 – Résultat de l'algorithme de voxel coloring sur un dinosaure à partir de 21 images en entrée. Les résultats sont donnés pour différentes résolutions de grille, de 13920 à 7696922 voxels

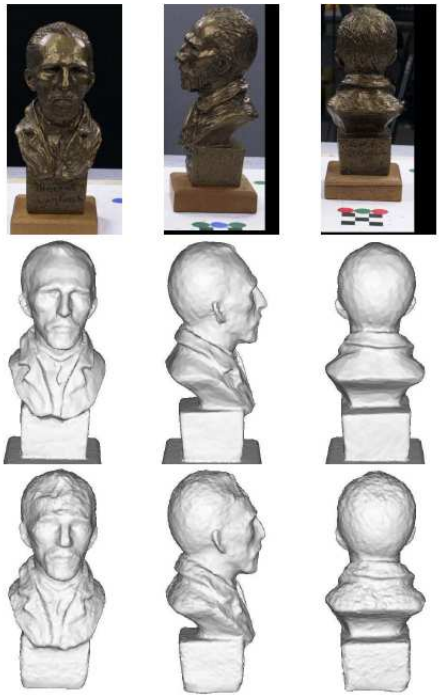


FIG. 2.11 – Résultat de l'algorithme de PONS et al. 2005 sur un buste sculpté, obtenu à partir de 24 images sources

de profondeurs sont déterminées par corrélation multi-images par balayage de plans dans les directions prinipales identifiées dans la scène (façade fronto-parallèle, façades fuyantes...) de manière très rapide en utilisant le processeur de la carte graphique. Ensuite, une étape de fusion est ajoutée pour obtenir une surface unique à partir des cartes de profondeurs.

Les travaux un peu plus anciens de Fua (FUA 1997) portaient sur la reconstruction basée sur une représentation en "particules" orientées. Ces particules voient leur position et leur orientation se modifier sous l'effet d'une force d'attache aux données (fonction de coût image) et d'interactions entre particules qui tendent à régulariser la surface. Cette approche convient mieux à des surfaces régulières, telles que les visages.

Les méthodes présentées donnent toutes de bons résultats lorsqu'elles sont utilisées dans les conditions pour lesquelles elles sont adaptées. Les méthodes de reconstruction par estimation de cartes de profondeurs semblent bien adaptées à notre contexte de reconstruction de façades qui sont généralement constituées d'éléments plans par morceaux en première approximation. La méthode développée dans la première partie de ce chapitre peut être adaptée au cas multi-stéréoscopique, ce que nous allons présenter maintenant.

2.4.2 Méthode de reconstruction proposée

Le passage à un ensemble d'images multiples en recouvrement peut se dérouler suivant deux scénarios : soit passer dans une géométrie guidée par la scène, que nous appelons "géométrie objet", soit rester dans une géométrie guidée par l'image ("géométrie image"). Nous présentons dans un premier temps chacune de ces deux variantes puis ensuite nous comparerons les résultats obtenus dans ces deux cas.

2.4.3 Géométrie objet

Nous reprenons la recherche du meilleur appariement au sens du score de corrélation vu dans le cas stéréoscopique, mais cette fois cela se fait avec des images multiples et en parcourant un axe du repère terrain lié aux objets de la scène : c'est la géométrie objet. Dans le cadre de cette thèse, les objets que nous cherchons à décrire sont des façades, il est donc possible de définir un repère tel qu'un des axes soit perpendiculaire au plan directeur de celle-ci. Les deux autres sont la verticale et l'horizontale dans

la direction du plan de façade, qui correspond d'ailleurs souvent approximativement à la direction de déplacement du véhicule (voir figure 2.12).

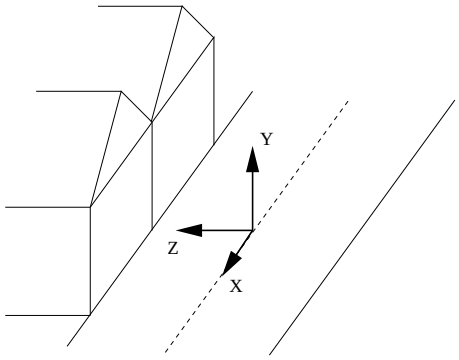


FIG. 2.12 – Repère choisi pour décrire l'espace objet

Dans cette géométrie, il est possible de décrire une façade par une surface du type $z = f(x,y)$. Le parcours de l'espace objet se fait donc à x et y fixés, en discrétisant l'axe z, et en reprojetant chaque point 3D dans toutes les images. Nous sommes ramenés à un problème de calcul de modèle numérique de surface sur la façade des bâtiments.

Nous avons choisi d'implémenter une technique basée sur la géométrie objet de la manière suivante. Soit (X,Y,Z) le système de coordonnées, choisi de manière à ce que l'axe Z soit perpendiculaire à la façade. Pour être plus précis, l'axe Z est perpendiculaire à la trajectoire du véhicule qui a capturé les images, en faisant l'hypothèse que la rue dans laquelle il roule est parallèle aux façades de bâtiments. L'axe Y est choisi vertical, l'axe X complète le système d'axe pour que le repère (X,Y,Z) soit direct.

En géométrie objet, l'espace de recherche est le volume défini par :

$$\mathscr{V} = \{(X,Y,Z) \in [X_{min},X_{max}] \times [Y_{min},Y_{max}] \times [Z_{min},Z_{max}]\} \tag{2.12}$$

X_{min} et X_{max} correspondent aux limites de recherche dans le sens du déplacement du véhicule, Y_{min} et Y_{max} correspondent respectivement à l'altitude du sol et à l'altitude maximale des bâtiments, Z_{min} et Z_{max} à la plage de recherche en profondeur (voir également la figure 2.12). Cet espace est parcouru de manière discrétisée, avec des pas de discrétisation $pas_X = pas_Y$ et pas_Z. La recherche de la surface du bâtiment s'effectue en parcourant l'axe Z entre Z_{min} et Z_{max}, pour chaque X et Y. La figure 2.13 montre une vue

schématique de dessus de ce processus.

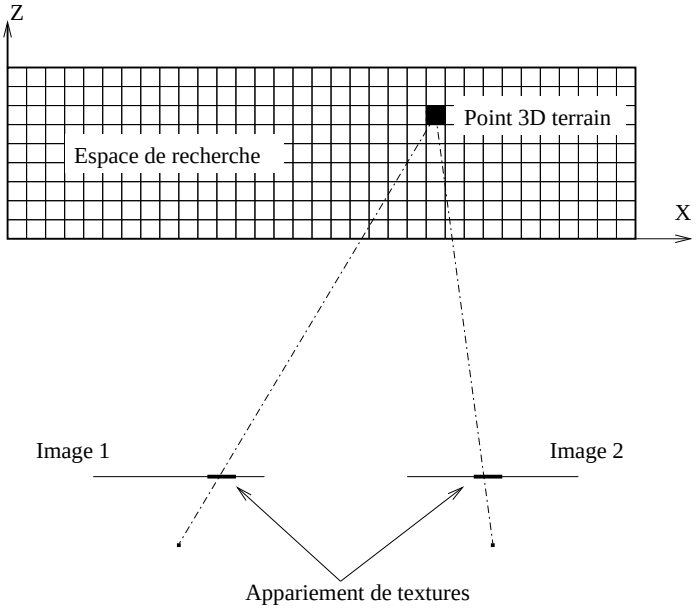


FIG. 2.13 – Reprojection d'un point 3D dans les images. Vue de dessus.

Lorsqu'il n'y a pas de parties du bâtiment masquées dans les images par d'autres éléments de la scène, il est possible de déterminer *a priori* les images dans lesquelles un point est visible. Pour un point $M(X,Y,Z)$, notons I_M le sous-ensemble des images qui voient ce point. Soit alors m_i la projection de M dans l'image i où $i \in I_M$. Soient c_i et l_i les coordonnées (colonne,ligne) sub-pixellaires du point m_i. Pour chaque point M de l'espace, nous pouvons alors calculer un coefficient de corrélation multi-images, en considérant les vignettes centrées sur les points m_i, $\forall i \in I_M$. Le ré-échantillonnage des images est réalisé a priori par le calcul d'une image ré-échantillonnée pour chaque plan $Z = cste$, afin d'éviter toute redondance de calcul.

Les niveaux de gris des pixels de la fenêtre de corrélation centrée en (c_i, l_i) sont notés v_{ij}, et sont représentées par le vecteur $\mathbf{V_i}$:

$$\mathbf{V_i} = \begin{pmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{in} \end{pmatrix} \tag{2.13}$$

et le coefficient de corrélation multi-images, issu des travaux PAPARODITIS et al. 2000 est donné par :

$$cor(X,Y,Z) = \frac{var(\sum_{i \in I_M} \mathbf{V_i})}{\sum_{i \in I_M} var(\mathbf{V_i})} \tag{2.14}$$

où $var(\mathbf{V_i})$ correspond à la variance des composantes du vecteur $\mathbf{V_i}$. Ce coefficient est symétrique par rapport à toutes les images.

A X et Y fixés, l'ensemble I_M des images voyant le point $M(X,Y,Z)$ dépend de Z. Or le coefficient de corrélation $cor(X,Y,Z)$ dépend de I_M. Afin d'éviter des discontinuités de $cor(X,Y,Z)$ quand Z varie, nous remplaçons I_M par $I_{X,Y}$ dans l'équation 2.14 où :

$$I_{X,Y} = \bigcap_{Z_{min}<Z<Z_{max}} I_M \tag{2.15}$$

Finalement, la meilleure profondeur $\hat{Z}_{X,Y}$ à X et Y donnés s'exprime comme :

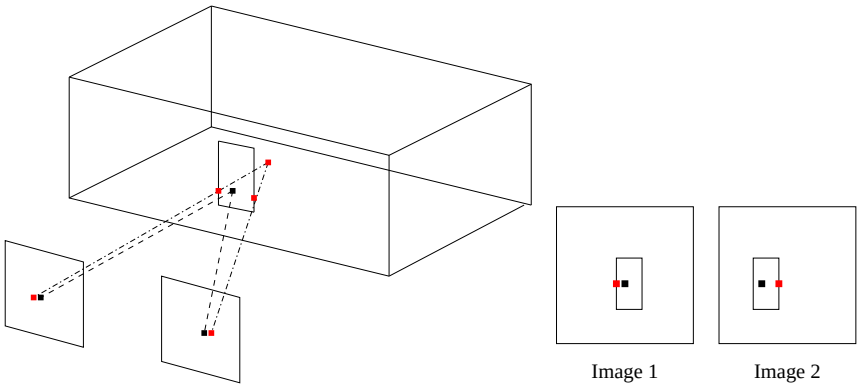
$$\hat{Z}_{X,Y} = \underset{Z_{min}<Z<Z_{max}}{\operatorname{argmax}} \; cor(X,Y,Z) \tag{2.16}$$

La figure 2.14 montre le modèle numérique de façade (MNF) obtenu, avec un pas de 2 *cm* dans le plan de la façade, et une taille de fenêtre de corrélation de 5×5 pixels. Ces résultats ne sont pas très bons et présentent plusieurs défauts. La profondeur trouvée sur la partie vitrée de la fenêtre est complètement erronée du fait de l'absence de texture et des effets spéculaires du verre. Ceci est un effet attendu lié à notre problème : on ne peut trouver des homologues lorsque les radiométries sont très différentes. En revanche, de nombreuses erreurs sont présentes sur des zones texturées, même lorsqu'il n'y a pas de parties cachées. Ce phénomène s'explique par le fait qu'un bâtiment possède de nombreuses structures spatialement répétitives (montants des fenêtres, joints de pierre,...). Une erreur d'appariement qui consisterait à prendre deux joints de pierre différents dans deux images peut conduire, du fait de la similarité de ces joints, à un score élevé. La figure 2.15 l'illustre schématiquement. Il est possible d'augmenter la taille des vignettes de corrélation pour pallier au moins partiellement ce problème, mais ceci nuirait fortement à une détection précise des discontinuités.

La recherche de surface dans la géométrie objet accentue ces erreurs. En effet, pour X et Y donnés,

(a) Partie d'une façade (b) Zoom sur une fenêtre

FIG. 2.14 – MNF partie droite de la façade. Pas planimétrique : 2 cm. Taille des fenêtres de corrélation : 5 pixels. Echelle des profondeurs : plus un pixel est sombre, plus le point est éloigné



(a) Vue 3D (b) Vue image

FIG. 2.15 – Un appariement faux peut avoir un score de corrélation plus élevé lorsque la position réelle correspond à une zone faiblement texturée. Points noirs : position réelle. Points rouges : position obtenant un score de corrélation élevé.

nous recherchons le meilleur Z. La reprojection du point M lorsque Z varie est donc modifiée dans toutes les images appartenant à $I_{X,Y}$. Or la corrélation favorise les zones contrastées, et le long du segment de recherche le score maximum peut être obtenu pour une profondeur distincte de celle de la façade réelle, comme dans le cas de la figure 2.15.

2.4.4 Géométrie image

Afin de tenter de pallier certains des problèmes apparus pour la géométrie objet, nous choisissons ici de reprendre la recherche du meilleur appariement en parcourant différemment l'espace terrain. Nous

utilisons un repère directement basé sur l'image avec les coordonnées (c,l,p), colonne, ligne et profondeur pour tout point 3D du cône de visibilité de l'image. La droite issue du pixel (c,l) est discrétisée. En stéréo-vision, le pas naturel de discrétisation est de un pixel le long de la droite épipolaire, ce qui fournit la mesure de disparité. Dans le cas multi-images, un pixel donné d'une image a n homologues, un pour chacune des n autres images voyant le même détail de la scène. La disparité, classique en stéréo-vision, ne peut alors être employée car le déplacement le long des n droites épipolaires est asservi et n'est pas uniforme en fonction de l'image considérée. Nous choisissons donc un pas de discrétisation en profondeur selon l'axe du rayon lumineux issu du pixel de référence.

Nous appelons "géométrie image" ce type de discrétisation, car c'est la géométrie de la prise de vue de référence, et en particulier l'orientation des rayons lumineux issus de chacun de ses pixels, qui conduit le parcours de l'espace 3D. Il est possible de réaliser le même travail avec un parcours différent de l'espace 3D, que nous appelons géométrie "objet", décrite dans le paragraphe suivant.

La figure 2.16 montre la stratégie de parcours de l'espace en géométrie image.

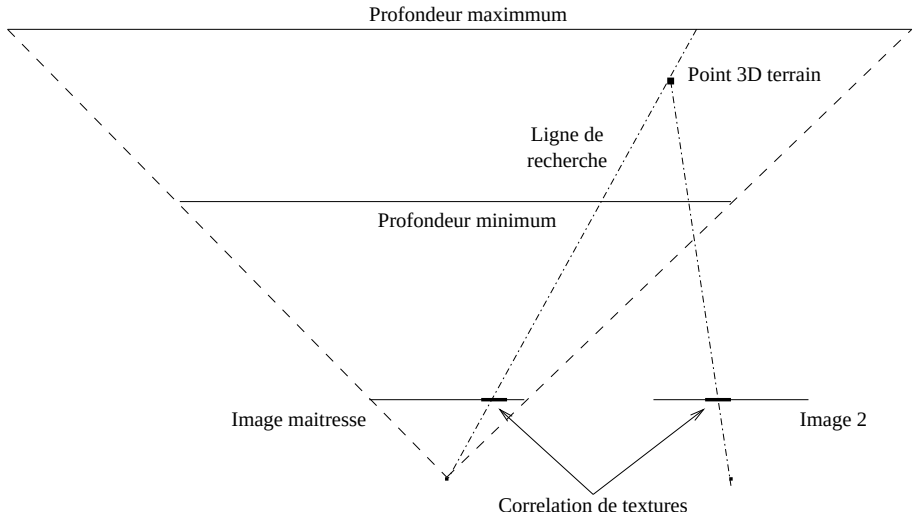


FIG. 2.16 – Ligne de parcours de l'espace de l'image maîtresse à (c,l) fixé. Reprojection d'un point terrain dans les autres images. Vue de dessus.

La stratégie suivie est la suivante. Une image maîtresse est choisie. Pour chacun de ses pixels (colonne, ligne) nous cherchons à retrouver la profondeur à laquelle se situe l'image réelle. Un point M de l'espace en trois dimensions sera repéré par ses cordonnées (c,l,p) liées à l'image maîtresse : colonne,

ligne, profondeur. Le principe est alors le suivant : à c et l fixés, nous recherchons le "meilleur" p, noté $\hat{p}$ entre deux bornes p_{min} et p_{max}. Le "meilleur" p s'entend comme la profondeur qui maximise le score de corrélation :

$$\hat{p}_i(c,l) = \underset{p_{min}<p<p_{max}}{\operatorname{argmax}} \; cor_i(c,l,p) \tag{2.17}$$

Le coefficient de corrélation choisi mesure la similarité de chaque vignette avec la vignette de l'image maîtresse. Nous avons choisi de faire la moyenne simple du coefficient de corrélation classique à deux images :

$$cor_i(c,l,p) = \frac{1}{N-1} \sum_{j \neq i} ZNCC(\mathbf{V_i}, \mathbf{V_j}) \tag{2.18}$$

où i est l'indice de l'image maîtresse, N le nombre total d'images de $I_{c,l}$, $\mathbf{V_i}$ le vecteur texture de la fenêtre de corrélation centrée sur la projection de (c,l,p) et $ZNCC(\mathbf{V_i}, \mathbf{V_j})$ le coefficient de corrélation normée et centrée.

Le calcul des scores de corrélation nécessite de ré-échantillonner dans la géométrie de l'image maîtresse les fenêtres de corrélation pour chaque c, l, p, et pour toutes les images. Or chacun des pixels des fenêtres de corrélation de taille t intervient t^2 fois : il convient d'éviter de faire t^2 fois chaque rééchatillonage. Afin de limiter ces calculs, nous procédons de la manière suivante : à p fixé, nous ré-échantillonons chacune des images dans la géométrie de l'image maîtresse. Les calculs de corrélation peuvent ensuite directement se servir de ces images ré-échantillonnées, sans nouvelle interpolation.

Discrétisation Les pas de discrétisation en colonne et ligne sont de 1, suivant la grille image. Comment choisir le pas de discrétisation dans le sens de la profondeur ? Un pas constant n'est pas adapté car il mène à une description trop fine aux grandes profondeurs. Dans le cas stéréoscopique, le pas naturel est celui de la disparité. Avec des images multiples, nous souhaitons également que le pas augmente avec la profondeur. Nous choisissons d'utiliser des voxels de forme constante et dont la taille augmente avec la profondeur (figure 2.17).

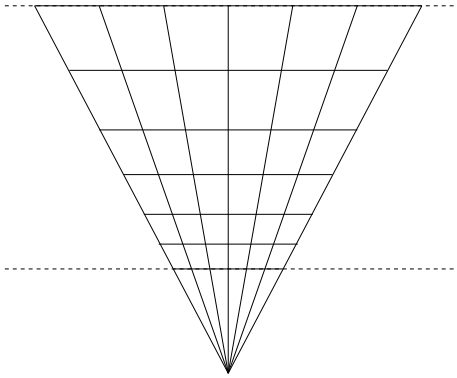


FIG. 2.17 – Ligne de parcours de l'espace de l'image maîtresse à (c,l) fixé. Reprojection d'un point terrain dans les autres images. Vue de dessus.

2.4.5 Comparaison de géométries

Dans les deux paragraphes précédents, nous avons vu qu'il était possible de déterminer une carte de profondeur sur une façade de bâtiment en géométrie objet ou image.

