

Reconstruction 3D automatique de façades de bâtiments en multi-vues

Automatic 3D Building Facade Reconstruction from Multiple Views

L. Pénard

N. Paparoditis

M. Pierrot-Deseilligny

Institut Géographique National
2-4 avenue Pasteur 94165 Saint-Mandé
(lionel.penard, nicolas.paparoditis, marc.pierrot-deseilligny)@ign.fr

Résumé

Dans cet article, nous présentons une méthode permettant d'extraire automatiquement, à partir d'un ensemble de plusieurs vues géoréférencées et en recouvrement acquises par un véhicule de cartographie mobile, une surface triangulée et texturée de façades de bâtiments. Nous nous plaçons dans un contexte de production de masse de données, et nous visons une très forte précision géométrique. L'idée centrale de cet article est d'extraire pour chaque vue un nuage de points 3D. Ceci est réalisé en géométrie image : pour chaque pixel, nous calculons la position du point 3D correspondant par une technique de corrélation multi-images en multi-résolution. Une étape d'optimisation est ajoutée, afin de régulariser la surface résultat. Un post-traitement permet de ne garder de cette surface qu'un ensemble de points 3D en lesquels nous avons une forte confiance, puis de les trianguler. Le résultat est un maillage 3D de la portion de façade vue par chaque image. Sa forme et sa position sont très précises. Le plaquage de texture est correct.

Mots Clef

Reconstruction 3D de surfaces, optimisation globale, corrélation multi-images, multi-résolution.

Abstract

In this paper, we present an algorithm, which automatically generates textured meshes of building facades from a set of multi overlapping calibrated and oriented images acquired by a mobile mapping system. We are in a context of massive data production and aim at high geometric accuracy. The central idea is to create a 3D point cloud. We compute the depths of the pixels in each image using a correlation-based method in the image-space geometry. This is achieved with a multi-resolution scheme. We then perform a global optimization to find a regularized surface. A post-processing step is necessary to filter points according to a confidence index and remove the sparse speckle noise. Finally, the 3D point cloud is triangulated. The resulting mesh is an accurate representation of the facade surface from each image view point. We obtain promising

results, with a correct texture projection on the reconstructed model.

Keywords

3D Surface Reconstruction, Global Optimization, Multi-Image Matching, Multi-Resolution.

1 Introduction

L'automatisation de la reconstruction de modèles 3D urbains suscite beaucoup d'intérêt depuis plusieurs années. Des techniques à base d'images aériennes, de scanners laser aéroportés ou d'imagerie satellite sont couramment utilisées pour retrouver la géométrie en 3D des bâtiments (par exemple [1, 6]). Leur résolution ne permet cependant pas de fournir des détails géométriques précis sur les façades des bâtiments, ce qui nécessite des données acquises depuis le sol. Les techniques de reconstruction à partir de données terrestres dépendent du but recherché ainsi que du système d'acquisition employé. De nombreux auteurs traitent de la reconstruction d'un bâtiment remarquable ou d'un ensemble peu étendu de bâtiments [4, 17, 7]. Dans ce cas, le temps d'acquisition n'est pas une contrainte, et l'on peut utiliser des images de scanners laser, des images de nombreux points de vue... En revanche, dans notre cas où l'information géométrique doit être extraite à l'échelle d'une ville entière, la mise en oeuvre de l'acquisition de données (laser ou image) ne doit pas être lourde. La reconstruction de modèles urbains très étendus nécessite donc d'utiliser un véhicule mobile, à bord duquel sont installés les capteurs souhaités. Plusieurs auteurs ont exploré le champ des systèmes mobiles [8, 21]. Dans ces articles, les méthodes décrites pour retrouver la géométrie des bâtiments reposent sur l'enregistrement de données laser. Notre système d'acquisition n'est constitué que de caméras, la géométrie doit être extraite des images qui sont calibrées et géoréférencées [3].

L'extraction de géométrie à partir d'un ensemble de plusieurs images est un problème classique. Le cas stéréo simple (2 images) a fait l'objet d'un article de comparaison entre la plupart des méthodes existantes [15]. Pour le cas multi stéréo qui nous intéresse, de nombreuses approches

ont été proposées, avec de bons résultats : méthodes basées sur les level-sets [5], techniques de voxel coloring [16] ou de space carving [12], méthodes basées sur des particules de surface orientées [9] ou encore méthodes utilisant un formalisme bayésien [18, 10]. Cependant, nous souhaitons un algorithme simple et rapide compte tenu de la taille des données, et qui ne se limite pas à un nombre restreint d'images. Wang *et al.* proposent dans [20] de rechercher la surface de la façade et calculant la différence de profondeur entre la surface réelle et le plan principal de la façade sous la forme de rectangles périodiques. Cette technique, bien adaptée pour les fenêtres par exemples, est trop restrictive. Nous avons choisi de calculer directement, à partir des images, des cartes denses de profondeur par une technique de coupe de graphe puis d'en extraire des nuages de points 3D, et de les trianguler.