La géométrie objet est la plus naturelle pour décrire l'espace 3D de la scène. Tout d'abord, elle fournit un repère unique pour décrire l'ensemble de la façade ce qui n'est pas le cas de la géométrie objet. En effet, il faut plusieurs images (voire plusieurs dizaines) pour couvrir la totalité d'une façade. Pour chaque image maitresse considérée dans la géométrie image, un nouveau repère est utilisé. De plus, dans le cadre de la géométrie image, pour chacune des images maîtresses un morceau de la surface de la façade est reconstruit. Il y aura nécessité dans un second temps d'assembler ces morceaux en une description unique. Enfin, dans la géométrie objet, les images jouent un rôle parfaitement symétrique, alors que cette symétrie est brisée pour la géométrie image, car une des images joue le rôle de référence pour la recherche d'appariement. Néanmoins, l'analyse des résultats montre un net avantage de la géométrie image.

2.4.6 Algorithme proposé

L'algorithme 2.2 donne une vue d'ensemble de la stratégie proposée. Il reprend les modules présentés dans ce chapitre : appariement en géométrie image, en multi vues, avec régularisation de surface et stratégie multi-résolution.

Algorithme 2.2 : Stratégie d'appariement dense en multi-résolution

Données : $\mathscr{I}$, ensemble de N images géoréférencées de la façade
Sorties : Carte de profondeur pour chaque image

```
pour tous les im ∈ 𝒥 faire
    Création de la pyramide d'images sous-échantillonnées jusqu'à la résolution resol_min
fin
pour resol = resol_min à resol_max faire
    pour tous les im ∈ 𝒥 faire
        // im est l'image maîtresse
        si resol ≠ resol_min alors
            Calcul de la zone de recherche par dilatation de nappe_courante
        fin
        Découpage de l'espace image en dalles;
        pour tous les dalles faire
            pour tous les voxels v ∈ zone de recherche faire
                Détermination de 𝒥_v ⊂ 𝒥, sous-ensemble des images voyant le voxel v ;
                Calcul des scores de corrélation cor_im(v) restreint à 𝒥_v;
            fin
            Optimisation;
        fin
        nappe_courante ← assemblage_dalles() ;
    fin
    resol = resol + 1 ;
fin
```

Numéro de point	81	82	83	91	92	93	101	102	103
Erreur (*mm*)	2,5	0,5	3,2	4,4	4,8	2,7	17,7	5,9	5,6

TAB. 2.2 – Erreur de reconstruction en profondeur sur neuf points de contrôle.

La figure 2.18 montre les résultats obtenus avec notre stratégie. La surface topographique du bâtiment est correctement reconstruite. Nous extrayons des détails de taille inférieure au centimètre en profondeur et au demi centimètre dans le plan de la façade. Nous disposons par ailleurs sur cette portion de façade de neuf points de contrôle mesurés avec grande précision (de l'ordre de $0,1$ *mm*). La mesure de l'erreur en profondeur sur ces points est donnée dans le tableau 2.2.

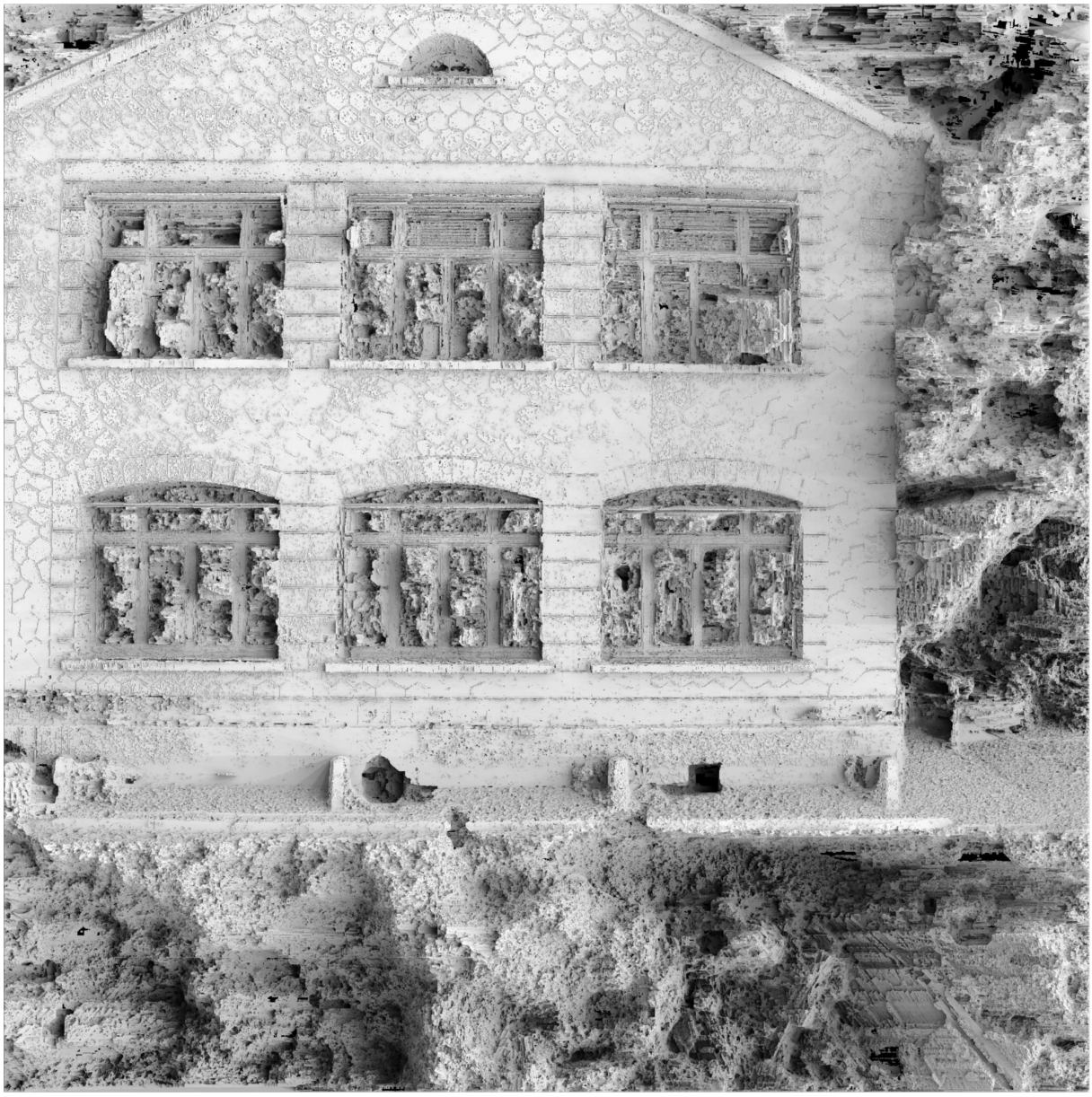


FIG. 2.18 – Carte de profondeur en multi-vues

2.5 Bilan

Nous avons proposé dans ce chapitre un algorithme de reconstruction topographique de façades de bâtiments. Nous avons présenté dans un premier temps l'utilisation des images de la base verticale du véhicule, pour lesquelles nous savons que la scène est identique. Le bruit de reconstruction observé nous a conduit à proposer une régularisation de surface, par une méthode de coupe de graphe. Nous avons ensuite présenté une version multi-résolution, qui permet de réaliser des gains de temps de calcul, de place mémoire et de robustesse sur les résultats. Enfin, nous reprenons le principe de cet algorithme en images multiples cette fois. La comparaison entre les géométries image et objet donne l'avantage à la première.

Nous disposons de cartes de profondeurs robustes et précises pour chacune des images en entrée de l'algorithme. Dans le chapitre suivant, nous proposons une méthode de fusion de ces cartes en une surface triangulée unique.

CHAPITRE 3

RECONSTRUCTION DE SURFACE

3.1 Introduction

Dans ce qui précède, nous avons vu comment obtenir la carte de profondeur dense d'une image de façade de bâtiment. Après cette première étape, la description de la scène dont nous disposons ne répond pas totalement aux objectifs fixés. Tout d'abord, les surfaces obtenues se limitent aux portions visibles depuis chaque point de vue, c'est à dire que nous disposons d'autant de portions de façade qu'il y a d'images en entrée. Il serait bien sûr pertinent de fournir une description par façade complète ou par bâtiment complet, soit en fusionnant les surfaces de points de vue voisins, soit en segmentant la surface si celle-ci est en fait composée de deux éléments de façades non connexes. Ceci est la directe conséquence du choix de travailler en géométrie image.

Ensuite, nous avons une description de type raster ou 2,5 D : à chaque pixel correspond une valeur de profondeur de la scène. Pour une scène aérienne supposée sans surplomb, une telle représentation de l'espace est pertinente. Ce n'est clairement pas le cas pour des bâtiments vus depuis le sol car on peut en faire le tour. On peut également remarquer que la représentation obtenue est assimilable à une description ponctuelle de la scène, en considérant pour chaque pixel le point 3D situé à la profondeur correspondante et placé sur le rayon perspectif passant par le centre du pixel. Afin de palier ces insuffisances, nous proposons dans ce chapitre de passer à une représentation de la surface par maillage triangulé. Nous extrayons tout d'abord les pixels fiables de la carte de profondeur (section 3.2). Ensuite, afin de tirer partie de la connaissance de la topologie image, la triangulation sera menée dans un premier temps pour chaque point de vue (voir section 3.3) puis un algorithme de fusion des maillages pour des points de vue en recouvrement est proposé (section 3.4). Les maillages triangulés sont par ailleurs très bien adaptés au

rendu de scènes tridimensionnelles, notamment pour la texturation.

3.2 Extraction des points fiables de la carte de profondeur

Au chapitre précédent, pour chaque image, la profondeur a été calculée pour tous les pixels. Cependant, cette valeur n'a pas de sens réel pour les pixels de ciel par exemple, et toutes les parties en dehors de la plage de recherche en profondeur (avant plan, bâtiments en arrière plan). Les vitres des fenêtres sont un autre exemple de zones pour lesquelles la profondeur trouvée est erronée, du fait des réflexions qui diffèrent d'une image à l'autre, et de l'absence de textures significatives. En conséquence, une étape de filtrage est proposée. Le but est d'enlever tous les pixels pour lesquels la fiabilité est faible.

Cette fiabilité est traduite par deux mesures. Tout d'abord, dans l'image maîtresse, la texture au voisinage du pixel doit être suffisamment élevée. Dans le cas contraire, l'appariement avec des zones similaires sera relativement indéterminé. La texture est mesurée par la variance des valeurs de niveau de gris du voisinage $var(\mathbf{V_i})$. Nous pouvons noter au passage que le choix de travailler dans la géométrie image apporte ici un avantage important par rapport à la géométrie objet, pour laquelle ce type de mesures est difficile à mettre en place. Ensuite, le score de corrélation doit être suffisamment élevé. Dans le cas contraire, l'objet correspondant à notre pixel est probablement en dehors de l'espace de recherche. L'ensemble $\mathscr{P}$ des pixels que nous cherchons à supprimer est défini par le respect d'au moins une des conditions suivantes :

$$\begin{cases} \max_{p_{min}<p<p_{max}} cor_i(c,l,p) < seuil_{cor} \\ var(\mathbf{V_i}) < seuil_{var} \end{cases} \tag{3.1}$$

où $seuil_{cor}$ est un seuil sur la valeur du score de corrélation et $seuil_{var}$ un seuil sur la variance de la fenêtre de corrélation de l'image maîtresse. La mise en oeuvre de ce filtrage se fait en deux temps. Tout d'abord, pour ces pixels, les scores de corrélation en entrée de l'optimisation sont tous fixés à une constante (par exemple 0) afin qu'ils n'influencent pas la surface régularisée :

$$\forall (c,l) \in \mathscr{P} \forall p \in [p_{min}; p_{max}] cor_i(c,l,p) = 0$$

Généralement ces zones peu texturées correspondent à des zones planes. Dans un deuxième temps, après

l'étape d'optimisation, nous générons un nuage de points 3D en ignorant les pixels filtrés. Ces points 3D sont remis dans un repère terrain, afin de pouvoir réaliser un mélange entre toutes les vues maîtresses. La figure 3.1(b) montre le résultat après filtrage en géométrie image : dans l'images de droite, seuls les pixels fiables au sens de nos critères sont conservés.

(a) MNF après optimisation (b) MNF pour les points validés après l'étape de filtrage

FIG. 3.1 – Post-traitement en géométrie image : extraction d'un ensemble de points fiables. L'image maîtresse de départ se trouve figure 2.1 en bas à droite

La figure 3.2 montre un autre résultat de l'extraction de points fiables sur une carte de profondeur filtrée ainsi que quelques vues 3D du nuages de points correspondant. Ces résultats ont été obtenus à partir d'une vue stéréoscopique à base horizontale de la chapelle du château de Vincennes.

3.3 Triangulation par point de vue

3.3.1 Nuage de points 3D

Nous considérons dans cette partie la surface obtenue sous la forme d'un nuage de points 3D, représentation équivalente de la géométrie de la scène. En pratique, pour chaque pixel fiable (c,l) de la carte de profondeur, affecté de sa profondeur p, nous considérons le point $M(X,Y,Z)$ en trois dimensions tel que le rayon perspectif (OM) – où O désigne le centre de la prise de vue – passe par le centre du pixel

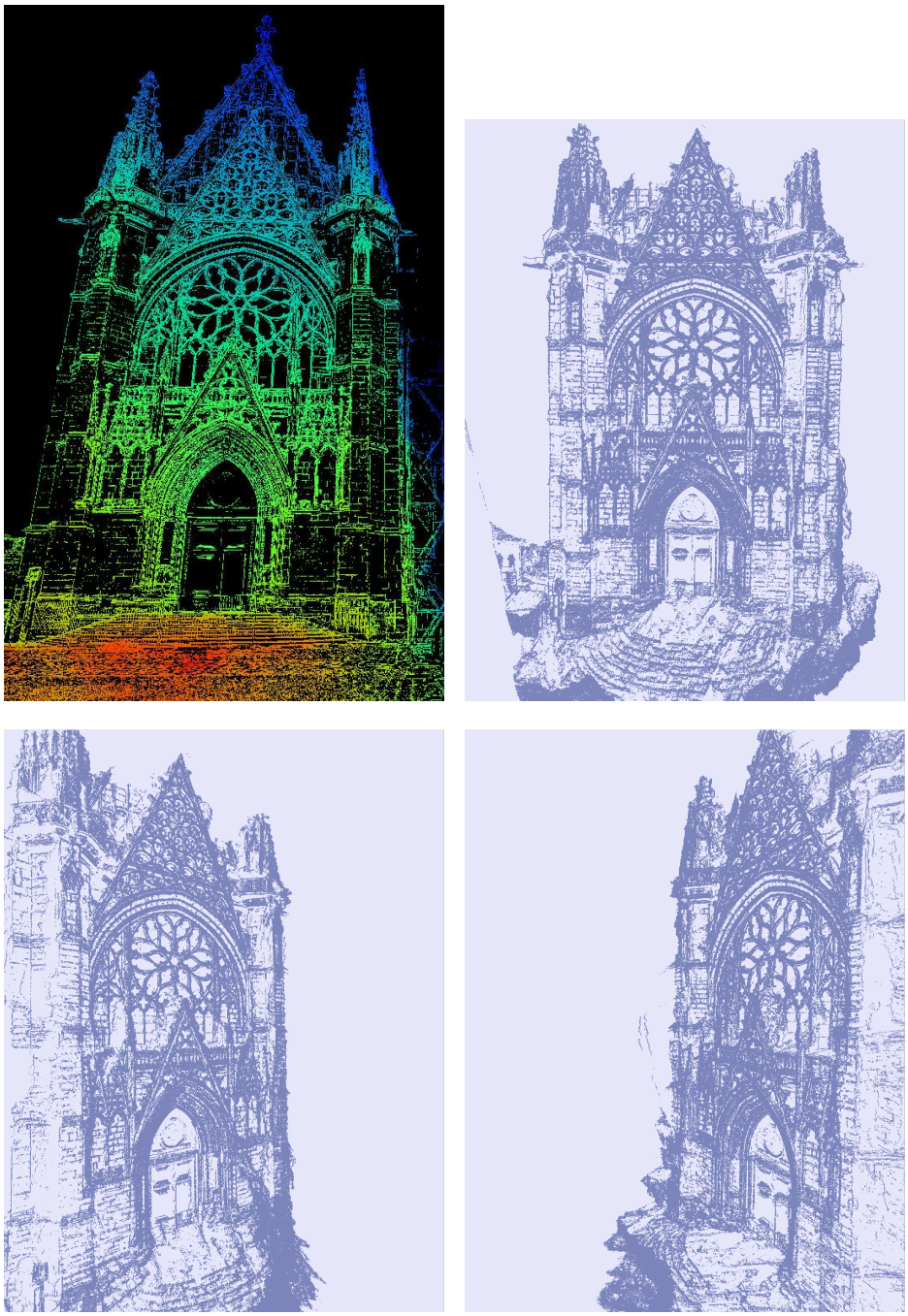


FIG. 3.2 – Vues du nuage de points dérivé de la carte de profondeur obtenue sur la façade de la chapelle du château de Vincennes

(c,l) et $\|OM\| = p$. La figure 3.2 montre quelques vues 3D du nuage de points correspondant à une carte de profondeur filtrée. Le nuage de points ainsi obtenu à partir d'un point de vue conserve la topologie de l'image maîtresse support de la carte de profondeur. C'est une caractéristique importante, que l'on retrouve dans le cas de nuages acquis par des scanners LASER à balayage. Cette topologie image permet de réaliser une triangulation bidimensionnelle classique du nuage.

Le passage à un nuage de points permet de réaliser un premier mélange de la géométrie obtenue pour chaque point de vue. En effet, toutes les images sont géoréférencées dans le même système de coordonnées, donc tous les nuages de points calculés peuvent également être exprimés dans ce système de coordonnées commun. La figure 3.3 montre les différents nuages de points obtenus sur le jeu de données photothèque de l'IGN. Une couleur différente a été attribuée pour les nuages de chaque point de vue.

Il n'y a pas de décalage visible entre les nuages de points. Ce résultat était attendu pour deux raisons :

- toutes les images ont été géoréférencées par un ajustement par faisceaux global, avec des points d'appui communs entre les images ;
- les points sont issus d'un processus de corrélation multi-images, donc pour chacun des nuages, les points sont par construction cohérents avec toutes les images voisines.

3.3.2 Triangulation de points 3D

Obtenir une triangulation adaptée à la scène à partir de points 3D n'est pas toujours aisé dans un cadre général. En effet, un nuage de points 3D n'a pas de topologie, et les relations de voisinages au plus proche voisin ne répondent pas toujours au critère d'appartenance à la même surface. Comme cela a été développé au paragraphe précédent, nous sommes dans le cas d'un nuage de points possédant une topologie image. La relation de voisinage en 3D est directement déduite de l'adjacence des pixels d'où sont issus les faisceaux perspectifs passant par les points 3D dans l'image maîtresse qui a été utilisée. Ainsi, la triangulation des points se ramène à une triangulation en deux dimensions dans l'image sur laquelle se projettent les points.

A l'issue de cette étape, nous disposons d'un modèle de surface pour la façade, sous la forme d'un maillage triangulé ayant pour sommets les points 3D reconstruits. Dans le cas de discontinuités en pro-

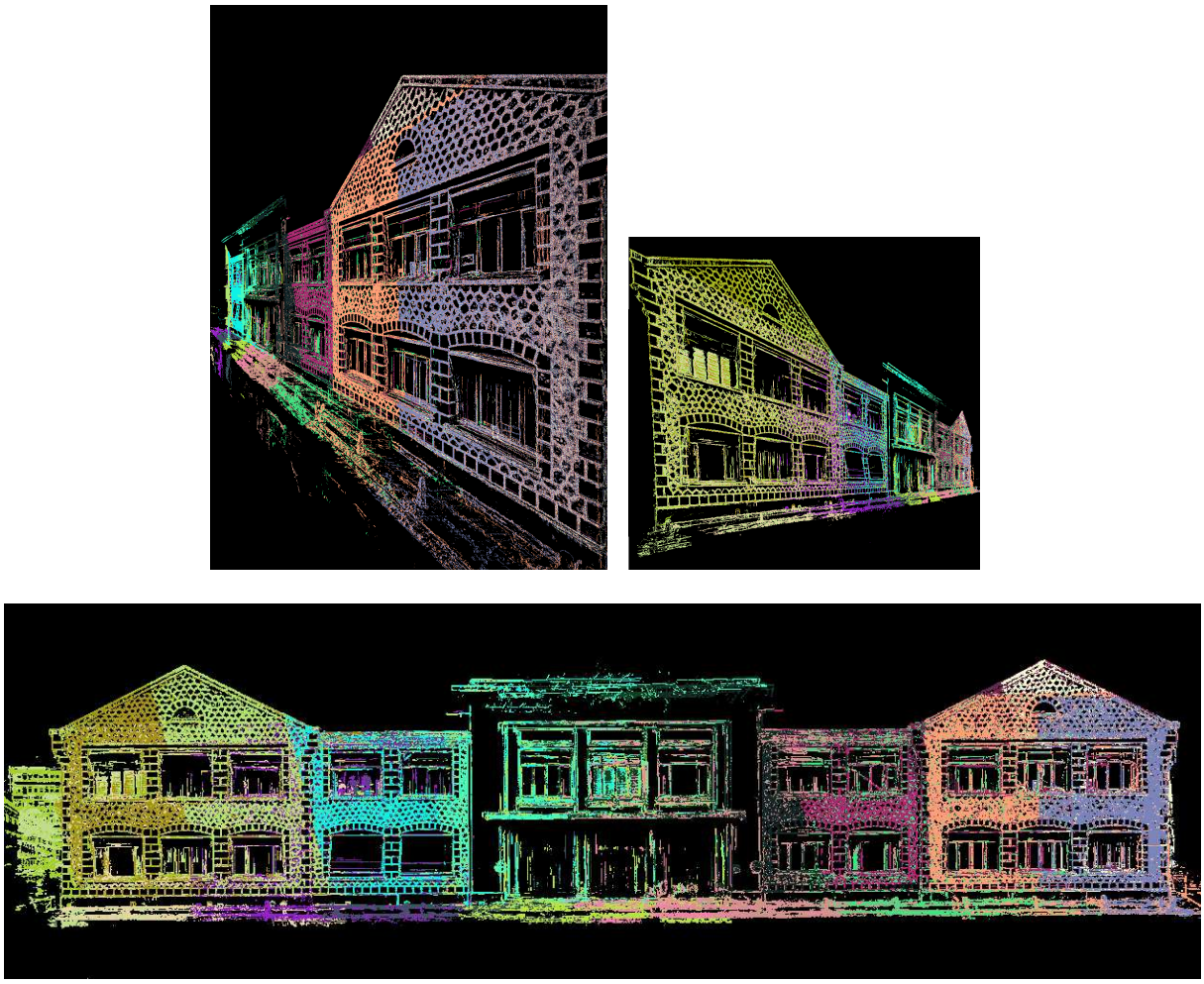


FIG. 3.3 – Nuage de points 3D. Les points extraits à partir d'images maîtresses différentes ont des couleurs différentes.

fondeur, des triangles ont été connectés des points qui ne sont pas voisins dans la réalité. Ces triangles peuvent être filtrés à ce stade en imposant que la normale au plan 3D support de la facette triangulaire ne soit pas orthogonale à la direction du rayon perspectif.

Il est enfin possible de texturer le modèle ainsi obtenu à l'aide de l'image maîtresse. Ceci achève l'obtention d'un modèle 3D pour chaque point de vue.

3.4 Fusion de maillages triangulés

La représentation d'une portion de façade de bâtiment vue sur une image par un maillage triangulé permet de disposer d'une surface dans l'espace tridimensionnel. Il nous faut maintenant passer à une description "façade par façade", pour cela il faut fusionner les portions précédemment reconstruites. Dans cette partie, nous présentons un rapide état de l'art des méthodes de fusion de surfaces triangulées avant de proposer notre algorithme.

3.4.1 Etat de l'art

L'intégration ou la fusion de maillages triangulés a été fréquemment abordée dans la littérature. Généralement, il s'agit de la dernière étape du processus de traitement de données acquises par des systèmes actifs tel des scanners à balayage laser à partir de plusieurs stations ou points de vue. Avec de tels dispositifs, l'enchaînement des étapes est le suivant (voir par exemple ZHANG et al. 2000) :

- acquisition des données ;
- recalage des nuages (géoréférencement relatif des données) ;
- fusion ou intégration des données.

Dans notre cas, seule la dernière de ces étapes reste à réaliser, la mise en place relative des nuages étant assurée.

Comme mentionné dans CURLESS et LEVOY 1996, la classification des algorithmes de fusion peut se faire selon deux axes transversaux :

- nuages structurés ou non structurés ;
- fusion surfacique ou volumique.

Ces termes seront explicités dans la suite.

Le lecteur intéressé par une description très détaillée des principales méthodes de fusion pourra se reporter au chapitre 8 de la thèse de Benjemaa (BEN-JEMAA 1998) qui dresse un état de l'art fouillé datant d'une dizaine d'années.

Nuages non structurés

La reconstruction basée sur des nuages de points non structurés doit se passer de toute connaissance a priori sur le voisinage entre points, autre que leur position en 3D. Des auteurs ont proposé des algorithmes pour réaliser la fusion de telles données. Boissonat (BOISSONNAT 1984) par exemple passe par une triangulation de Delaunay de l'ensemble des points 3D. D'autres auteurs passent par la représentation de la surface comme surface de niveau d'une fonction implicite. Cette méthode est introduite par Hoppe (HOPPE et al. 1992), la fonction utilisée est une distance signée. Elle inspirera la plupart des auteurs utilisant une fusion volumique dans le cas de nuages structurés.

Nuages structurés

La reconstruction basée sur des nuages de points structurés pourra au contraire tirer partie de l'information topologique disponible, ce qui est le cas avec les nuages de points issus d'acquisitions par scanner à balayage ou calculés à partir de cartes de profondeur.

Méthodes volumiques Ces méthodes ont été baptisées volumiques car les traitements qu'elles contiennent se déroulent directement dans l'espace 3D, à la différence des méthodes surfaciques décrites plus bas. Les deux principales méthodes sont décrites dans CURLESS et LEVOY 1996 et HILTON et al. 1996. Elles sont toutes deux basées sur le même principe de construction d'une surface implicite dans l'espace 3D discrétisé puis de l'extraction d'une iso-surface. Une fonction f est définie telle que sa courbe de niveau 0, c'est à dire $L_0 = \{X \in \mathbb{R}^3 | f(X) = 0\}$, contiennent les points du nuage (ou que ses points soient proches de L_0). Cette opération est réalisée pour chacun des nuages à fusionner et donne donc un ensemble de fonctions f_i, $0 \leqslant i \leqslant nb_nuages$. Il reste ensuite à définir une fonction f, "fusion" des f_i puis à extraire la surface recherchée.

Hilton et al. (HILTON et al. 1996) se servent d'une fonction de type distance signée, de manière similaire à HOPPE et al. 1992 mais ils ajoutent une fonction binaire pour prendre en compte les discontinuités des nappes à fusionner. Un ensemble de règles leur permettent ensuite d'obtenir une fonction f unique, calculée en tout point de l'espace, en réalisant un compromis entre toutes les données redondantes. Ensuite, l'extraction de l'iso-surface est réalisée par l'algorithme de "marching triangles".

Curless et al. (CURLESS et LEVOY 1996) se basent également sur une distance signée, mais ils tirent partie de leur capteur par translation pour calculer la distance suivant la direction du capteur. Les nappes sont ré-échantillonnées pour accélérer le calcul des distances signées qui sont intégrées en une seule fonction par une moyenne pondérée. Enfin, l'extraction de l'iso-surface se fait par l'algorithme classique des "marching cubes" (LORENSEN et CLINE 1987).

Méthodes surfaciques Par opposition aux méthodes volumiques, les méthodes surfaciques sont des approches basées sur la 2,5 D. Les images de profondeurs sont triangulées et superposées, puis localement fusionnées pour éliminer les données redondantes. Plusieurs approches ont été proposées.

Rutishauser et al. RUTISHAUSER et al. 1994 utilisent une approche probabiliste basée sur les erreurs le long de la direction d'acquisition du capteur afin d'établir un consensus pour la position de la surface.

D'autres méthodes (SOUCY et LAURENDEAU 1995 ; TURK et LEVOY 1994 ; PITO 1996) travaillent à partir de surfaces issus des nuages initiaux triangulés indépendamment chacun avec leur topologie. Ces trois méthodes respectent le même schéma algorithmique général :

- détection des zones en recouvrement ;
- suppression des données redondantes ;
- raccordement des surfaces pour créer une surface unique.

Dans le cas de SOUCY et LAURENDEAU 1995, les recouvrements sont déterminés à l'aide d'un diagramme de Venn. Le diagramme de Venn de N nappes se définit comme un graphe d'adjacence de régions qui sont toutes les zones en recouvrement à exactement 1 nappe, 2 nappes, 3 nappes,...,N nappes, et que l'on appelle sous-ensemble canonique (CS). Formellement, pour un ensemble S de K nappes ($S = \{V_1, \ldots, V_K\}$), le sous-ensemble canonique associé est :

$$CS = \left\{ X \in \mathbb{R}^3 | X \in \bigcap_{a=1}^{K} V_a \text{ et } X \notin \bigcup_{a=K+1}^{N} V_a \right\} \tag{3.2}$$

La figure 3.4 montre un exemple de diagramme de Venn à trois nappes.

L'appartenance d'un point à une nappe est déterminée par un critère de voisinage et un critère de visibilité. Dans les travaux de Turk et al., la détermination de la redondance de deux triangles appartenant à des surfaces issues de la triangulation de deux nuages distincts est faite par application d'un seuil sur la

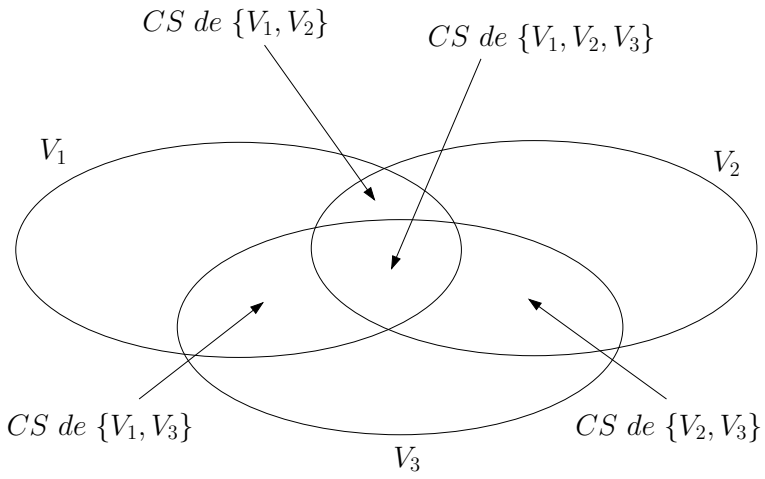


FIG. 3.4 – Diagramme de Venn pour trois nappes (d'après SOUCY et LAURENDEAU 1995).

distance séparant les sommets d'un triangle à la surface support de l'autre. La méthode utilisée par Pito est sensiblement identique.

La phase suivante consiste en la suppression des données redondantes.

Dans SOUCY et LAURENDEAU 1995, les auteurs tirent partie des sous-ensembles canoniques en recouvrement et considèrent une profondeur comme moyenne pondérée des valeurs de chaque nappe. Ils retriangulent ensuite le sous-ensemble canonique modifié. Turk et al. suppriment quant à eux les triangles détectés redondants dans la phase précédente jusqu'à ne conserver qu'un recouvrement minimal (aucun sommet en recouvrement). Les auteurs se servent des points d'intersection entre les deux nappes pour retrianguler localement, et obtenir une nappe unique.

3.5 Méthode proposée et résultats obtenus

Le principe de l'algorithme de fusion retenu est celui d'une fusion surfacique, bien adaptée aux cartes de profondeurs 2,5 D obtenues précédemment. La fusion est réalisée de manière itérative : à partir d'une première nappe, on cherche à ajouter / fusionner une nouvelle nappe. Ensuite, les nappes sont ajoutées / fusionnées une à une, venant grossir la nappe de départ qui devient la surface fusionnée recherchée. Pour chaque fusion d'une nouvelle nappe (notée S_2) avec la nappe courante (notée S_1), le déroulement est le suivant :

– détection et suppression des triangles redondants : ces triangles sont détectés sur la surface S_1 comme incompatibles avec la surface S_2. Les triangles correspondants de la surface S_2 sont également supprimés ;
– retriangulation des trous créés dans la surface S_1 par le processus précédent ;
– création du raccordement des deux surfaces.

L'algorithme sera illustré par les deux surfaces triangulées notées S_1 et S_2 dans la suite, représentées sur la figure 3.5.

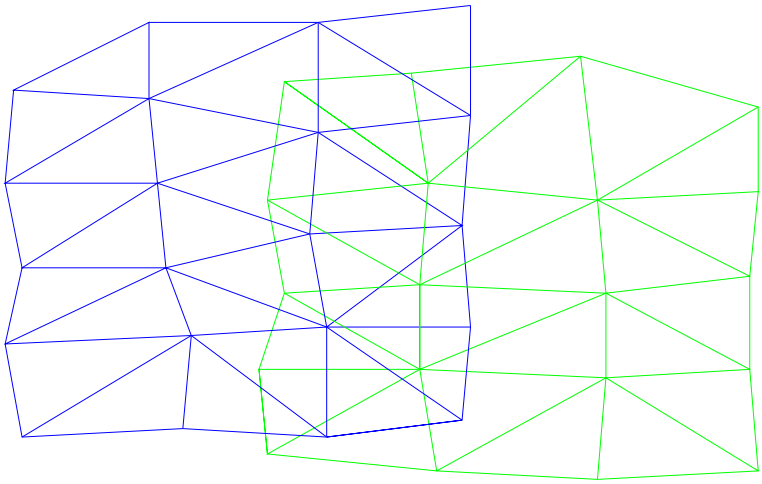


FIG. 3.5 – 2 surfaces en recouvrement. La surface S_1 est en vert, la surface S_2 en bleu.

Les paragraphes suivants détaillent ces étapes.

3.5.1 Filtrage des triangles incohérents

L'algorithme repose sur l'hypothèse suivante : les points 3D sur lesquels s'appuient les triangulations sont considérés comme fiables, c'est à dire sans faute. Cette hypothèse est évidemment trop optimiste dans le cadre d'une reconstruction à partir de techniques automatiques à partir d'images. En revanche elle serait bien adaptée à des nuages de points acquis par numérisation LASER. Nous conservons cette hypothèse pour exposer l'algorithme de fusion, en considérant que l'immense majorité des points 3D en entrée sont bien positionnés au voisinage de la surface réelle.

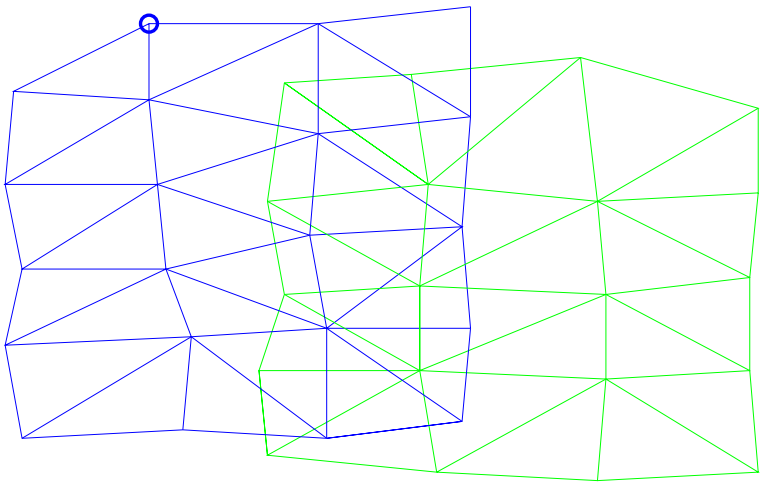
Nous considérons donc une surface S_1 à laquelle nous souhaitons fusionner une surface S_2. S_1 est

soit le résultat de fusions précédentes, soit la première surface triangulée avant d'itérer les fusions. S_2 est directement le résultat d'une triangulation de points issus d'un point de vue. Nous noterons C_2 le centre de prise de vue pour ce nuage, qui sera appelé N_2. Le principe de filtrage des triangles incohérents est le suivant : si pour un point M_2 du nuage N_2, le segment $[C_2M_2]$ intersecte la surface S_1, le triangle T_1 appartient à la surface S_1 où se produit l'intersection est dit *incohérent*. En effet, le point M_2 est vu depuis C_2, donc l'existence du triangle T_1 n'est pas cohérente. L'hypothèse de validité de tous les points 3D précédemment énoncée implique que le triangle T_1 a relié des points qui étaient topologiquement voisins dans la géométrie image associée, mais qui ne le sont pas sur la surface réelle de la scène. T_1 sera supprimé de la triangulation, et M_2 est gardé en mémoire comme point filtrant le triangle T_1. Ceci afin de retrianguler le trou créé dans le maillage.

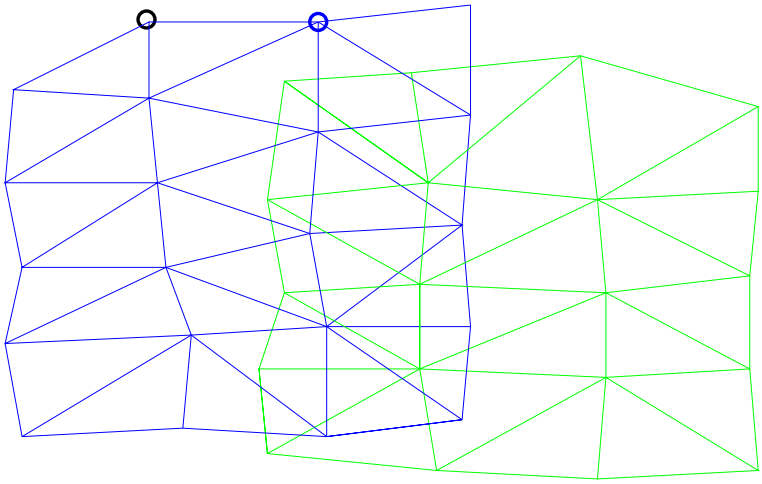
L'implémentation de ce qui précède peut se faire par lancer de rayon, c'est à dire que pour tout point $M_2 \in N_2$ on cherche tous les triangles T_1 tels que le segment $[C_2M_2]$ intersecte T_1. Cette technique a l'inconvénient d'être particulièrement lente. On peut lui préférer une technique basée sur un Z-buffer. Dans ce cas, on crée, dans la géométrie de la vue n°2 l'image du Z-buffer de la nappe n°1. Lors du parcours des pixels image, on dispose immédiatement du triangle intersecté par le segment $[C_2M_2]$, ce qui constitue un gain de temps considérable. Néanmoins en pratique, plusieurs cas de figure invalident la pertinence du Z-buffer. Le plus fréquent est la présence de nombreux petits triangles étirés. Ce problème est lié à la discrétisation de l'image 2, dont le pas n'est pas toujours suffisant pour distinguer 2 triangles. Nous conservons donc le lancer de rayon qui présente l'avantage de marcher dans tous les cas.