Cet article est organisé de la manière suivante : la partie 2 donne la stratégie générale adoptée, les parties 3 à 7 décrivent les étapes de l'algorithme annoncées en partie 2, la partie 8 présente quelques résultats et la partie 9 est la conclusion de ce travail.

2 Stratégie générale

2.1 Données

Le système d'acquisition mobile Stéréopolis [2] est constitué de 2 antennes GPS pour la localisation, et de 6 caméras numériques panchromatiques. Celles-ci sont disposées de la manière suivante : de chaque côté, 2 caméras forment une base stéréoscopique verticale permettant d'obtenir des images de façades ; à l'arrière, 2 caméras forment une base stéréoscopique horizontale (ces dernières ne sont pas utilisées dans le cadre de cet article). Ces caméras ont été construites et calibrées précisément par le laboratoire d'instrumentation de l'IGN (LOEMI), et nous supposons donc parfaitement connus l'ensemble des paramètres intrinsèques des caméras (focale, distorsion). Nous faisons également l'hypothèse pour ce travail que les paramètres de pose sont connus. Un travail est actuellement en cours pour automatiser cette étape (voir [3]).

A pleine résolution, nos images sont de taille 4096x4096, codées sur 4096 niveaux de gris, avec un rapport signal sur bruit de 300. La focale est de 28 mm, ce qui représente un angle de vue de 60 degrés environ. La taille du pixel sur les façades est en moyenne de l'ordre de quelques millimètres. Le recouvrement vertical de la base verticale et horizontal dans le sens de propagation du véhicule est suffisant pour assurer que chaque point des façades est vu dans au moins 2 images, et plus généralement dans 4 à 6 images.

Aucune donnée externe (telle qu'un plan cadastral, une orthophotographie ou un modèle numérique d'élévation) n'est nécessaire : géométrie et textures sont extraites des seules images.

2.2 Résumé des traitements

Le processus peut être résumé comme suit :

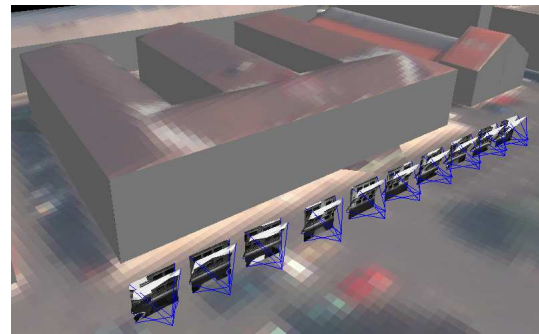


FIG. 1 – Données en entrée : images géoréférencées en vue 3D

- recherche d'une surface s'appuyant sur de forts scores de corrélation (parties 3 et 4) ;
- génération d'un nuage de points 3D fiables sur cette surface (partie 5) ;
- triangulation et texturation du nuage pour obtenir le modèle final (partie 7).

La première étape sera traitée en multi-résolution (partie 6) pour accélérer l'algorithme.

3 Espace objet vs. espace image

Il existe deux grands types d'algorithmes de recherche de surface 3D, basés sur l'appariement de points homologues dans des images, que ce soit dans le cas stéréo pure ou dans le cas multi-vues. Le premier recouvre toutes les techniques utilisant la géométrie image, c'est-à-dire les procédés qui cherchent à retrouver la profondeur d'objets repérés par les coordonnées image (colonne, ligne, profondeur). Cette stratégie revient à trouver la meilleure profondeur possible pour les pixels de l'image. Le repérage dans l'espace est lié à chaque image géoréférencée, et semble donc mal adapté à notre problème pour lequel plusieurs images sont nécessaires pour couvrir une façade. Il est plus naturel de choisir une technique basée sur la géométrie objet, où un unique repère lié au terrain sera utilisé pour calculer les coordonnées des points de la façade. Dans cette seconde approche, l'espace objet est discrétisé en voxels, et la recherche de la géométrie de la façade revient donc à trouver les meilleurs voxels support de cette surface.

3.1 Appariement en espace objet

Comme cela vient d'être dit, l'espace objet est logiquement un cadre plus naturel pour repérer une surface telle qu'une façade. Cette approche permet également de conserver une symétrie parfaite dans le traitement des images, aucune ne jouant le rôle d'image maîtresse. 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.