3.5.2 Retriangulation des trous

A l'issue de l'étape précédente, nous disposons d'une nappe S_1, avec la liste des triangles incohérents qu'il faut supprimer, et la liste associée des points issus du nuage N_2 à utiliser pour retrianguler les trous. La figure 3.8 montre schématiquement l'état des deux nappes à cet instant du déroulement de l'algorithme. Dans ce paragraphe, nous abordons uniquement les "trous" à l'intérieur de la nappe, pas les suppressions au bord de celle-ci. Nous procédons de la manière suivante : par croissance de région à partir d'un triangle à supprimer, nous considérons l'ensemble connexe maximal de triangles à supprimer. Nous réalisons la suppression, en conservant la liste des éventuels sommets rendus isolés par cette opération,



(a) Début du parcours des points : premier point.



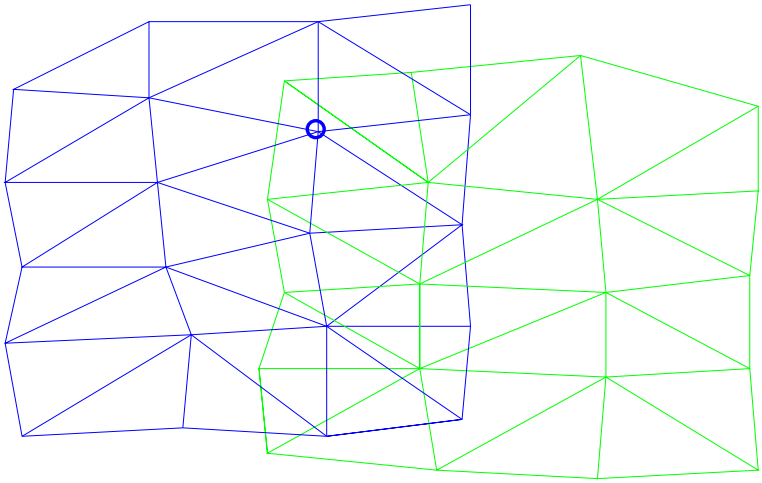
(b) Parcours des points : point suivant.

FIG. 3.6 – Parcours des points de S_2. Exemples de points qui ne sont pas en recouvrement avec S_1.

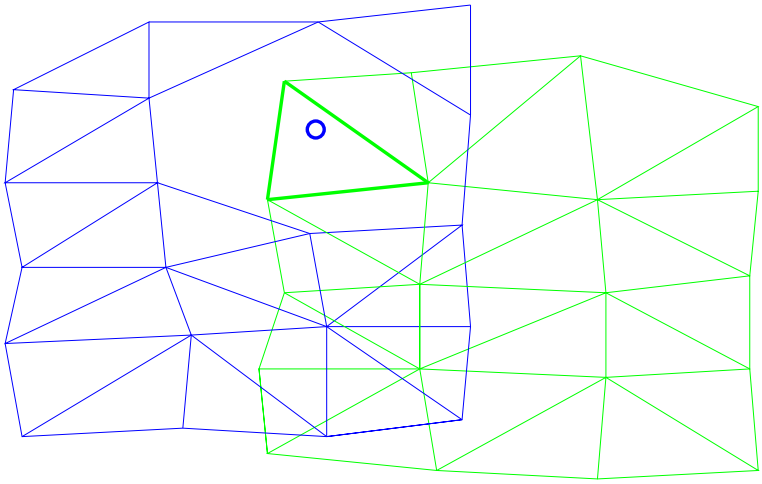
c'est à dire les points tels que tous ses triangles adjacents sont supprimés. Nous disposons également de l'ensemble des points du nuage N_2 responsables du filtrage de ces triangles.

Le "trou" ainsi créé est rebouché par la triangulation, dans la géométrie image n°2, de l'union de ces deux ensembles de points, contrainte par les segments frontières du trou, et ces nouveaux triangles sont ajoutés à la nappe S_1. La figure 3.9 illustre l'étape de retriangulation des trous. Le trou situé en bas montre est un exemple pour lequel plusieurs triangles connexes font partie du même trou.

Ceci termine la fusion des parties en recouvrement. Le paragraphe suivant détaille la fin de la fusion.



(a) Parcours des points : point en recouvrement.



(b) Suppression des arrêtes de S_2 connectées à ce point.

FIG. 3.7 – Parcours des points de S_2. Exemple d'un point en recouvrement avec S_1.

3.5.3 Création du raccord entre surfaces

Nous disposons à ce stade de deux surfaces : la nappe n°1 modifiée par la suppression de triangles puis retriangulation des trous, et la nappe n°2, limitée aux parties hors recouvrement. Dans ce paragraphe, nous proposons une méthode pour raccorder ces deux surfaces entre elles, afin de terminer la fusion des nappes. Nous utilisons une méthode voisine de l'approche dite "soudure", introduite par PITO 1996. Pour cela, nous avons besoin de connaître la partie de la frontière des deux nappes qui doivent être fusionnées. A cette fin, des marqueurs sont introduits lors de l'étape de filtrage exposée ci-dessus pour

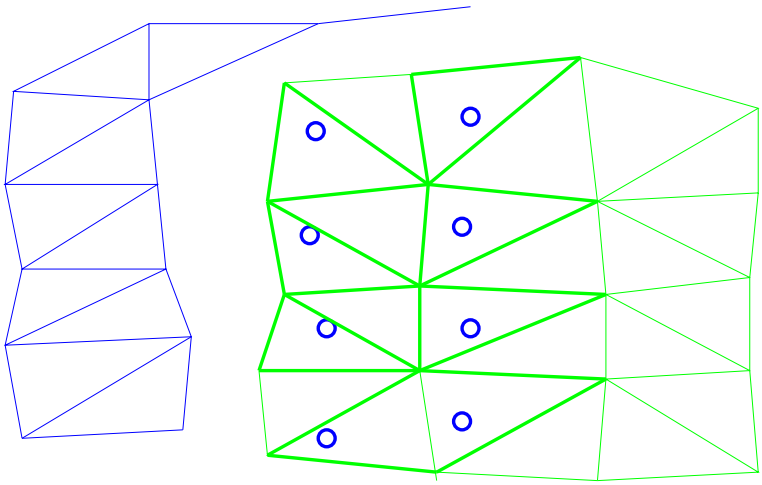


FIG. 3.8 – Fin du parcours des points de S_2.

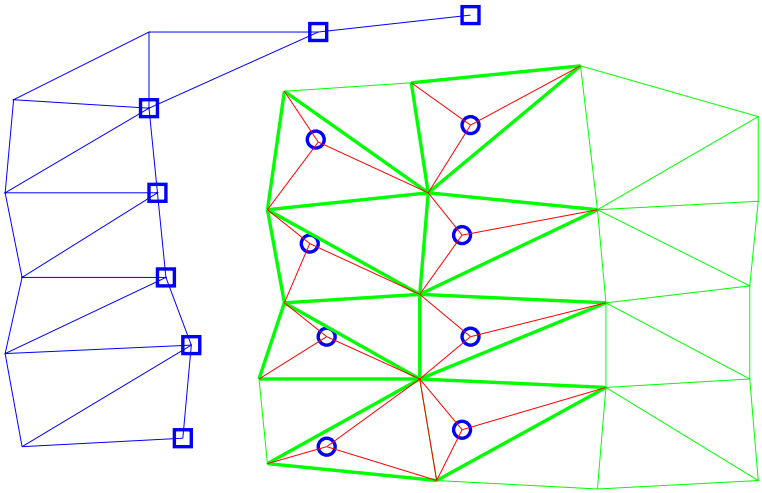


FIG. 3.9 – Retriangulation des trous de S_1.

suivre l'évolution de la frontière. L'algorithme de filtrage est donc modifié comme suit. Avant le début du filtrage, des ensembles sont créés : l'ensemble des points sommets de la surface S_2 et l'ensemble des points situés à la frontière S_2. Le premier est le nuage N_2, le second est noté F_2. On a bien sûr $F_2 \subset N_2$. Comme décrit plus haut, on réalise le parcours des points $M_2 \in N_2$. On teste si ce point est en recouvrement avec la surface S_1, c'est à dire si la demi-droite $[C_2M_2)$ intersecte S_1. Dans ce cas, les tests d'incohérence se poursuivent normalement, mais d'autres opérations sont ajoutées. Le point est retiré de N_2, et de F_2 s'il lui appartient. Ensuite F_2 est mis à jour en lui ajoutant les voisins de M_2 sur S_2 qui sont

encore dans N_2. A la fin du parcours des points de S_2, l'ensemble F_2 est constitué de la frontière de partie en non recouvrement de la surface S_1. Le raccordement entre les deux surfaces constitue la dernière étape de l'algorithme de fusion. Le principe en est simple. Nous disposons de l'ensemble F_2 des points de la surface S_2, qui sont la frontière de la partie qui n'est pas en recouvrement avec S_1. Pour chaque point de cette frontière, nous recherchons le segment de la frontière de S_1 qui en est le plus proche. L'objectif est de créer un triangle entre ce segment et ce point. Si plusieurs points de F_2 sont en concurrence pour un même segment de S_1, seul le plus proche sera raccordé dans un premier temps. Les autres sont reliés au sommet du segment le plus proche. Les faces correspondantes sont créées, c'est à dire les triangles avec pour base le segment, et pour sommet le point. Dans un second temps, les triangles inverses sont créés.

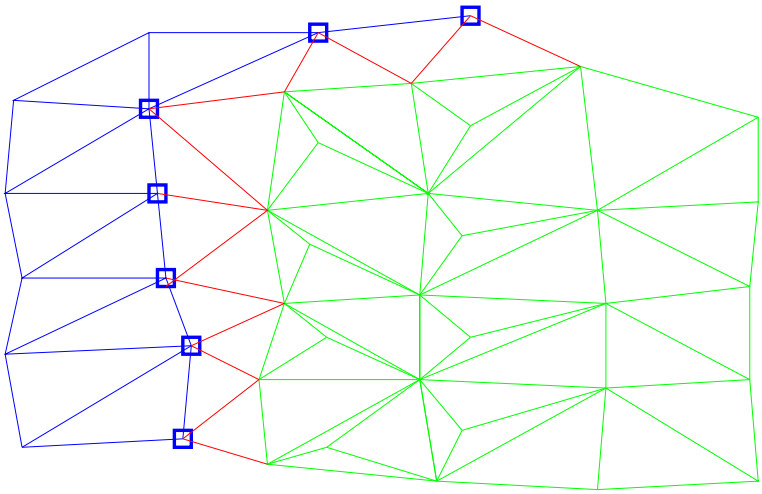


FIG. 3.10 – Création du raccordement entre les deux nappes.

A l'issue de ces traitements, le raccordement entre les deux nappes est terminé. Le résultat est le maillage donné figure 3.11

Cet algorithme, est utilisé de manière itérative, pour chaque surface correspondant à un nouveau point de vue. Ainsi, l'ensemble d'une façade peut être fusionnée sous la forme d'une surface triangulée

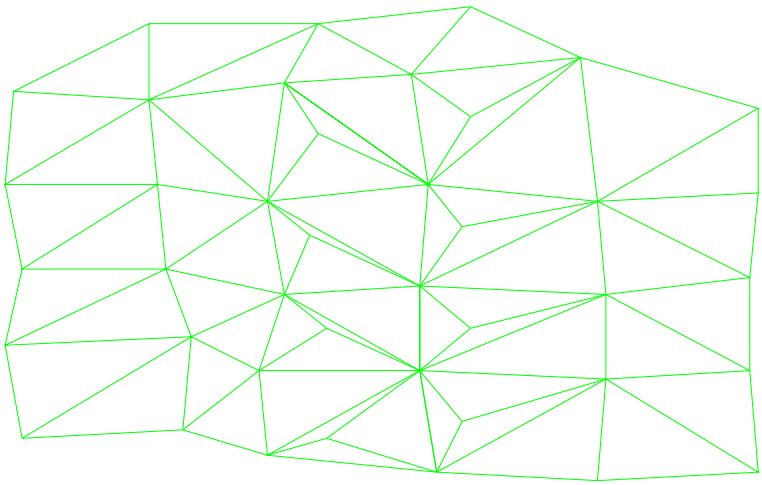


FIG. 3.11 – Résultat de l'algorithme de fusion (schéma)

unique. Le processus complet est résumé par l'algorithme 3.1

Algorithme 3.1 : Stratégie de fusions de maillages triangulés issus de différents points de vue

Données : L'ensemble des maillages triangulés issus des vues en recouvrement d'une façade

Sorties : Maillage triangulé de l'ensemble de la façade

```
// L'algorithme suppose que chaque nouvelle vue est en
   recouvrement partiel avec au moins l'une des précédentes
```

surface_courante $\leftarrow S_0$;

pour $i = 1$ **à** $N - 1$ **faire**

Suppression des triangles incohérents de surface_courante ;

Suppression des triangles correspondants de S_i;

Retriangulation des trous formés dans surface_courante ;

Création de triangles pour raccorder les surfaces surface_courante et S_i;

surface_courante $\leftarrow$ surface_courante $\bigcup S_i \bigcup raccord$;

fin

3.5.4 Résultats

Dans cette partie, nous montrons des résultats de fusion obtenus sur des données réelles. Les images maîtresses de départ sont celles montrées figure 3.12. Les figures 3.13 à 3.19 illustrent la fusion itérative des surfaces.

(a) (b) (c) (d)

FIG. 3.12 – Images maîtresses utilisées pour produire les résultats présentés

En détail, la surface initiale S_1 est montrée figure 3.13, elle s'appuie sur l'image maitresse de la figure 3.12(d). Les points 3D du nuage voisin N_2, en recouvrement partiel avec S_1 sont représentés figure 3.14. Ces points s'appuient sur l'image maîtresse 3.12(c).

Le parcours des points 3D de N_2 donne lieu au filtrage de triangles incohérents de la surface initiale S_1 (s'il y en a), ne restent à l'issue de cette étape que les points qui ne sont pas en recouvrement avec S_1 (voir figure 3.15), ainsi que le maillage triangulé associé (figure 3.16).Parmi ces points 3D, on s'intéresse aux points situés à la frontière des deux maillages à partir desquels le maillage de raccord est alors créé (voir figure 3.17).

Enfin, la figure 3.18 montre la surface fusionnée issue des deux points de vue. De manière itérative, les autres points de vue sont progressivement fusionnés, complétant la façade (figure 3.19).

Les illustrations sont présentées ici avec des modèles non texturés afin de mieux percevoir la triangulation.

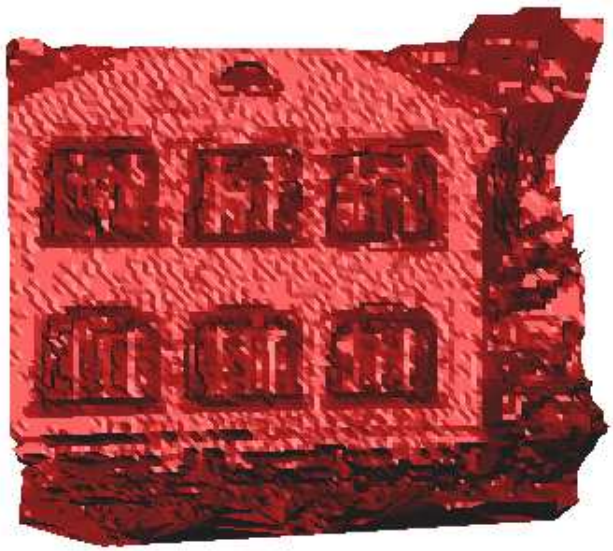


FIG. 3.13 – Surface initiale S_1.

3.6 Bilan

Dans cette partie, nous avons présenté une méthode pour passer des cartes de profondeurs en géométrie image à une surface unifiée et triangulée. Cette méthode s'articule en plusieurs points :

- le passage d'une carte de profondeur à un nuage de points. Ceci repose sur un indice de confiance du point de la carte, estimé à partir de la texture locale de l'image et du score de corrélation obtenu. Cela permet de filtrer les points peu fiables : ciel, zones spéculaires, zones trop uniformes.
- la triangulation de ce nuage, guidée par la géométrie de l'image source.
- la fusion par une méthode surfacique et itérative.

Le temps de calcul est donne une moyenne de 26 minutes par image source ajoutée ; ce temps est essentiellement du au lancer de rayon utilisé pour la fusion.

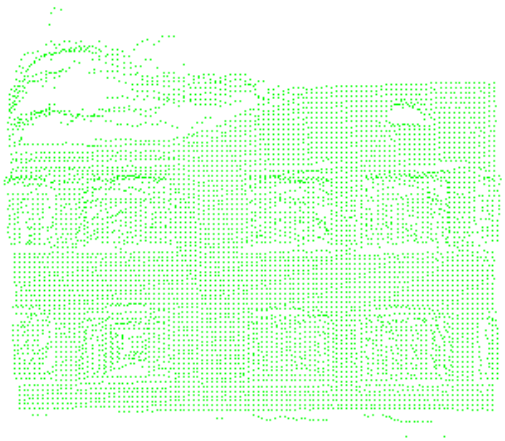


FIG. 3.14 – Points du nuage N_2.

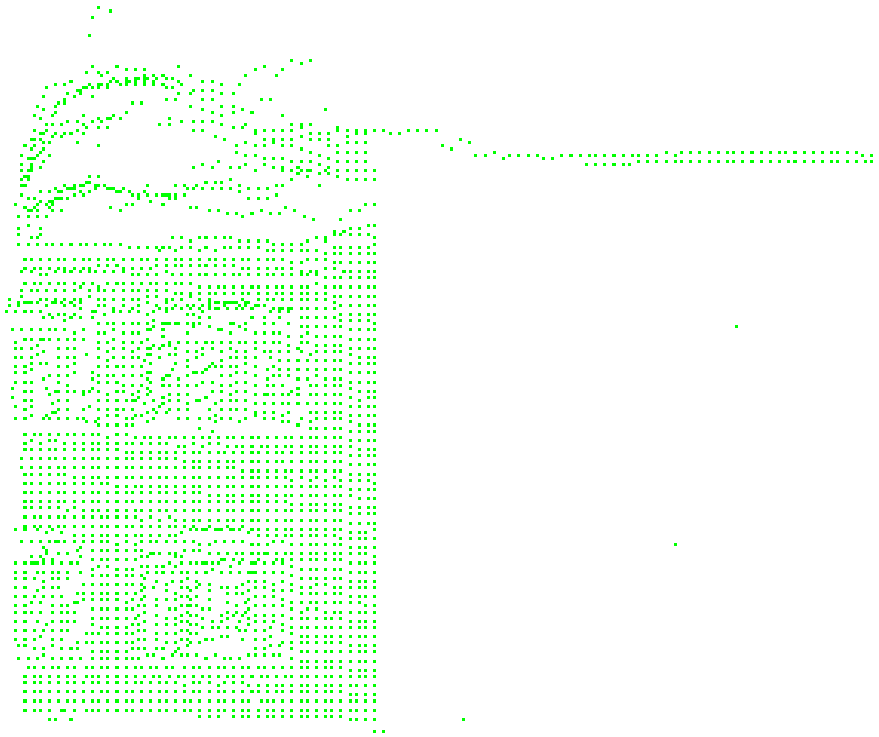


FIG. 3.15 – Points du nuage N_2 qui ne sont pas en recouvrement avec la surface initiale S_1.

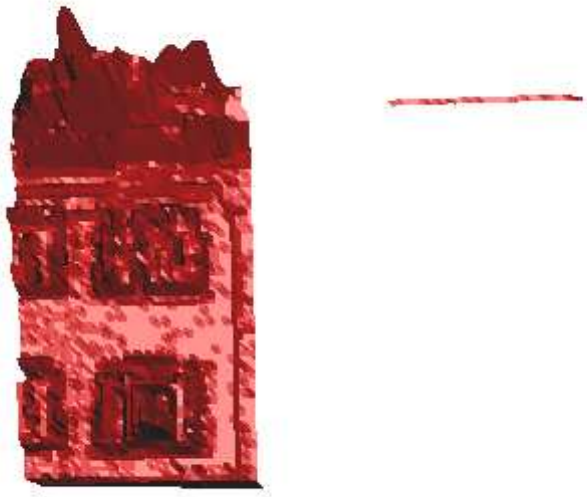


FIG. 3.16 – Surface S_2 triangulée après suppression de triangles en recouvrement avec S_1. Ici, cette surface présente deux composantes connexes.

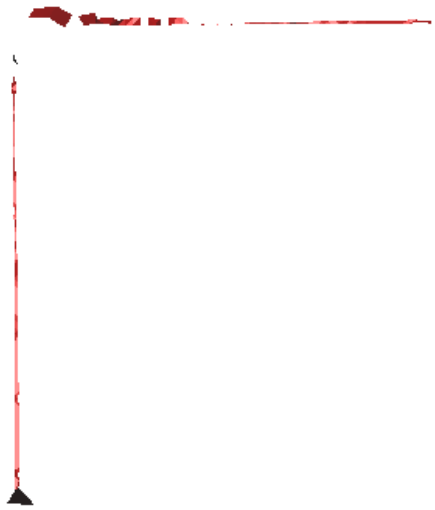


FIG. 3.17 – Triangles du raccord créés à l'interface entre les surfaces S_1 et S_2.

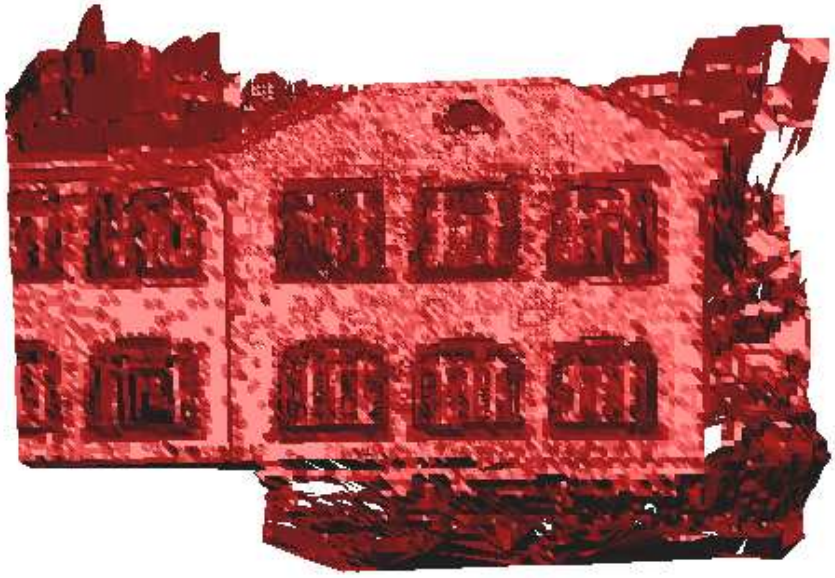


FIG. 3.18 – Résultat de l'algorithme de fusion de S_1 avec S_2.

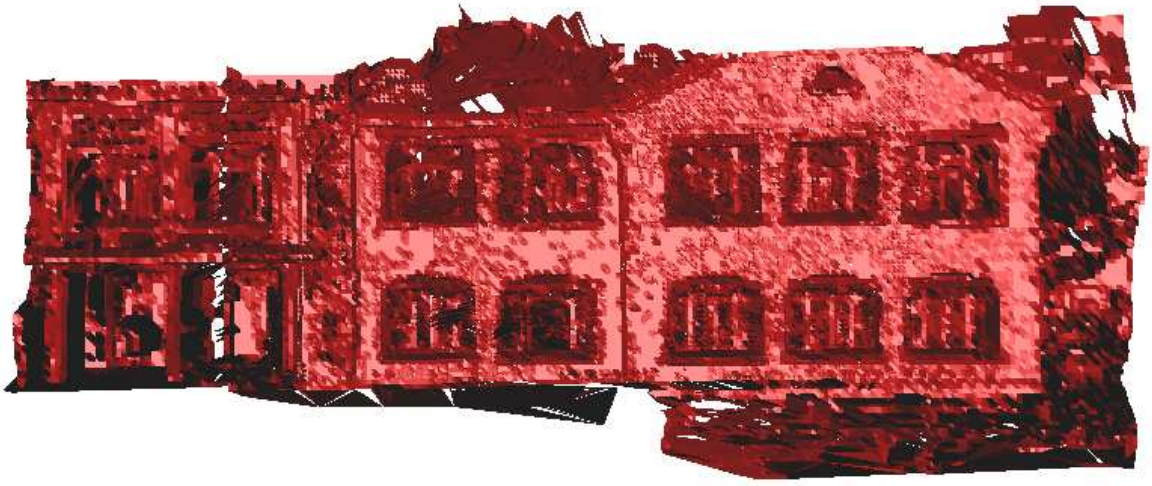


FIG. 3.19 – Résultat de l'algorithme de fusion en intégrant plusieurs surfaces.

CHAPITRE 4

VERS UNE GESTION DES PARTIES CACHÉES POUR L'APPARIEMENT

4.1 Introduction

Plusieurs images d'une même scène, prises depuis des points de vue distants, présentent des similarités et des différences. L'algorithme décrit au chapitre 2 repose sur la corrélation et donc sur la similarité locale des images. Plus les points de vue sont éloignés, moins cette similarité est grande. En effet, les images sont modifiées par des facteurs géométriques d'une part et radiométriques d'autre part.

La déformation géométrique est due au changement de perspective : la forme apparente d'un objet varie en fonction de la position d'observation. L'effet radiométrique est dû au caractère non lambertien des matériaux, c'est à dire que la rediffusion de la lumière par les objets n'est pas uniforme selon la direction d'observation.

Cependant, un autre effet majeur du changement de point de vue réside dans la présence d'occultations partielles de la scène. En effet, à moins d'observer un objet quasi-plan, on trouvera toujours un point de vue pour lequel une partie de la scène en masque une autre, placée derrière. Dans une telle image, il sera donc impossible de trouver un homologue pour les points images issus de détails appartenant à la zone cachée. Le schéma de la figure 4.1 illustre l'existence de parties cachées dans les images qui peut conduire à une reconstruction erronée de la surface. Sur cette figure, la reconstruction à partir de l'image maîtresse représentée en bleu au centre est perturbée par l'image de gauche dans laquelle le point terrain à reconstruire est occulté.

Les occultations représentent donc un problème majeur et souvent délicat, dans le cadre de la mise en

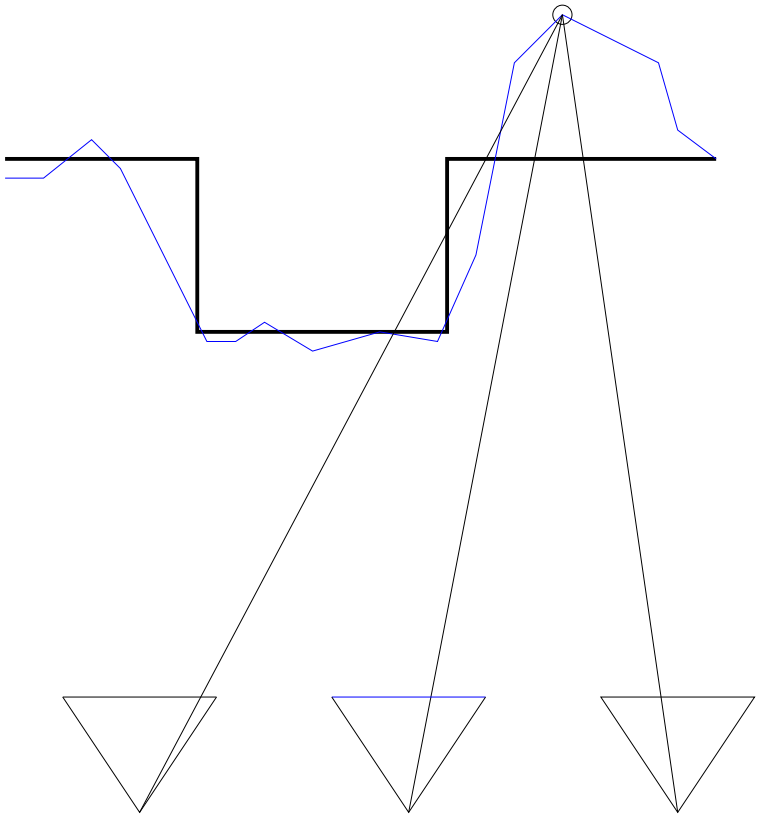


FIG. 4.1 – Surface reconstruite erronée

correspondance par corrélation. De nombreuses propositions ont été faites dans la littérature pour pallier ce problème.

4.2 Etat de l'art

De manière générale en vision, la prise en compte des occultations est une préoccupation majeure car elles perturbent le fonctionnement d'algorithmes automatiques. Un grand nombre de méthodes ont été proposée qui les prennent en compte de manière intrinsèque. On peut citer la technique du space carving (KUTULAKOS et SEITZ 2000), ou du voxel coloring (SEITZ et DYER 1999) déjà détaillées dans la partie 1. Pour ces deux méthodes, le volume englobant l'objet à reconstruire est respectivement creusé et colorié, jusqu'à obtenir la surface recherchée.

Dans le cadre particulier de la mise en correspondance par corrélation, une revue détaillée de mé-

thodes existantes dans la littérature est présentée dans CHAMBON 2005. Nous adoptons la même typologie de ces méthodes : méthodes avec prétraitement, méthodes locales, méthodes globales, méthodes à passages multiples et méthodes avec affinement.

Méthodes avec prétraitement

Une première catégorie d'approches repose sur la détection *a priori* de zones d'occultations, étape qui constitue donc un prétraitement visant à obtenir un masque des occultations. En effet, celles-ci sont délimitées par un changement de profondeur de la scène, souvent associé au franchissement d'un contour dans l'image. Les techniques de détection de contour classiques, ou de segmentation en régions sont alors utilisées (voir par exemple HOFF et AHUJA 1989).

Méthodes locales

D'autres approches locales utilisent des mesures de corrélation plus complexes. On peut par exemple citer la mesure *Increment Sign Correlation* ou encore *Selective Correlation Coefficient* (KANEKO et al. 2002 ; KANEKO et al. 2003) et enfin les mesures de rang (ZABIH et WOODFILL 1994). Ces mesures sont des transformations non paramétriques. Elles ont la particularité d're très robustes aux occultations.

Méthodes globales

Des méthodes globales ont également été développées. Dans ces méthodes, la recherche de la carte de profondeur se fait par minimisation d'une énergie somme d'un terme de correspondance (attache aux données) et d'un terme de lissage. Dans ce dernier terme est intégré un coût supplémentaire correspondant aux contraintes des occultations. Ce coût peut être basé sur la contrainte d'unicité (par exemple ISHIKAWA et GEIGER 1998), sur la contrainte de symétrie (par exemple BELHUMEUR et MUMFORD 1992), sur les sauts de disparités avec les voisins (par exemple TORR et CRIMINISI 2004).

Méthodes à passages multiples

Enfin, les approches à passages multiples sont assez populaires. Les algorithmes utilisant des fenêtres adaptatives en sont un premier exemple. Après une première passe normale, l'idée est de modifier la taille

et/ou la forme des fenêtres de corrélation pour que celles-ci ne contiennent pas de contours (par exemple SCHARSTEIN et SZELISKI 1998 ; ZHANG et KAMBHAMETTU 2002) ceci afin d'éviter les discontinuités de surface à l'intérieur de la fenêtre de corrélation. D'autres approches proposent de pondérer les mesures de corrélation obtenues. Une première carte de profondeur permet de calculer les pondérations des corrélations utilisées dans le calcul de la carte de profondeur finale.

Méthodes avec affinement

Ces méthodes, très courantes, tentent de déterminer les occultations. Plusieurs approches ont été proposées (FAUGERAS et al. 1993 ; SZELISKI et SCHARSTEIN 2002, utilisant par exemple la contrainte de symétrie, l'ambiguïté, la contrainte d'ordre, la contrainte d'unicité ou encore un seuillage sur le score de corrélation. Le principe est de déclarer occulté tout pixel ne vérifiant pas la condition fixée, ce qui permet de calculer la carte des occultations en même temps que la carte de profondeur sur les pixels non occultés.

Méthodes à deux coefficients

D'autres méthodes (CHAMBON et CROUZIL 2004) utilisent deux coefficients de corrélation différents suivant si la zone traitée est sous l'influence d'occultations (mesure robuste) ou non (mesure classique de type ZNCC).

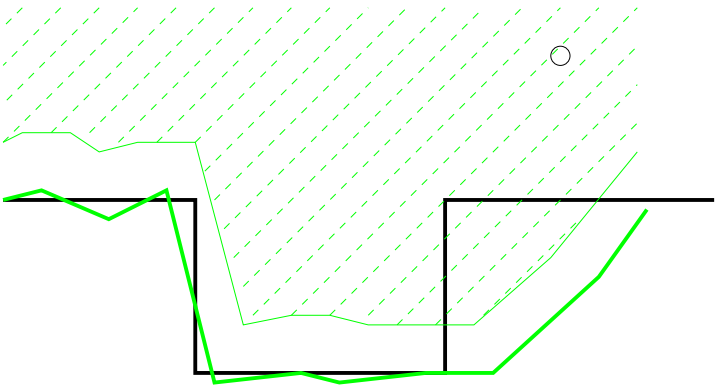
4.3 Algorithme

Les méthodes de prise en compte des effets des occultations qui utilisent un schéma à passages multiples peuvent donner de très bons résultats. En revanche, elles ont l'inconvénient d'augmenter le temps de calcul de façon importante du fait des multiples passages. Cependant, dans le cadre d'un processus multi-résolution, il est tout à fait naturel d'introduire ce type d'algorithme qui n'augmente pas significativement le temps de calcul, car les passages multiples sont déjà faits en pratique à des résolutions croissantes. Nous proposons donc une méthode dont le principe est le suivant : à une résolution donnée le calcul de la carte de profondeur prendra en compte le résultat de la résolution immédiatement inférieure

pour prédire les zones d'occultations. Pour rappel, nous utilisions déjà cette information mais seulement pour prédire la zone de recherche.

4.3.1 Calcul des zones d'occultation

Le cas où le meilleur score de corrélation n'est pas obtenu exactement au point réel de la scène est fréquent. La régularisation de surface (présentée chapitre 2, section 2) permet de pallier ces défauts de reconstruction dans une grande partie des cas. En revanche, lorsque le phénomène est dû à une occultation dans une ou plusieurs images du point de la scène, l'algorithme ne parvient pas dans tous les cas à une solution satisfaisante. La figure 4.5(a) à la fin de ce chapitre en donne un exemple sur données réelles, exemple illustré par le schéma de la figure 4.2.



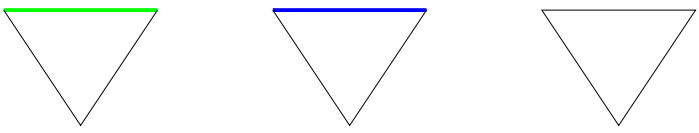


FIG. 4.2 – Zone d'occultation calculée pour l'image de gauche (en vert)

Dans un premier temps, nous nous attachons à décrire le principe du calcul des zones d'occultation

étant donné une carte de profondeur initiale. Cette carte de profondeur initiale est une solution approchée de la surface à reconstruire. Dans notre contexte multi-échelle, elle est fournie par le calcul de la carte de profondeur à la résolution immédiatement inférieure. Toutefois, elle peut également provenir d'une base de données exogène (limites des zones bâties sur le plan cadastral par exemple) ou d'une utilisation d'un scanner balayage laser embarqué lors de l'acquisition des données.

Pour une image donnée, nous définissons une nouvelle carte de profondeur, appelée limite de visibilité, construite à partir de la carte de profondeur initiale par dilatation d'une part dans les deux directions de la façade, et d'autre part dans le sens des profondeurs croissantes dans la direction de la visée image.

Le schéma de la figure 4.3 illustre cette dilatation. A partir de la surface initiale reconstruite en utilisant l'image verte comme image maîtresse, nous construisons la limite de visibilité associée. Ainsi toute la zone hachurée représente le volume dans lequel les calculs de corrélation ne feront pas intervenir l'image représentée en vert.

Nous choisissons de faire correspondre cette limite de visibilité avec la limite maximale de la zone de recherche présentée au chapitre 2, dans le cadre du passage de l'algorithme en multirésolution. Dans ces conditions, l'obtention de la limite de visibilité se fait donc sans aucun calcul supplémentaire.

4.3.2 Modification des scores de corrélation

Une fois que les zones de visibilité et d'occultation sont définies, nous pouvons injecter cette connaissance dans le calcul des coefficients de corrélation. Le calcul des scores de corrélation multi-images pour un voxel v dans la géométrie de l'image maîtresse i a été défini au chapitre 2 comme suit :

$$cor_i(v) = \frac{1}{\#\mathscr{I}_v} \sum_{j \in \mathscr{I}_v, j \neq i} ZNCC(\mathbf{V_i}, \mathbf{V_j}) \tag{4.1}$$

où $\mathscr{I}_v$ est l'ensemble des images qui voient le voxel v. Pour tenir compte des occultations, nous restreignons $\mathscr{I}_v$ au sous-ensemble $\mathscr{I}_v^{occ}$ des images dans lesquelles v n'est pas occulté. Un voxel v sera dit occulté dans une image s'il est au delà de la zone de visibilité, construite par dilatation de la carte de

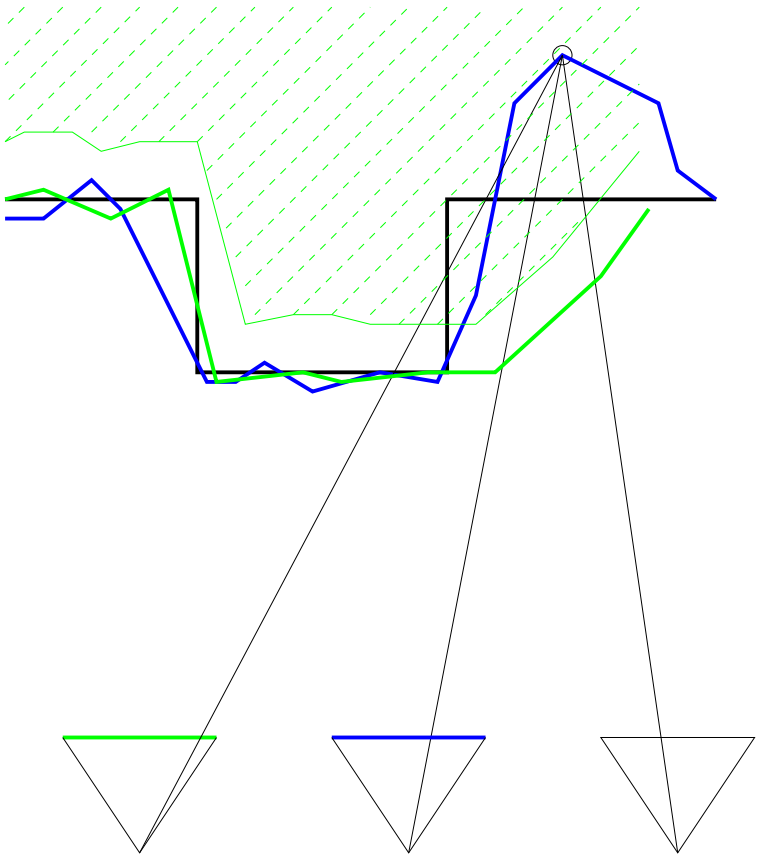


FIG. 4.3 – L'image gauche est eclue de l'ensemble des images visibles : elle ne pertubera pas les calculs de reconstruction

profondeur calculée dans sa géométrie image maîtresse. En pratique, nous utilisons l'algorithme 4.1.