L'espace objet de recherche est le volume défini par : $X_{min} < X < X_{max}$, $Y_{min} < Y < Y_{max}$, $Z_{min} < Z < Z_{max}$. 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, 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 montre une vue schématique de dessus de ce processus.

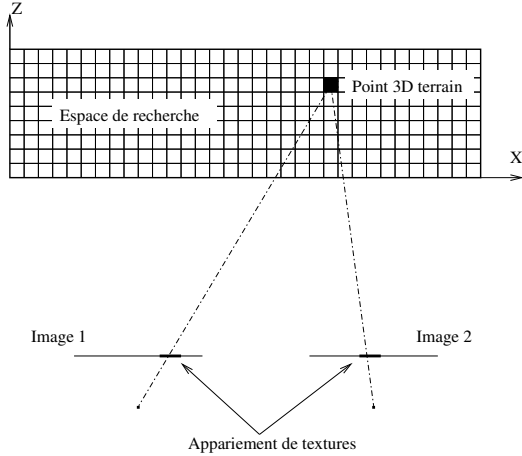


FIG. 2 – Reprojection d'un point 3D dans les images. Vue de dessus.

Lorsqu'il n'y a pas de parties du bâtiment cachées dans les images, 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 rééchantillonnage des images est réalisé *a priori* par le calcul d'une image rééchantillonnée 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 :

$$\mathbf{V}_i = \begin{pmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{in} \end{pmatrix} \quad (1)$$

et le coefficient de corrélation multi-images, issu des tra-

vaux [13] est donné par :

$$cor(X, Y, Z) = \frac{var(\sum_{i \in I_M} \mathbf{V}_i)}{\sum_{i \in I_M} var(\mathbf{V}_i)} \quad (2)$$

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 où :

$$I_{X,Y} = \bigcap_{Z_{min} < Z < Z_{max}} I_M \quad (3)$$

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

$$\hat{Z}_{X,Y} = \underset{Z_{min} < Z < Z_{max}}{\operatorname{argmax}} cor(X, Y, Z) \quad (4)$$

L'implémentation de cet algorithme ne donne pas des résultats aussi bons qu'escompté. La figure 3 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 5x5 pixels. 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 était un effet attendu de notre processus. 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 2 joints de pierre différents dans 2 images peut conduire, du fait de la similarité de ces joints, à un score élevé. La figure 4 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 que nous avons adoptée accentue ces erreurs. En effet, pour X et Y donnés, 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 4. En revanche, dans le cas de la recherche en géométrie image, le parcours des profondeurs à (colonne, ligne) fixées garantit que la vignette de corrélation de l'image maîtresse reste identique. Nous proposons alors de changer le score de corrélation. En géométrie image, il ne s'agit plus de comparer plusieurs vignettes entre elles, mais de comparer une vignette de l'image référence à une vignette de chacune des autres images. Ainsi



(a) Partie d'une façade

(b) Zoom sur une fenêtre

FIG. 3 – MNF partie droite de la façade. Pas planimétrique : 2cm. Taille des fenêtres de corrélation : 5 pixels. Echelle des profondeurs : plus un pixel est sombre, plus le point est éloigné

on impose que la reprojection du point 3D dans les images doit ressembler à celle dans l'image maîtresse, ce qui évite les problèmes de l'approche géométrie objet. L'approche géométrie image est donc adoptée dans la suite.

3.2 Appariement en espace image

Remarque préliminaire : l'utilisation d'une approche en géométrie image contraindra à fusionner les données extraites lorsque chaque image joue le rôle d'image maîtresse. Cette étape est en cours de réalisation et n'est pas décrite ici.

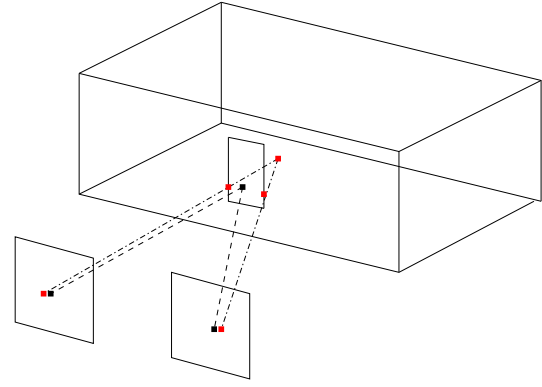
La stratégie suivie est la suivante. Une image maîtresse est choisie. Pour chacun de ses pixels (colonne, ligne) nous cherchons à retrouver la profondeur à laquelle se situe l'image réelle. Un point M de l'espace 3D sera repéré par ses coordonnées (c, l, p) liées à l'image maîtresse : colonne, ligne, profondeur.