Algorithme 4.1 : Détermination des images dans lesquelles un voxel v est occulté

Données : voxel, ensemble $\mathscr{I}_{voxel}$, cartes de profondeur des images à la résolution inférieure

Sorties : ensemble $\mathscr{I}^{occ}_{voxel}$

$\mathscr{I}^{occ}_{voxel} = \emptyset$;

pour tous les $im \in \mathscr{I}_{voxel}$ **faire**

```
// Reprojection du voxel en coordonnées subpixelliques
```

$(colonne, ligne) \leftarrow$ `reprojection`(im, v);

si `dilatation`$(im, colonne, ligne) >$ `distance`(im, v) **alors**

$\mathscr{I}^{occ}_{voxel} \leftarrow \mathscr{I}^{occ}_{voxel} \cup \{im\}$

fin

fin

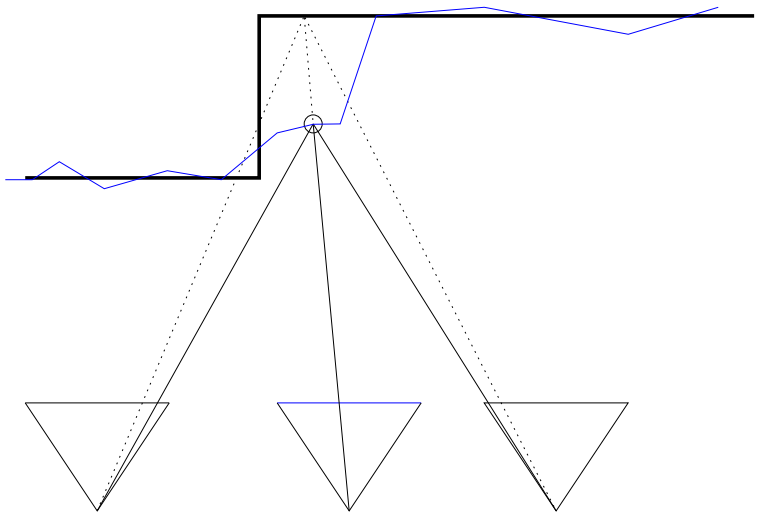


FIG. 4.4 – Parties cachées

L'algorithme 4.2 illustre le fait que le processus global n'est affecté qu'à la marge par la prise en compte des parties cachées (voir la version précédente, algorithme 2.2). Elle n'intervient que dans la détermination du sous-ensemble d'images à prendre en compte dans le calcul du score de corrélation, qui ne coûte qu'une reprojection dans les cartes de profondeurs dilatées issues des images voisines à la résolution inférieure. L'étape de calcul de dilatation étant déjà implémentée pour délimiter l'espace de recherche, elle n'ajoute aucun calcul si les paramètres de dilatation sont choisis identiques (choix que nous faisons).

4.4 Résultats

Dans cette partie, nous présentons quelques résultats obtenus avec la gestion des parties cachées. L'image 4.5 montre deux cartes de profondeurs en niveaux de gris. La première a été réalisée sans tenir compte des occultations, la seconde utilise la limite de visibilité calculée à partir des cartes de profondeurs de la résolution inférieure (voir image 4.7). Elle est produite à partir de l'image maîtresse représentée image 4.6(b), et se situe à l'extrémité de la façade à reconstruire du bâtiment. Le mur perpendiculaire, vu sous un angle proche de 90 degrés, n'est donc pas vu sur le couple d'images prises lorsque le véhicule était quelques mètres plus à gauche (voir image 4.6(a)). Cette occultation entraîne une mauvaise estima-

Algorithme 4.2 : Stratégie d'appariement dense et d'estimation des parties cachées en multi-résolution

Données : $\mathscr{I}$, ensemble de N images géoréférencées de la façade
Sorties : Carte de profondeur pour chaque image

pour tous les $im \in \mathscr{I}$ **faire**
  Création de la pyramide d'images sous-échantillonnées jusqu'à la résolution resol_min
fin
pour resol = resol_min **à** resol_max **faire**
  pour tous les $im \in \mathscr{I}$ **faire**
    `// im est l'image maîtresse`
    si resol ≠ resol_min **alors**
      Calcul de la zone de recherche par dilatation de nappe_courante
    fin
    Découpage de l'espace image en dalles;
    pour tous les *dalles* **faire**
      pour tous les *voxels* $v \in$ *zone de recherche* **faire**
        Détermination de $\mathscr{I}_v \subset \mathscr{I}$, sous-ensemble des images voyant le voxel v par prise en compte des occultations;
        Calcul des scores de corrélation $cor_{im}(v)$ restreint à $\mathscr{I}_v$;
      fin
      Optimisation;
    fin
    nappe_courante ← `assemblage_dalles()` ;
  fin
  resol = resol + 1 ;
fin

tion de la surface à reconstruire (image 4.5(a)). En ajoutant la contrainte de zone de visibilité, le résultat obtenu est nettement amélioré (image 4.5(b)).

(a) Surface obtenue sans gestion des parties cachées. La façade fuyante au centre de l'image est mal reconstruite car elle est masquée par la façade principale dans plusieurs images.

(b) Surface obtenue avec gestion des parties cachées. La façade fuyante est mieux reconstruite.

FIG. 4.5 – Exemple d'amélioration de la description de la surface par prise en compte des occultations. L'image maîtresse est représentée figure 4.6(b)

FIG. 4.6 – Images maîtresses

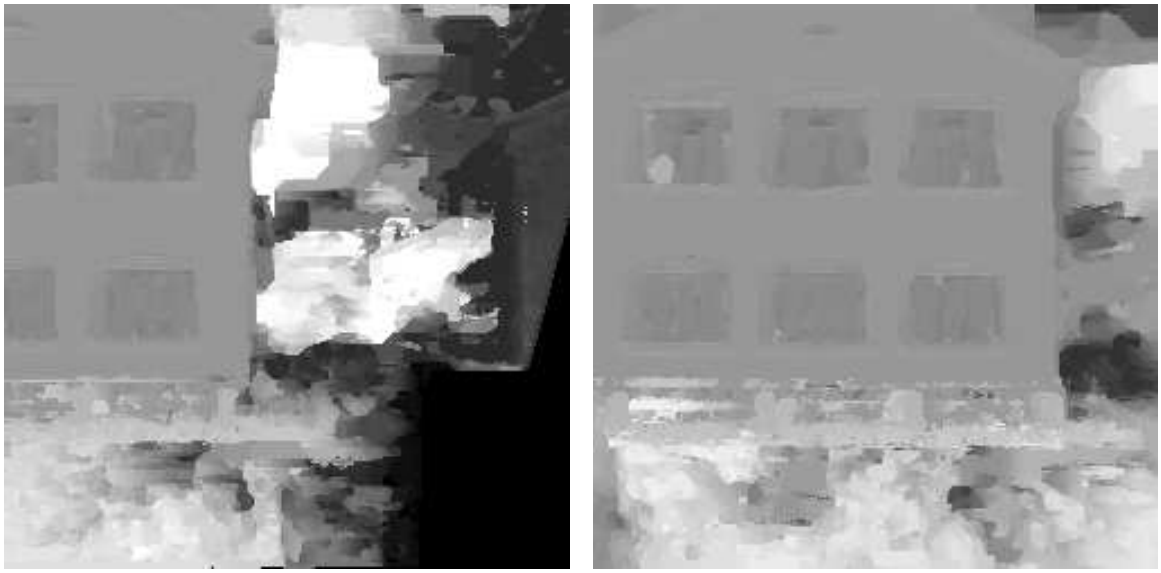


FIG. 4.7 – Carte de profondeur à la résolution inférieure

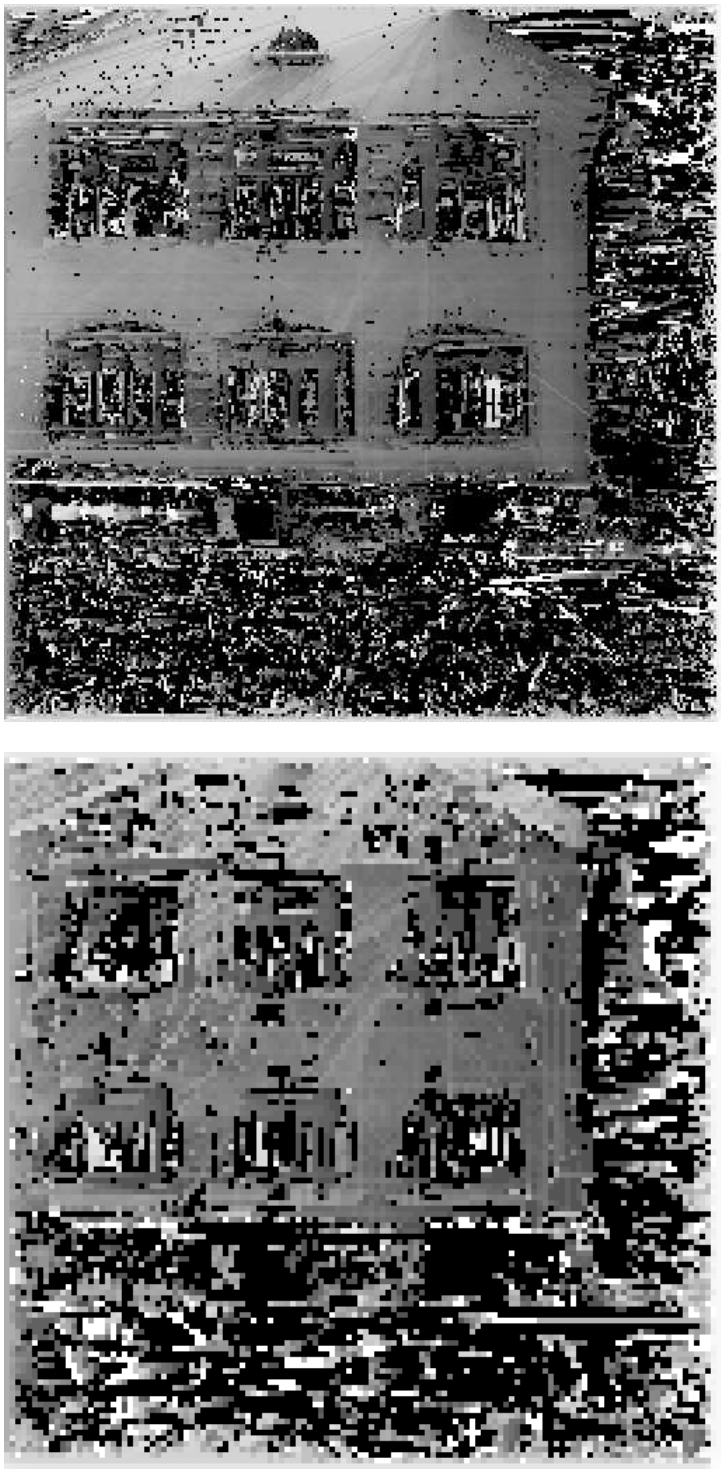


FIG. 4.8 – Résultats de reconstruction avec une résolution croissante (8 *cm*, 4 *cm*, 2*cm*, 1 *cm* et 5 *mm* au niveau de la façade). Modèles présentés en vue ombrée.

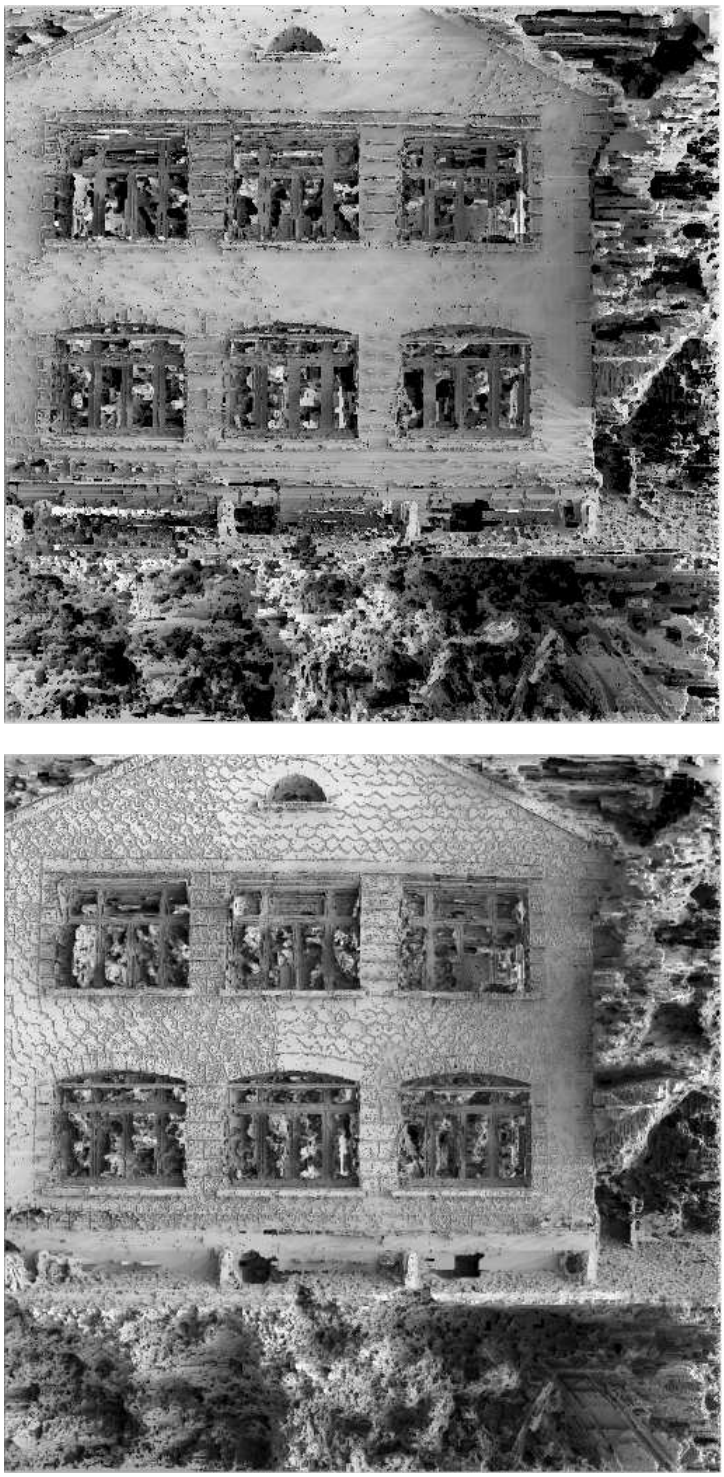


FIG. 4.8 – Résultats de reconstruction avec une résolution croissante (8 *cm*, 4 *cm*, 2*cm*, 1 *cm* et 5 *mm* au niveau de la façade). Modèles présentés en vue ombrée.

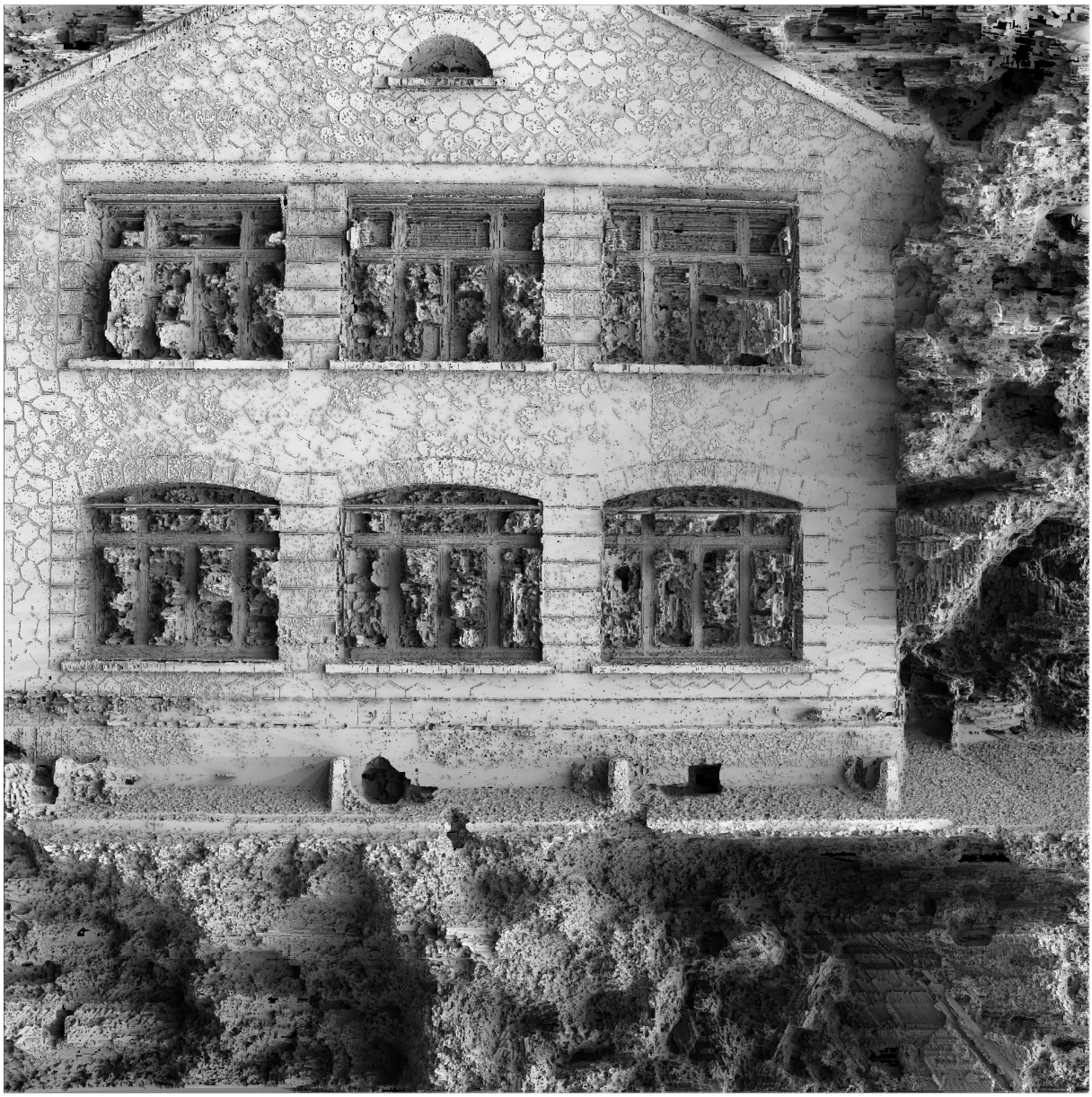
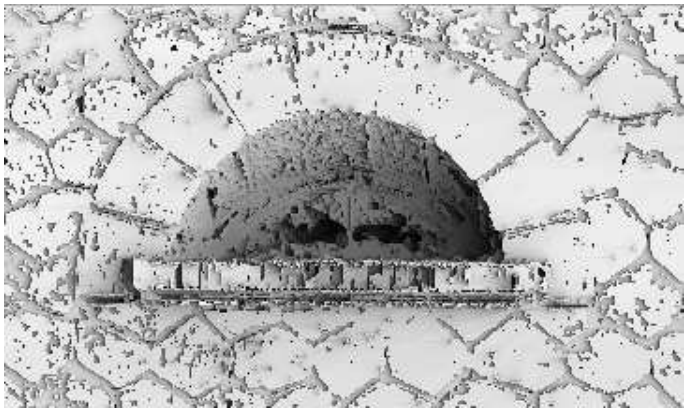
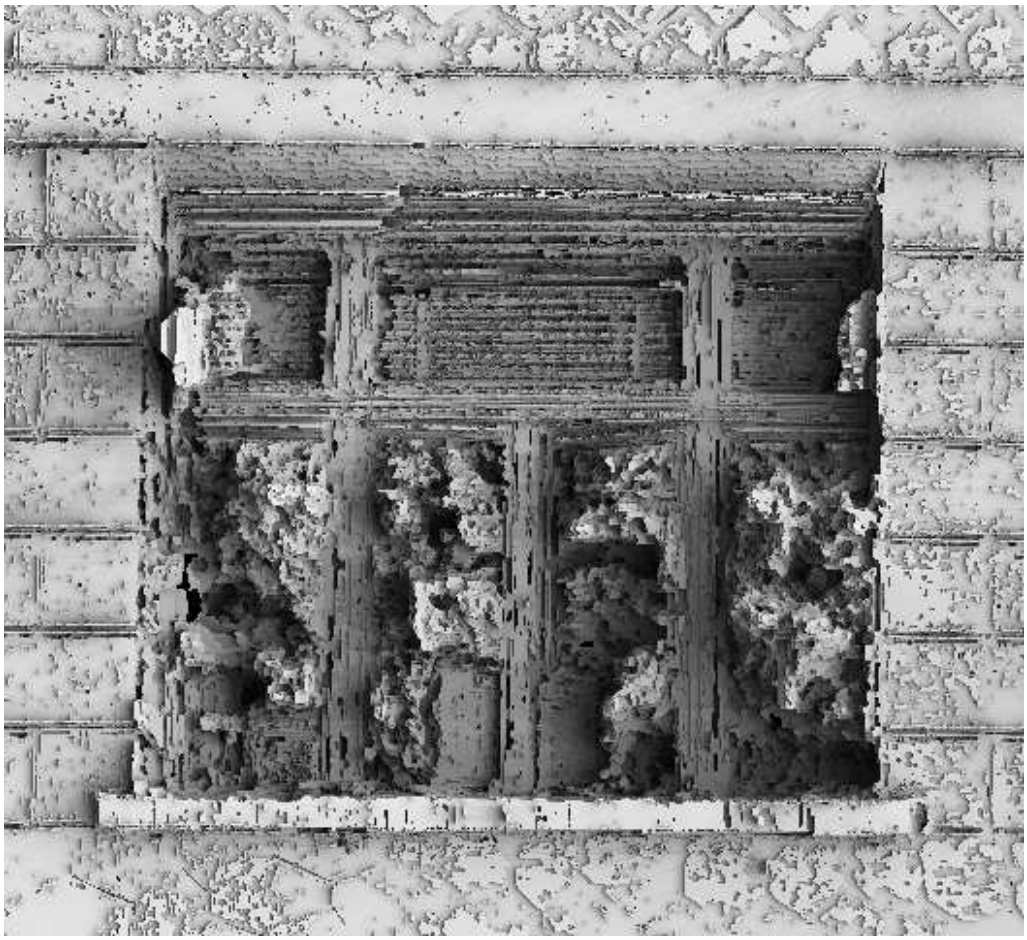