Le problème se ramène au cas précédent, et à c et l fixés, nous recherchons le meilleur p entre deux bornes p_{min} et p_{max} . Le point 3D est reprojeté dans les autres images du sous-ensemble $I_{c,l} = \bigcap_{p_{min} < p < p_{max}} I_M$ (voir figure 5). Le coefficient de corrélation multi-images utilisé dans la partie 3.1 était symétrique, il était une mesure de similarité de toutes les images entre elles. Ici, la symétrie entre les images est brisée, un nouveau coefficient est donc choisi, mesurant 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 2 images :

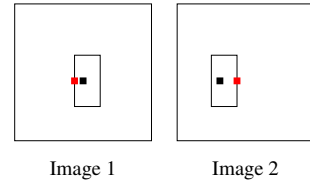
$$cor_i(c, l, p) = \frac{1}{N-1} \sum_{j \neq i} CCS(\mathbf{V}_i, \mathbf{V}_j) \quad (5)$$

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 $CCS(\mathbf{V}_i, \mathbf{V}_j)$ le coefficient de corrélation classique.

Les résultats sont meilleurs que dans le cas de la géométrie objet (voir figure 6). Pour la majorité des pixels, la profondeur a correctement été retrouvée, mais des erreurs généra-



(a) Vue 3D



(b) Vue image

FIG. 4 – 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 de score élevé.

lement ponctuelles subsistent. Par exemple, les limites des images créent des perturbations visibles (figure 6(b)). Ceci s'explique par le changement du facteur de normalisation $N - 1$ dans l'équation 5 lors du franchissement de ces limites.

4 Optimisation globale

Le choix a été fait de garder le meilleur score pour sélectionner la profondeur 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 attendons également un résultat robuste. Une régularisation de la surface est donc nécessaire. Pour ce faire, plusieurs approches de flot maximal ou de coupe de graphes ont été proposées [19, 11, 14]. Le processus d'optimisation globale par flot maximal développé par Roy et Cox [14] 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 est choisi volontairement très bas, de manière à préserver les discontinuités réelles des façades. Le résultat après application de cette technique sur nos données est présenté figure 7(a). La surface est lissée,

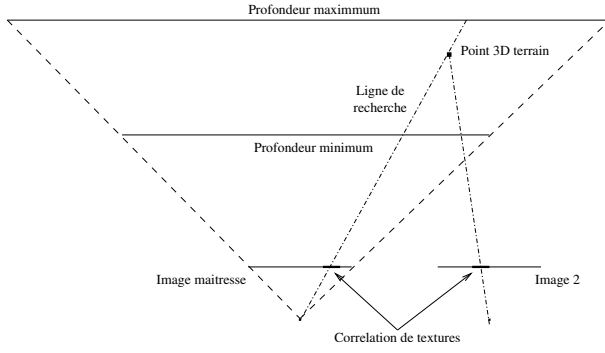


FIG. 5 – Ligne de parcours de l’espace de l’image maîtresse à (c, l) fixé. Reprojection d’un point terrain dans les autres images. Vue de dessus.

les discontinuités sont bien conservées.

5 Extraction d’un nuage de points fiables

Pour chaque image, la profondeur a été calculée pour tous les pixels. Cependant, cette valeur n’a pas de sens réel pour le 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 lesquels 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 ajoutée. Le but est d’enlever tous les pixels pour lesquels la fiabilité du résultat est faible. Cette fiabilité est traduite par 2 mesures. Tout d’abord la texture au voisinage du pixel doit être suffisante, car sinon l’appariement avec des zones similaires sera relativement indéterminé. La texture est mesurée par la variance des valeurs de niveau de gris du voisinage. 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. Les pixels que nous cherchons à supprimer vérifient au moins une des conditions suivantes :

$$\begin{cases} \max_{p_{min} < p < p_{max}} cor_i(c, l, p) < seuil_{cor} \\ var(\mathbf{V}_i) < seuil_{var} \end{cases} \quad (6)$$

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 fixés à 0 afin qu’ils n’influencent pas la surface régularisée. 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



(a) Image en entrée

(b) Profondeur du maximum de corrélation



(c) Profondeur du maximum de corrélation : zoom sur une fenêtre

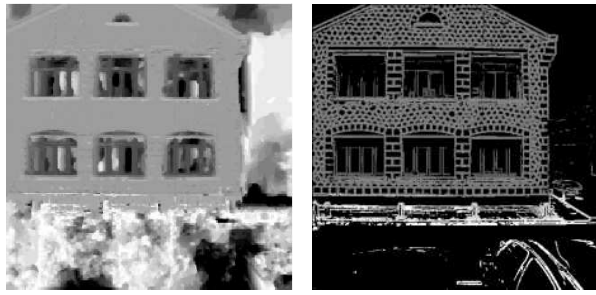
FIG. 6 – Carte de profondeur en espace image. Taille des fenêtres de corrélation : 3 pixels, pas en profondeur : 2 cm.

toutes les vues maîtresses. La figure 7(b) montre le résultat après filtrage en géométrie image.

6 Multi-résolution

6.1 Utilisation de la multi-résolution

L’algorithme précédent fournit une estimation robuste et précise de la position de points 3D sur la façade d’un bâtiment. Cependant le temps de calcul est long. A titre d’exemple, pour une image sous-échantillonnée 4 fois (donc de taille 1024x1024), pour une plage de recherche en profondeur de 5 m, et un pas de discrétisation de 2 cm, le temps de traitement tout compris est de 1h30 environ sur un pentium 4 à 3 GHz. Ce temps est prohibitif lorsque l’on sait que le système d’acquisition capture environ 2 images tous les 5 m de façade, et qu’il faut toutes les traiter. L’algorithme a donc 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 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’es-



(a) MNF après optimisation

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

FIG. 7 – Post-processing en géométrie image. L'image de départ se trouve figure 6(a)

pace de recherche à la résolution supérieure (voir figure 8).

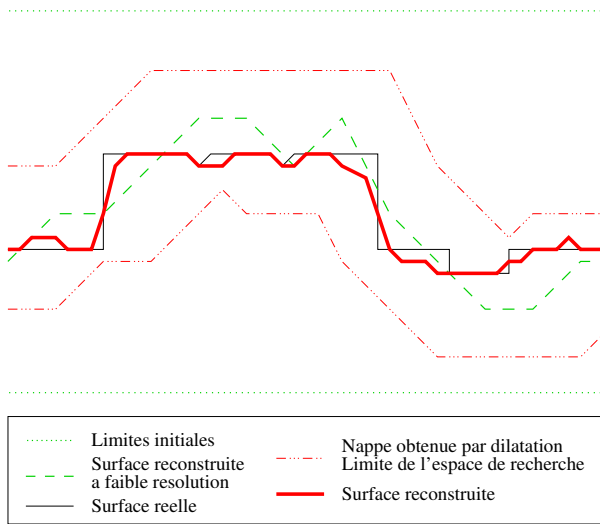


FIG. 8 – Dilatation à partir d'une surface pour délimiter l'espace de recherche à la résolution supérieure

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 2 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. Notre choix est de prendre les dilations minimales qui vérifient le premier impératif.

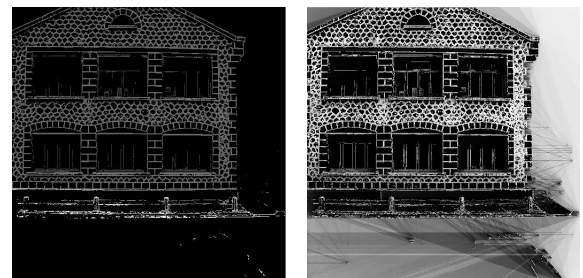
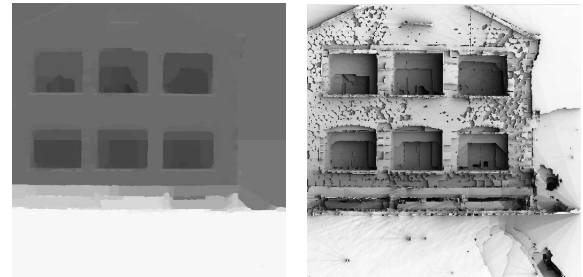
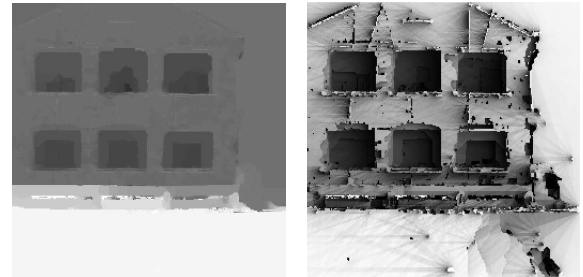
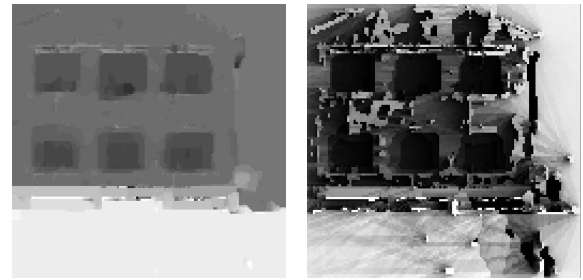