FIG. 4.8 – Résultats de reconstruction avec une résolution croissante (8 *cm*, 4 *cm*, 2*cm*, 1 *cm* et 5 *mm* au niveau de la façade). Modèles présentés en vue ombrée.



(a) Lucarne



(b) Fenêtre complète. Montants et rebords sont correctement reconstruits. L'intérieur des vitres est bruité : surface spéculaire et transparente.

FIG. 4.9 – Résultats de reconstruction à résolution maximale : zoom sur quelques détails.

(c) Coin bas droit du bâtiment. L'avant d'une voiture en stationnement est également reconstruit malgré l'occultation partielle (2 vues sur 6).

FIG. 4.9 – Résultats de reconstruction à résolution maximale : zoom sur quelques détails.

CONCLUSION

Le travail présenté dans ce document propose une méthode de reconstruction de façades de bâtiments, en multi-stéréoscopie. Nous suivons une stratégie basée sur l'appariement dense d'images, avec une régularisation par coupe de graphe. Nous proposons une méthode de fusions de surfaces pour unifier la reconstruction à l'échelle d'une façade. Nous avançons également en direction d'une meilleure gestion des parties cachées par nos algorithmes.

Contributions principales

Reconstruction par appariement dense d'images

Le système mobile utilisé comporte une base stéréoscopique verticale. Nous avons exploité cette caractéristique en proposant une reconstruction avec ces deux images pour chacune des prises de vue. Nous étendons ensuite la méthode avec des images multiples. Nous avons mis en évidence la plus grande qualité des résultats dans la géométrie image par rapport à la géométrie objet.

La stratégie suivie utilise une régularisation de surface par coupe de graphe, basée sur l'algorithme de ROY et COX 1998, pour rendre moins bruitée la surface obtenue en prenant directement le maximum du score de corrélation. De plus, afin de réduire le temps de calcul et la place mémoire requise, le processus s'utilise en multi-résolution. Les résultats obtenus montrent une bonne reconstruction des façades de bâtiments. La résolution de reconstruction n'est pas dégradée par notre méthode, nous nous rapprochons donc de celles des images : dans les exemples présentés, la résolution est meilleure que le centimètre.

Fusion de surfaces

Nous avons développé une stratégie de fusion de surfaces triangulées. Cette fusion se classe dans la catégorie des méthodes surfaciques, bien adaptées pour les nuages structurés. Elle permet de fusionner les cartes de profondeurs estimées dans chaque géométrie image en une surface unique par façade. Les

résultats montrent le bon comportement de l'algorithme, au prix d'un temps relativement long du fait d'un lancer de rayon.

Gestion des parties cachées

Enfin, nous proposons une méthode pour limiter les effets des parties cachées dans les images. Notre méthode exploite l'algorithme multi-résolution développé pour fournir un indicateur des volumes de la scène masqués dans les images à la résolution supérieure. Cela permet d'améliorer les résultats obtenus précédemment.

Perspectives

Dans un premier temps, nous dégageons des perspectives d'amélioration directe des résultats de nos travaux.

Le véhicule d'acquisition du projet Stéréopolis a été profondément remanié au cours de la dernière année. L'augmentation du débit d'acquisition s'est accompagnée d'une réduction de la taille des capteurs et d'une augmentation du nombre d'images par façade de bâtiment. Pleinement opérationnel, le véhicule peut maintenant faire des acquisitions à l'échelle de villes entières. Or, une des faiblesses des algorithmes présentés réside dans le temps de calcul important. Trois étapes sont particulièrement longues : calcul des scores de corrélation, régularisation par coupe de graphe, fusion des surfaces triangulées. En ce qui concerne les deux premières étapes, il est possible de réaliser les calculs par un système distribué car nous gérons le découpage en dalles indépendantes. Le gain de temps escompté dépend directement du nombre de processeurs disponibles. L'étape de fusion fait appel à un lancer de rayon, ce qui est lent comme nous l'avons dit. Si l'objectif de rapidité d'exécution devient prépondérant, il sera nécessaire de modifier la méthode de fusion.

Dans le cadre d'un objectif de modèle destiné à la visualisation, le seul critère de faible écart géométrique avec la réalité n'est pas suffisant. L'oeil humain est très sensible aux ruptures de contours par exemple. Remplacer une ligne droite de contour par une reconstruction en ligne brisée autour de la ligne théorique est particulièrement perturbant. Nous proposons donc d'introduire, dans l'étape de production

des maillages triangulés, les contours détectés dans les images et reconstruits en trois dimensions et non plus seulement les points. Ainsi, une triangulation s'appuyant sur ces contours et ces points fournira une qualité de rendu supérieure.

Pour aller plus loin Tout d'abord, comme nous l'avons mentionné au chapitre 1, la grande majorité des équipes qui travaillent sur la cartographie mobile automatique utilisent un ou plusieurs laser embarqués pour fournir la géométrie de la scène. La nouvelle version du véhicule Stéréopolis sera équipée de tels laser, ce qui amène naturellement à réfléchir à l'intégration de données de ce type dans le processus décrit ici. Nous voyons plusieurs intérêts à utiliser les points laser. En premier lieu, ils peuvent bien sûr fournir une reconstruction "basse résolution" (par rapport à ce que l'on peut obtenir avec des images) sans traitement particulier. Mais aussi, ils peuvent venir améliorer la présente chaîne de traitements. Nous pensons qu'ils pourraient fournir un excellent prédicteur de la position des façades en entrée de la reconstruction multi-résolution. Leur apport serait moins grand en terme de temps de calcul qu'en terme de robustesse. En effet, les étapes aux basses résolution image que l'on pourrait éviter sont les moins gourmandes en temps de calcul. En revanche, la qualité homogène des points laser permettrait de pallier les situations où la reconstruction image à basse résolution est trop mauvaise localement pour que le processus finisse par converger vers la bonne estimation de la profondeur à la résolution maximale. L'utilisation d'un prédicteur issu des points laser nous semble donc favorable à une amélioration de la robustesse des résultats.

Ensuite, nous proposons de dépasser le stade de la description topographique des façades pour aller vers une véritable modélisation en éléments simples : facettes planes de façades, fenêtres, portes, balcons,... L'avantage de tels modèles est leur caractère compact et l'intégration plus facile de niveaux de détails lors de la visualisation.

Nous proposons également d'unifier les reconstructions de façades – typiquement faite à partir d'images terrestres – et les reconstructions de toits – à partir de données aériennes. Cette unification permettrait de produire des modèles de bâtiments complets avec une résolution bien adaptée à la visualisation : très forte sur les façades, vues à faible distance lors d'un parcours virtuel dans les rues, moyenne sur les toits, vus à plus grande distance lors d'un survol de la zone reconstruite. C'est bien une unification

(pas simplement une juxtaposition) qui s'avère nécessaire pour régler les problèmes de raccordements entre les modèles essentiellement dus à la différence de résolution des observations. Les reconstructions de façades peuvent par exemple servir de prédicteur pour l'emplacement des limites de toit, permettant une meilleure localisation et évitant aussi une trop grande généralisation des formes de toit.

BIBLIOGRAPHIE

ABUHADROUS, Iyad, Samer AMMOUN, Fawzi NASHASHIBI, François GOULETTE et Claude LAURGEAU, « Digitizing and 3D Modelling of Urban Environments using Vehicle-Borne Laser Scanner System », dans : *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004)*, Sendai, Japan 2004.

AKBARZADEH, Amir, Jan-Michael FRAHM et al., « Towards Urban 3D Reconstruction From Video », dans : *Proceedings 3DPVT'06 (Int. Symp. on 3D Data, Processing, Visualization and Transmission)*, Chapel Hill, NC, USA 2006.

ALLEN, Peter, Ioannis STAMOS, Atanas GUEORGUIEV, Ethan GOLD et Paul BLAER, « AVENUE : Automated Site Modeling in Urban Environments », dans : *Proc. of 3rd Conference on Digital Imaging and Modeling*, Quebec, Canada 2001, p. 357–364.

BAUER, Joachim, Andreas KLAUS, Konrad KARNER, Christopher ZACH et Konrad SCHINDLER, « METROPOGIS : A Feature Based City Modeling System », dans : *PCV '02 : Proceedings of Photogrammetric Computer Vision*, 2002, B :22.

BELHUMEUR, Peter et David MUMFORD, « A Bayesian Treatment of the Stereo Correspondence Problem Using Half-Occluded Regions », dans : *IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR -92*, Urbana-Champaign, Etats-Unis 1992, p. 506–512.

BEN-JEMAA, Raouf, « Traitement de données 3D denses acquises sur des objets réels complexes », Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications, 1998.

BENTRAH, Ouided, « Reconstruction de la géométrie d'acquisition d'une séquence d'images acquise par un véhicule de cartographie mobile terrestre », Thèse de doctorat, Institut National Polytechnique de Grenoble, 2006.

BENTRAH, Ouided, Nicolas PAPARODITIS et Marc PIERROT-DESEILLIGNY, « STEREOPOLIS : An Image Based Urban Environments Modeling System », dans : *MMT 2004. The 4th International Symposium on Mobile Mapping Technology, Kunming, China*, 2004.

BOISSONNAT, Jean-Daniel, « Geometric structures for three-dimensional shape representation », dans : *ACM Trans. Graph.* 3.4 (1984), p. 266–286.

BRENNER, Claus, Norbert HAALA et Dieter FRITSCH, « Towards fully automated 3D city model generation », dans : *Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images III*, 2001.

BRETAR, Frédéric et Michel ROUX, « Recognition of Building Roof Facets by Merging Aerial Images and 3D Lidar Data in a Hierarchical Segmentation Framework », dans : *Proc. of the International Conference on Pattern Recognition*, Hong-Kong, China 2006.

BRUN, Xavier, Jean-Emmanuel DESCHAUD et François GOULETTE, « On-the-way city mobile mapping using laser range scanner and fisheye camera », dans : *Mobile Mapping Technologies*, Padoue, Italie 2007.

CHAMBON, Sylvie, « Mise en correspondance stéréoscopique d'images couleur en présence d'occultations », Thèse de doctorat, Université Paul Sabatier - Toulouse III, 2005.

CHAMBON, Sylvie et Alain CROUZIL, « Towards correlation-based matching algorithms that are robust near occlusions », dans : *IEEE Conference Proceedings of International Conference on Pattern Recognition, ICPR '04*, t. 3, Cambridge, Royaume-Uni 2004, 20–23.

COORG, Satyan, Neel MASTER et Seth TELLER, « Acquisition of a Large Pose-Mosaic Dataset », dans : *CVPR '98 : Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA : IEEE Computer Society, 1998, p. 872.

CURLESS, Brian et Marc LEVOY, « A volumetric method for building complex models from range images », dans : *SIGGRAPH '96 : Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, New York, NY, USA : ACM, 1996, p. 303–312.

DEBEVEC, Paul, Camillo TAYLOR et Jitendra MALIK, « Modeling and rendering architecture from photographs : a hybrid geometry- and image-based approach », dans : *SIGGRAPH '96 : Proceedings of*

the 23rd annual conference on Computer graphics and interactive techniques, New York, NY, USA : ACM, 1996, p. 11–20.

DICK, Anthony, Philip TORR et Roberto CIPOLLA, « Modelling and Interpretation of Architecture from Several Images », dans : *Int. J. Comput. Vision* 60.2 (2004), p. 111–134.

FAUGERAS, Olivier, Bernard HOTZ, Zhengyou ZHANG et Pascal FUA, *Real time correlation-based stereo : Algorithm, implementation and applications*, Rapport de recherche, Institut National de Recherche en Informatique et en Automatique, 1993.

FORD, Lester Randolph et Delbert Ray FULKERSON, *Flows in Networks*, Princeton University Press, 1962.

FRÜH, Christian, « Automated 3D model generation for urban environments », Thèse de doctorat, University of Karlruhe, Germany, 2002.

FRÈRE, David, Jan VANDEKERCKHOVE, Theo MOONS et Luc VANGOOL, « Automatic modelling and 3D reconstruction of urban buildings from aerial imagery », dans : *Geoscience and Remote Sensing Symposium Proceedings*, t. 5, 1998, p. 2593 –2596.

FRÜH, Christian et Avideh ZAKHOR, « An Automated Method for Large-Scale, Ground-Based City Model Acquisition », dans : *Int. J. Comput. Vision* 60.1 (2004), p. 5–24.

FRÜH, Christian, Siddharth JAIN et Avideh ZAKHOR, « Data Processing Algorithms for Generating Textured 3D Building Facade Meshes from Laser Scans and Camera Images », dans : *Int. J. Comput. Vision* 61.2 (2005), p. 159–184.

FUA, P., « From Multiple Stereo Views to Multiple 3-D Surfaces », dans : *International Journal of Computer Vision* 24.1 (1997), p. 19–35.

GARGALLO, Pau et Peter STURM, « Bayesian 3D Modeling from Images using Multiple Depth Maps », dans : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, t. 2, San Diego, Californie 2005, p. 885–891.

GOESELE, Michael, Brian CURLESS et Steven M. SEITZ, « Multi-View Stereo Revisited », dans : *CVPR '06 : Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA : IEEE Computer Society, 2006, p. 2402–2409.

GOOGLE, *Street View*, URL : http://maps.google.com.

HAALA, Norbert et Claus BRENNER, « Generation of 3D city models from airborne laser scanning data », dans : *3rd EARSL Workshop on LIDAR Remote Sensing on Land and Sea*, Tallinn 1997, 105112.

HERNÁNDEZ, Carlos et Francis SCHMITT, « Silhouette and stereo fusion for 3D object modeling », dans : *Comput. Vis. Image Underst.* 96.3 (2004), p. 367–392.

HILTON, Adrian, Andrew STODDART, John ILLINGWORTH et Terry WINDEATT, « Reliable Surface Reconstruction from Multiple Range Images », dans : *ECCV '96 : Proceedings of the 4th European Conference on Computer Vision-Volume I*, London, UK : Springer-Verlag, 1996, p. 117–126.

HOFF, William et Narendra AHUJA, « Surfaces from Stereo : Integrating Feature Matching, Disparity Estimation, and Contour Detection », dans : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11.2 (fév. 1989), 121136.

HOPPE, Hugues, Tony DEROSE, Tom DUCHAMP, John MCDONALD et Werner STUETZLE, « Surface reconstruction from unorganized points », dans : *SIGGRAPH '92 : Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, New York, NY, USA : ACM, 1992, p. 71–78.

HUERTAS, Andres, Ramakant NEVATIA et David LANDGREBE, « Use of hyperspectral data with intensity images for automatic building modeling », dans : *Second International Conference on Information Fusion*, t. 2, 1999, p. 680–687.

ISHIKAWA, Hiroshi et Davi GEIGER, « Occlusions, Discontinuities, and Epipolar Lines in Stereo », dans : *Proceedings of European Conference on Computer Vision, ECCV '98*, t. 1, Fribourg, Allemagne 1998, 232–248.

KANEKO, Shun'ichi, Ichiro MURASE et Satoru IGARASHI, « Robust Image registration by Increment Sign Correlation », dans : *The Journal of the Pattern Recognition Society, PR* 35.10 (oct. 2002), p. 2223–2234.

KANEKO, Shun'ichi, Yutaka SATOH et Satoru IGARASHI, « Using selective correlation coefficient for robust image registration », dans : *The Journal of the Pattern Recognition Society, PR* 36.5 (mai 2003), p. 1165–1173.

KERIVEN, Renaud et Olivier FAUGERAS, « Complete dense stereovision using level set methods », dans : *5th European Conference on Computer Vision*, 1998.

KOLMOGOROV, Vladimir et Ramin ZABIH, « Multi-camera Scene Reconstruction via Graph Cuts », dans : *ECCV '02 : Proceedings of the 7th European Conference on Computer Vision-Part III*, London, UK 2002, p. 82–96.

KUTULAKOS, Kirakos et Steven SEITZ, « A Theory of Shape by Space Carving », dans : *Int. J. Comput. Vision* 38.3 (2000), p. 199–218.

KUTULAKOS, Kiriakos et Steven SEITZ, « A Theory of Shape by Space Carving », dans : *International Conference on Computer Vision*, Corfu 1999.

LORENSEN, William et Harvey CLINE, « Marching cubes : A high resolution 3D surface construction algorithm », dans : *SIGGRAPH '87 : Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, New York, NY, USA : ACM, 1987, p. 163–169.

MARR, David et Tomaso POGGIO, *A Theory of Human Stereo Vision*, Rapport technique, Cambridge, MA : Massachusetts Institute of Technology, 1977.

MARTIN HUBER Wolfgang Schickler, Stefan Hinz et Albert BAUMGARTNER, « Fusion of Lidar Data and Aerial Imagery for Automatic Reconstruction of Building Surfaces », dans : *URBAN 2003*, Berlin, Allemagne 2003.

MOURAGNON, Etienne, Maxime LHUILLIER, Michel DHOME, Fabien DEKEYSER et Patrick SAYD, « Real-Time Localization and 3D Reconstruction », dans : *IEEE Conference of Vision and Pattern Recognition*, New-York, USA 2006.

NISTÉR, David, « Automatic Passive Recovery of 3D from Images and Video », dans : *3DPVT '04 : Proceedings of the 3D Data Processing, Visualization, and Transmission, 2nd International Symposium*, Washington, DC, USA : IEEE Computer Society, 2004, p. 438–445.

PAPARODITIS, Nicolas, Ouided BENTRAH, Lionel PÉNARD, Olivier TOURNAIRE, Bahman SOHEILIAN et Mathieu DEVEAU, « Automatic 3D Recording and Modeling of Large Scale Cities : the Archipolis Project », dans : *First International Workshop on Recording, Modelling and Visualization of Cultural Heritage*, 2005, p. 227–235.

PAPARODITIS, Nicolas, Jean-Philippe SOUCHON, Gilles MARTINOTY et Marc PIERROT-DESEILLIGNY, « High-end aerial digital cameras and their impact on the automation and quality of the produc-

tion workflow », dans : *ISPRS Journal of Photogrammetry and Remote Sensing* 60 (sept. 2006), p. 400–412.

PAPARODITIS, Nicolas, Christian THOM et Hassan JIBRINI, « Surface reconstruction in urban areas from multiple views of aerial digital frames », dans : *Proceedings of the XIXth ISPRS Congress*, t. 33, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS, Amsterdam 2000.

PIERROT-DESEILLIGNY, Marc, *Autocalibration des caméras numériques par ajustements de faisceaux sur des scènes planes texturées*, Rapport technique, IGN, 2005.

PITO, Richard, « Mesh integration based on co-measurements », dans : *In Proc. IEEE Int. Conf. on Image Processing, Special Session on Range Image Analysis*, 1996.

POLLEFEYS, Marc, David NISTÉR et al., « Detailed Real-Time Urban 3D Reconstruction from Video », dans : *International Journal of Computer Vision* 78.2-3 (juil. 2008), p. 143–167.

PONS, Jean-Philippe, Renaud KERIVEN et Olivier FAUGERAS, « Modelling Dynamic Scenes by Registering Multi-View Image Sequences », dans : *CVPR '05 : Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, Washington, DC, USA : IEEE Computer Society, 2005, p. 822–827.

ROY, Sébastien et Ingemar COX, « A Maximum-Flow Formulation of the N-camera Stereo Correspondence Problem », dans : *International Conference on Computer Vision*, Bombai 1998, p. 142.

RUTISHAUSER, Martin, Markus STRICKER et Marjan TROBINA, « Merging range images of arbitrarily shaped objects », dans : *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994, p. 573 –580.

SCHARSTEIN, Daniel et Richard SZELISKI, « A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms », dans : *Int. J. Comput. Vision* 47.1-3 (2002), p. 7–42, ISSN : 0920-5691.

— « Stereo Matching with Non-Linear Diffusion », dans : *International Journal of Computer Vision* 28.2 (juin 1998), 155–174.

SCHINDLER, Konrad et Joachim BAUER, « A Model-Based Method For Building Reconstruction », dans : *HLK '03 : Proceedings of the First IEEE International Workshop on Higher-Level Know-*

ledge in 3D Modeling and Motion Analysis, Washington, DC, USA : IEEE Computer Society, 2003, p. 74.

SEITZ, Steven et Charles DYER, « Photorealistic Scene Reconstruction by Voxel Coloring », dans : *Int. J. Comput. Vision* 35.2 (1999), p. 151–173.

SEITZ, Steven, Brian CURLESS, James DIEBEL, Daniel SCHARSTEIN et Richard SZELISKI, « A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms », dans : *CVPR '06 : Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA : IEEE Computer Society, 2006, p. 519–528.

SOHEILIAN, Bahman, Nicolas PAPARODITIS, Didier BOLDO et Jean-Paul RUDANT, « Automatic 3D Extraction Of Rectangular Roadmarks With Centimeter Accuracy From Stereo-pairs Of A Ground-based Mobile Mapping System », dans : *International Symposium on Mobile Mapping Technology*, Padua, Italia 2007.

SOUCY, Marc et Denis LAURENDEAU, « A General Surface Approach to the Integration of a Set of Range Views », dans : *IEEE Trans. Pattern Anal. Mach. Intell.* 17.4 (1995), p. 344–358.

STAMOS, Ioannis et Peter ALLEN, « Geometry and Texture Recovery of Scenes of Large Scale », dans : *Computer Vision and Image Understanding (CVIU)* 88.2 (nov. 2002), p. 94–118.

STRECHA, Christoph, Rik FRANSENS et Luc Van GOOL, « Combined depth and outlier estimation in multi-view stereo », dans : *Proceedings IEEE computer society conference on computer vision and pattern recognition - CVPR2006*, t. 2, New York, USA 2006, p. 2394–2401.

SZELISKI, Richard et Daniel SCHARSTEIN, « Symmetric Sub-Pixel Stereo Matching », dans : t. 2, Copenhague, Danemark 2002, 525–540.

TELLER, Seth, Matthew ANTONE, Zachary BODNAR, Michael BOSSE, Satyan COORG, Manish JETHWA et Neel MASTER, « Calibrated, Registered Images of an Extended Urban Area », dans : *Int. J. Comput. Vision* 53.1 (2003), p. 93–107.

TORR, Philip et Antonio CRIMINISI, « Dense stereo using pivoted dynamic programming », dans : *International Journal of Image and Vision Computing* 22.10 (sept. 2004), 795–806.

TOURNAIRE, Olivier, Nicolas PAPARODITIS, Franck JUNG et Bernard CERVELLE, « 3D Road-marks reconstruction from multiple calibrated aerial images », dans : *Proc. of the ISPRS Commission III Symposium on Photogrammetric and Computer Vision*, ISPRS, Bonn, Germany 2006.

TOURNAIRE, Olivier, Bahman SOHEILIAN et Nicolas PAPARODITIS, « Towards a sub-decimetric geo-referencing of ground-based Mobile Mapping Systems in urban areas : matching ground-based and aerial-based imagery using roadmarks », dans : *Proc. of the ISPRS Commission I Symposium*, t. Part A, Marne-la-Vall ee, France 2006.

TURK, Greg et Marc LEVOY, « Zippered polygon meshes from range images », dans : *SIGGRAPH '94 : Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, New York, NY, USA : ACM, 1994, p. 311–318.

VOGIATZIS, George, Carlos Hernández ESTEBAN, Philip H. S. TORR et Roberto CIPOLLA, « Multiview Stereo via Volumetric Graph-Cuts and Occlusion Robust Photo-Consistency », dans : *IEEE Trans. Pattern Anal. Mach. Intell.* 29.12 (2007), p. 2241–2246.

WERNER, Tomás et Andrew ZISSERMAN, « New Techniques for Automated Architectural Reconstruction from Photographs », dans : *ECCV '02 : Proceedings of the 7th European Conference on Computer Vision-Part II*, London, UK : Springer-Verlag, 2002, p. 541–555.

ZABIH, Ramin et John WOODFILL, « Non-parametric Local Transforms for Computing Visual Correspondence », dans : *Proceedings of European Conference on Computer Vision, ECCV*, Stockholm, Suède 1994, 151158.

ZHANG, Ye et Chandra KAMBHAMETTU, « Stereo Matching with Segmentation-based Cooperation », dans : *Proceedings of European Conference on Computer Vision, ECCV '02*, t. 2, Copenhague, Danemark 2002, 556–571.

ZHANG, Zonghua, Xiang PENG, Weiqiang SHI et Xiaotang HU, « A Survey of Surface Reconstruction from Multiple Range Images », dans : *2nd Asia-Pacific Conf. on Systems Integrity and Maintenance (ACSIM)*, t. 8, 2000, p. 519–523.

ZHAO, Huijing et Ryosuke SHIBASAKI, « Reconstructing Textured CAD Model of Urban Environment Using Vehicle-Borne Laser Range Scanners and Line Cameras », dans : *ICVS '01 : Proceedings of the Second International Workshop on Computer Vision Systems*, Springer-Verlag, 2001, p. 284–297.

TABLE DES FIGURES

1 Image aérienne et modèle tridimensionnel d'une portion du centre ville de Lyon. 6

2 Vue de rues de San Francisco dans Street View (*source : Google Maps*) 8

1.1 Vue d'un modèle 3D de Paris (*source PagesJaunes/IGN/ArchiVidéo*) 9

1.2 Véhicule et reconstruction de l'équipe de Berkeley (*source EECS - UC Berkeley*) 13

1.3 Résultats de la reconstruction automatique (*source : Center for Spatial Information Science, Tokyo University*) 15

1.4 Reconstruction d'une portion de rue de Paris (*source : CAOR - ENSMP*) 16

1.5 Reconstruction de façade par le projet Urbanscape (*source : Urbanscape project UNC / UK*) 17

1.6 Véhicule du projet Stéréopolis 18

1.7 Schéma montrant la géométrie d'acquisition de la base verticale de Stéréopolis 20

1.8 Deux prises de vues consécutives de la base stéréoscopique verticale de Stéréopolis. . . . 21

1.9 Vues géoréférencées acquises par la base verticale du véhicule Stéréopolis replacées dans le modèle 3D 22

1.10 Résultats des algorithmes développés dans DICK et al. 2004 23

1.11 Résultats des algorithmes développés dans WERNER et ZISSERMAN 2002 24

2.1 Images prises par le véhicule en multi-recouvrement sur une portion de façade. Photothèque de l'IGN, Saint-Mandé. 28

2.2 Géométrie épipolaire 31

2.3 Résultats obtenus avec deux images 32

2.4 Couple stéréoscopique d'images 33

2.5 Carte de profondeur obtenue par l'algorithme de Roy et Cox. D'après ROY et COX 1998 36

2.6 Résultats de régularisation de surface pour différentes valeurs du coefficient k. 38

2.6 Résultats de régularisation de surface pour différentes valeurs du coefficient k. 39

2.7 Dilatation à partir d'une surface pour délimiter l'espace de recherche à la résolution supérieure 40

2.8 Surface à plusieurs résolutions successives en vue ombrée. La résolution augmente de haut en bas. L'image maîtresse se trouve figure 2.1, en bas à droite. 41

2.8 Surface à plusieurs résolutions successives. La résolution augmente de haut en bas. Colonne de gauche : profondeur en niveaux de gris. Colonne de droite : même surface en vue ombrée. Dernière ligne : résultat après filtrage. L'image de départ se trouve figure 2.1, en bas à droite. 42

2.9 Résultat de l'algorithme de space carving sur une statuette en bois obtenu avec 21 images en entrée 45

2.10 Résultat de l'algorithme de voxel coloring sur un dinosaure à partir de 21 images en entrée. Les résultats sont donnés pour différentes résolutions de grille, de 13920 à 7696922 voxels 47

2.11 Résultat de l'algorithme de PONS et al. 2005 sur un buste sculpté, obtenu à partir de 24 images sources 47

2.12 Repère choisi pour décrire l'espace objet 49

2.13 Reprojection d'un point 3D dans les images. Vue de dessus. 50

2.14 MNF partie droite de la façade. Pas planimétrique : 2 cm. Taille des fenêtres de corrélation : 5 pixels. Echelle des profondeurs : plus un pixel est sombre, plus le point est éloigné 52

2.15 Un appariement faux peut avoir un score de corrélation plus élevé lorsque la position réelle correspond à une zone faiblement texturée. Points noirs : position réelle. Points rouges : position obtenant un score de corrélation élevé. 52

2.16 Ligne de parcours de l'espace de l'image maîtresse à (c,l) fixé. Reprojection d'un point terrain dans les autres images. Vue de dessus. 53

2.17 Ligne de parcours de l'espace de l'image maîtresse à (c,l) fixé. Reprojection d'un point terrain dans les autres images. Vue de dessus. 55
2.18 Carte de profondeur en multi-vues 57

3.1 Post-traitement en géométrie image : extraction d'un ensemble de points fiables. L'image maîtresse de départ se trouve figure 2.1 en bas à droite 61
3.2 Vues du nuage de points dérivé de la carte de profondeur obtenue sur la façade de la chapelle du château de Vincennes 62
3.3 Nuage de points 3D. Les points extraits à partir d'images maîtresses différentes ont des couleurs différentes. 64
3.4 Diagramme de Venn pour trois nappes (d'après SOUCY et LAURENDEAU 1995). 68
3.5 2 surfaces en recouvrement. La surface S_1 est en vert, la surface S_2 en bleu. 69
3.6 Parcours des points de S_2. Exemples de points qui ne sont pas en recouvrement avec S_1. . 71
3.7 Parcours des points de S_2. Exemple d'un point en recouvrement avec S_1. 72
3.8 Fin du parcours des points de S_2. 73
3.9 Retriangulation des trous de S_1. 73
3.10 Création du raccordement entre les deux nappes. 74
3.11 Résultat de l'algorithme de fusion (schéma) 75
3.12 Images maîtresses utilisées pour produire les résultats présentés 76
3.13 Surface initiale S_1. 77
3.14 Points du nuage N_2. 78
3.15 Points du nuage N_2 qui ne sont pas en recouvrement avec la surface initiale S_1. 78
3.16 Surface S_2 triangulée après suppression de triangles en recouvrement avec S_1. Ici, cette surface présente deux composantes connexes. 79
3.17 Triangles du raccord créés à l'interface entre les surfaces S_1 et S_2. 79
3.18 Résultat de l'algorithme de fusion de S_1 avec S_2. 80
3.19 Résultat de l'algorithme de fusion en intégrant plusieurs surfaces. 80

4.1 Surface reconstruite erronée 82

4.2 Zone d'occultation calculée pour l'image de gauche (en vert) 85
4.3 L'image gauche est eclue de l'ensemble des images visibles : elle ne pertubera pas les calculs de reconstruction 87
4.4 Parties cachées 88
4.5 Exemple d'amélioration de la description de la surface par prise en compte des occultations. L'image maîtresse est représentée figure 4.6(b) 90
4.6 Images maîtresses 91
4.7 Carte de profondeur à la résolution inférieure 91
4.8 Résultats de reconstruction avec une résolution croissante (8 *cm*, 4 *cm*, 2*cm*, 1 *cm* et 5 *mm* au niveau de la façade). Modèles présentés en vue ombrée. 92
4.8 Résultats de reconstruction avec une résolution croissante (8 *cm*, 4 *cm*, 2*cm*, 1 *cm* et 5 *mm* au niveau de la façade). Modèles présentés en vue ombrée. 93
4.8 Résultats de reconstruction avec une résolution croissante (8 *cm*, 4 *cm*, 2*cm*, 1 *cm* et 5 *mm* au niveau de la façade). Modèles présentés en vue ombrée. 94
4.9 Résultats de reconstruction à résolution maximale : zoom sur quelques détails. 95
4.9 Résultats de reconstruction à résolution maximale : zoom sur quelques détails. 96

LISTE DES TABLEAUX

2.1 Performances de l'algorithme avec ou sans l'utilisation de la multi-résolution pour un point de vue (figure 2.1, en bas à droite). Machine utilisée : P4 3GHz 43

2.2 Erreur de reconstruction en profondeur sur neuf points de contrôle. 56

LISTE DES ALGORITHMES

2.1 Stratégie d'appariement dense en multi-résolution à deux images 43
2.2 Stratégie d'appariement dense en multi-résolution 56
3.1 Stratégie de fusions de maillages triangulés issus de différents points de vue 75
4.1 Détermination des images dans lesquelles un voxel v est occulté 87
4.2 Stratégie d'appariement dense et d'estimation des parties cachées en multi-résolution . . . 89

TABLE DES MATIÈRES

Introduction **5**

1 Contexte et état de l'art **9**
1.1 Contexte général de la reconstruction de bâtiments . . . 9
1.2 Contexte particulier de la cartographie mobile . . . 11
1.2.1 Etat de l'art : reconstruction par systèmes de cartographie mobiles . . . 12
1.2.2 Le système Stéréopolis . . . 17
1.3 Etat de l'art en reconstruction de bâtiments à partir de données terrestres . . . 23
1.4 Bilan . . . 25

2 Appariement dense d'images pour la reconstruction 3D **27**
2.1 Appariement dense de deux images . . . 27
2.1.1 Etat de l'art . . . 28
2.1.2 Mesures de similarité . . . 28
2.1.3 Mise en place par corrélation . . . 30
2.1.4 Mise en oeuvre et résultats . . . 31
2.2 Régularisation de la surface . . . 33
2.2.1 Etat de l'art . . . 35
2.2.2 Algorithme de Roy et Cox . . . 35
2.2.3 Mise en oeuvre et résultats . . . 36
2.3 Multi-résolution . . . 37
2.3.1 Utilisation de la multi-résolution . . . 37
2.3.2 Apports de la multi-résolution . . . 40
2.4 Appariement dense d'images multiples . . . 44
2.4.1 Etat de l'art : reconstruction à partir d'images multiples . . . 44
2.4.2 Méthode de reconstruction proposée . . . 48
2.4.3 Géométrie objet . . . 48
2.4.4 Géométrie image . . . 52
2.4.5 Comparaison de géométries . . . 55
2.4.6 Algorithme proposé . . . 55
2.5 Bilan . . . 58

3 Reconstruction de surface **59**
3.1 Introduction . . . 59
3.2 Extraction des points fiables de la carte de profondeur . . . 60
3.3 Triangulation par point de vue . . . 61
3.3.1 Nuage de points 3D . . . 61
3.3.2 Triangulation de points 3D . . . 63
3.4 Fusion de maillages triangulés . . . 65

3.4.1 Etat de l'art . . . 65
3.5 Méthode proposée et résultats obtenus . . . 68
3.5.1 Filtrage des triangles incohérents . . . 69
3.5.2 Retriangulation des trous . . . 70
3.5.3 Création du raccord entre surfaces . . . 72
3.5.4 Résultats . . . 76
3.6 Bilan . . . 77

4 Vers une gestion des parties cachées pour l'appariement **81**
4.1 Introduction . . . 81
4.2 Etat de l'art . . . 82
4.3 Algorithme . . . 84
4.3.1 Calcul des zones d'occultation . . . 85
4.3.2 Modification des scores de corrélation . . . 86
4.4 Résultats . . . 88

Conclusion **97**
Conclusion de la thèse . . . 97
Reconstruction par appariement dense d'images . . . 97
Fusion de surfaces . . . 97
Gestion des parties cachées . . . 98
Perspectives . . . 98

Bibliographie **101**

Table des figures **112**

Liste des tableaux **113**

Liste des algorithmes **115**