FIG. 9 – Surface à plusieurs résolutions successives. La résolution augmente de haut en bas. Colonne de gauche : profondeur en niveaux de gris. Colonne de droite : même surface en vue ombrée. Dernière ligne : résultat après filtrage. L'image de départ se trouve figure 6(a)

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

TAB. 1 – Performances de l'algorithme avec ou sans l'utilisation de la multi-résolution pour un point de vue (figure 6(a)). Machine utilisée : P4 3GHz

6.2 Apports de la multi-résolution

Dans un premier temps, les traitements décrits dans les parties précédentes ont été réalisés à pleine résolution. Ceci avait introduit des limitations dont nous souhaitons nous affranchir. Le problème clé était celui du 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 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 pourra 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 modèle 3D grossier du bâtiment ou 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.

7 Reconstruction de la surface

Les étapes précédentes de l'algorithme ont permis de générer, pour chaque vue maîtresse, un nuage de points 3D. La figure 10 montre un exemple de résultat sur une façade. Pour chaque vue maîtresse, une couleur différente est utilisée pour représenter le nuage de points. Le mélange est parfait dans le sens où dans les zones de recouvrement les nuages sont bien ajustés les uns aux autres.

L'étape de filtrage a permis de supprimer presque tous les points étrangers au bâtiment (sol, ciel) mais aussi des points peu fiables sur la façade (fenêtres, mais aussi pierres homogènes, zones peintes uniformément...). Le nuage de points n'est donc plus dense partout. Une triangulation est donc menée pour parvenir à une description surfacique. Le choix de la géométrie image nous procure l'avantage de pouvoir faire une triangulation de Delaunay dans cette géométrie.

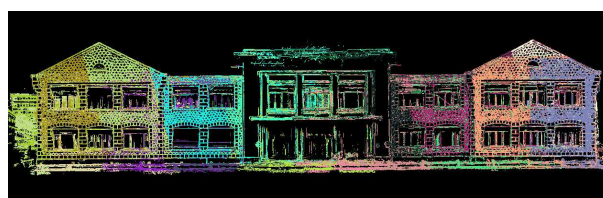
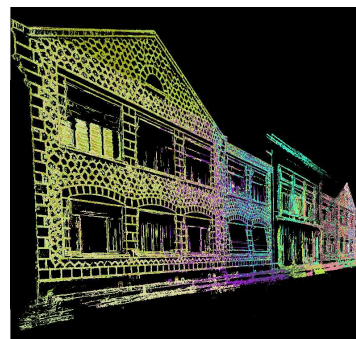
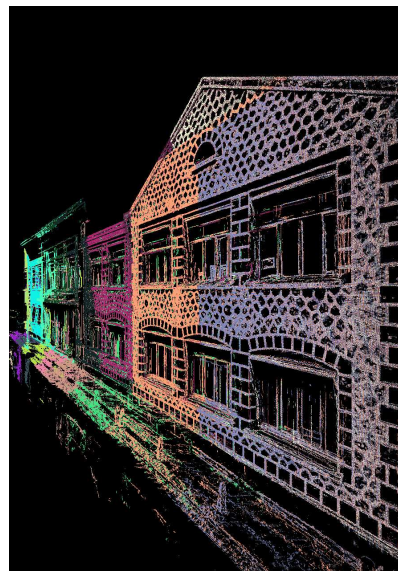
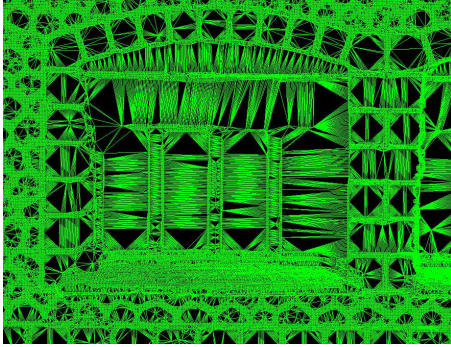


FIG. 10 – Nuage de points 3D. Les points extraits à partir d'images maîtresses différentes ont des couleurs différentes.



(a) Partie d'image



(b) Triangulation correspondante

FIG. 11 – Triangulation du nuage de points sur une partie de l'image.

8 Résultats et perspectives

Notre algorithme produit un modèle surfacique pour chacune des vues images, il y en a donc plusieurs par façade. La figure 12 montre l'un de ces modèles. Les zones planes sont bien reconstruites. La position des limites de fenêtres est très précise. Les vitres, peu texturées, sont reconstruites comme des surfaces planes entre les montants : ceci met en évidence l'importance de l'étape de filtrage. En revanche, des triangles erronés apparaissent entre les montants des fenêtres et les murs proches (voir figure 12(a)). Le nuage de points 3D n'est pas en cause, c'est l'étape de triangulation qui lie des points proches en géométrie image, alors que la topologie 3D n'est pas celle-ci. Au bas du bâtiment, la zone claire et peu texturée est responsable de la mauvaise qualité de la reconstruction : en effet, la densité de points fiables sur cette zone est très faible, la régularisation globale est trop faible sur cette zone. Pour évaluer quantitativement notre méthode, nous disposons de quelques points de contrôle, dont les coordonnées sont d'une précision meilleure que le millimètre, qui ont servi à mettre en place les images. Puisque c'est à partir de ces points

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

TAB. 2 – Erreur de reconstruction pour 9 points de contrôle dans la vue figure 6(a).

que le géoréférencement a été obtenu, on peut considérer que les erreurs constatées sur ces points mesurent uniquement l'erreur de reconstruction. Le tableau 2 montre les erreurs sur les 9 points de contrôle d'une vue (figure 6(a)). Pour ces points, l'erreur de reconstruction est très faible : maximum 1,8 cm pour un pas de discrétisation en profondeur de 1 cm. Bien sûr, la qualité de ces résultats ne rend pas compte de la précision globale de la surface, car les points de contrôle sont des points d'intérêts dans les images, c'est-à-dire des points pour lesquels notre technique est la plus fiable. Néanmoins, ils permettent de vérifier que nous n'avons pas introduit de biais important. En termes de temps de calcul, l'apport de la multi-résolution a déjà été présenté section 6.

Jusqu'à présent, aucune contrainte n'a été appliquée à la triangulation. Les segments de la scène, présents en très grand nombre sur les bâtiments, pourraient être utilisés afin de garantir une reconstruction plus précise sur les éléments linéaires. Le rendu serait davantage photo-réaliste.

L'autre développement important pour la suite est d'assurer la fusion des surfaces issues des différents points de vue. Différentes vues d'une même zone permettront d'améliorer la qualité de l'ensemble en palliant au moins partiellement les problèmes de parties cachées. De plus, nous pouvons espérer obtenir une scène fusionner avec une densité de points 3D supérieure à celle de chacune des vues, c'est à dire obtenir une reconstruction finale en super-résolution.

9 Conclusion

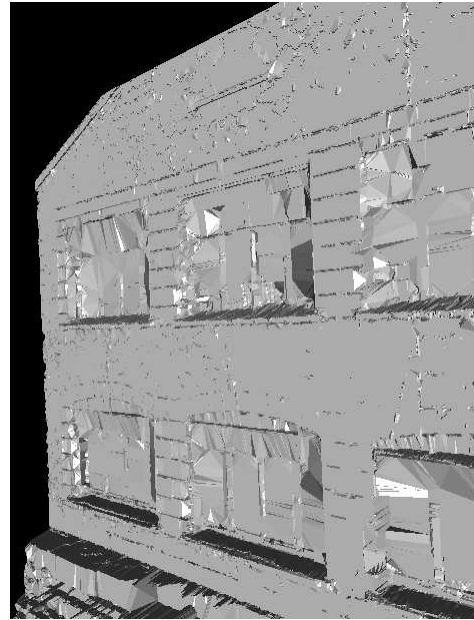
Nous avons présenté une méthode permettant, à partir d'un ensemble de prises de vues en recouvrement et dont tous les paramètres (intrinsèques et extrinsèques) sont connus, de reconstruire précisément la surface de façades de bâtiments. L'utilisation de la multi-résolution permet de nous affranchir d'une étape d'initialisation fournie soit par un plan cadastral, soit par une acquisition laser. Le processus est entièrement automatique.

Nous commençons par extraire un nuage de points 3D par corrélation en géométrie image. Une étape d'optimisation globale de surface et de filtrage permet, à partir des scores de multi-corrélation de ne retenir qu'un ensemble de points

3D robuste et précis. Le nuage ainsi obtenu est ensuite triangulé puis texturé pour obtenir la reconstruction finale. Nos résultats montrent une très bonne précision géométrique sur nos données. Cette méthode est robuste, précise et nous procure des nuages de points très denses par rapport à ceux obtenus par des laser mobiles. Notons pour finir que notre méthode, ici appliquée à des façades, ne fait aucune hypothèse particulière sur la nature de la surface à reconstruire, et peut donc s'appliquer à tout type de scènes.

Références

- [1] C. Baillard, H. Maitre. *3-D Reconstruction of Urban Scenes from Aerial Stereo Imagery : A Focusing Strategy*. CVIU, 76(3), p. 244-258, 1999.
- [2] O. Bentrach, N. Paparoditis, M. Pierrot-Deseilligny. *STEREOPOLIS : An Image Based Urban Environments Modeling System*. Proceedings of the 4th International Symposium on Mobile Mapping Technology, Kunming, China, 2004
- [3] O. Bentrach, N. Paparoditis, M. Pierrot-Deseilligny, R. Horaud. *Estimating Sensor Pose from images of a stereo rig*. ISPRS, 2004.
- [4] S. Coorg, S. Teller. *Extracting textured vertical facades from controlled close-range imagery*. CVPR 1999.
- [5] Olivier Faugeras, Renaud Keriven. *Complete Dense Stereovision Using Level Set Methods*. Proceedings of the 5th European Conference on Computer Vision-Volume I, p. 379-393, 1998
- [6] A. Fischer, T. Kolbe, F. Lang, A. Cremers, W. Förstner, L. Plümer, V. Steinhage. *Extracting Buildings from Aerial Images Using Hierarchical Aggregation in 2D and 3D*. CVIU, 72(2), p. 163-185, 1998.
- [7] A. W. Fitzgibbon, A. Zisserman. *Automatic 3D Model Acquisition and Generation of New Images from Video Sequences*. Proceedings of European Signal Processing Conference, p. 1261-1269, 1998.
- [8] C. Frueh, S. Jain, A. Zakhor. *Data Processing Algorithms for Generating Textured 3D Building Facade Meshes from Laser Scans and Camera Images*. International Journal of Computer Vision, 61(2) p. 159-184, 2005.
- [9] P. Fua. *From Multiple Stereo Views to Multiple 3-D Surfaces*. International Journal of Computer Vision, 24(1) p. 19-35, 1997.
- [10] P. Gargallo, P. Sturm. *Bayesian 3D Modeling from Images using Multiple Depth Maps*. Proceedings of Computer Vision and Pattern Recognition, p. 885-891, 2005.
- [11] V. Kolmogorov, R. Zabih, S. J. Gortler. *Generalized Multi-camera Scene Reconstruction Using Graph Cuts*. Proceedings of EMMCVPR, p. 501-516, 2003.
- [12] K. N. Kutulakos, S. M. Seitz. *A Theory of Shape by Space Carving*. International Journal of Computer Vision, 38(3), p. 199-218, 2000.



(a) Vue de côté sans texture



(b) Vue de côté texturée

FIG. 12 – Surface reconstruite à partir d'un point de vue image

- [13] N. Paparoditis, C. Thom, H. Jibrini. *Surface reconstruction in urban areas from multiple views of aerial digital frames*. Proceedings of the XIXth ISPRS Congress, 2000.
- [14] S. Roy, I. J. Cox. *A Maximum-Flow Formulation of the N-camera Stereo Correspondence Problem*. International Conference on Computer Vision, p. 492-502, 1998.
- [15] D. Scharstein, R. Szeliski. *A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondance Algorithms*. International Journal of Computer Vision, 47(1-3), p. 7-42, 2002.
- [16] S. M. Seitz, C. R. Dyer. *Photorealistic Scene Reconstruction by Voxel Coloring*. International Journal of Computer Vision, 35(2), p. 151-173, 1999.
- [17] I. Stamos, P. Allen. *3d model construction using range and image data*. Proceedings of Computer Vision and Pattern Recognition, p. 531-536, 2000.
- [18] C. Strecha, R. Fransens, L. Van Gool. *Wide-Baseline Stereo from Multiple Views : A Probabilistic Account*. Proceedings of Computer Vision and Pattern Recognition, p. 552-559, 2004.
- [19] G. Vogiatzis, P.H.S. Torr and R. Cipolla. *Multi-view stereo via Volumetric Graph-cuts*. Proceedings of Computer Vision and Pattern Recognition, p. 391-398, 2005.
- [20] X. Wang, S. Totaro, F. Taillandier, A. Hanson, S. Teller. *Recovering Facade Texture and Microstructure from Real-World Images*. Proceedings of ISPRS Commission III Symposium on Photogrammetric Computer Vision, 2002.
- [21] H. Zhao, R. Shibasaki. *Reconstructing Textured CAD Model of Urban Environment Using Vehicle-Borne Laser Range Scanners and Line Cameras*. Proceedings of the Second International Workshop on Computer Vision Systems, p. 284-297, 2001.