

Correction et amélioration des résultats de la stéréoscopie par corrélation : application au MNE urbain*

Frédéric Devernay
Julien Devade

ISTAR, Les Espaces de Sophia F, 80 route des Lucioles
B.P. 282, 06905 Sophia Antipolis cedex
devernay@istar.fr

Résumé

La stéréoscopie par corrélation classique permet de restituer un Modèle Numérique d'Élévation urbain (MNE) sans avoir à faire d'hypothèse a priori sur les modèles de bâtiments observés, contrairement aux méthodes à base de primitives. Cependant, cette technique souffre de plusieurs défauts, notamment au niveau des discontinuités de profondeur (par ex. au bord des toits). La solution classique consiste à utiliser une méthode de stéréoscopie à base de fenêtres adaptatives [KO94], et nous proposons dans cet article une solution alternative qui consiste à faire des corrections automatiques sur le MNE *a posteriori*. Cette méthode est basée sur une combinaison de techniques de morphologie mathématique et de diffusion anisotrope, et utilise l'information de contraste local de l'orthoimage pour déterminer quels endroits ont été «accrochés» par la corrélation.

Mots-clés : stéréoscopie par corrélation, MNE urbain, diffusion anisotrope.

1 Introduction

Pour produire des modèles numériques d'élévation urbain par stéréoscopie à partir de photographies aériennes ou d'imagerie spatiale haute résolution, il existe essentiellement deux types de méthodes. Les méthodes à base de primitives, consistent à détecter puis à apparier entre deux images ou plus des segments, des courbes, ou des objets plus complexes (par ex. des modèles de polyèdres). Les méthodes à base de stéréoscopie par corrélation consistent à mettre en correspondance des surfaces ayant des textures proches.

Le défaut principal des méthodes par primitives est de reposer sur une détection plus ou moins fidèle de ces primitives dans les images, et d'être par construction limitées à la reconstruction d'objets qui sont constitués par ces primitives : un bâtiment circulaire ne peut pas être reconstruit par une méthode à base de segments. La stéréo par corrélation, au contraire, permet de

*publié dans *3D Geospatial Data Production : Meeting Application requirements*, ISPRS Workshop, Paris, France, avril 1999.

restituer des objets de forme quelconque, dont la surface possède des propriétés de régularité géométrique et radiométrique suffisantes. Les défauts majeurs de la stéréo par corrélation sont de mal restituer (a) les surfaces présentant des réflexions spéculaires, (b) les surfaces trop faiblement contrastées, (c) les surfaces de pente trop forte, et (d) les discontinuités de profondeur. Pour la restitution de milieux urbains à partir de photographies aériennes, ce sont surtout les défauts (b) et (d) qui dominent : les objets «bavent» hors de leur limite, notamment à cause des fortes différences de contraste entre l'arrière-plan (le sol, qui est souvent à l'ombre) et le premier plan (toits, végétation, etc.) [Bus97, FW98].

Pour certaines applications, comme les réseaux haut débit qui utilisent des liaisons point-à-point, l'information utile est l'élévation des bâtiments et des superstructures, et le fait d'avoir dans le MNE des objets plus larges que la réalité n'est pas gênant. Par contre, d'autres applications, notamment la simulation de propagation radio pour la téléphonie mobile, utilisent les contours des bâtiments, et il est donc nécessaire d'avoir le plus précisément possible leur position.

La solution la plus courante aux défauts (b) et (d) passe par des méthodes de corrélation utilisant des fenêtres dites «adaptatives» [KO94, Lot96], mais cette solution présente l'inconvénient d'être beaucoup plus complexe et moins facilement optimisable que la corrélation classique. Nous avons donc préféré une autre approche, qui consiste à utiliser la corrélation classique pour construire un MNE, éventuellement par fusion des MNE issus de plusieurs couples [GGR94], puis à tenter d'améliorer a posteriori le résultat de la corrélation.

La méthode que nous avons développée utilise le fait que ce sont les zones contrastées qui «accrochent» la fenêtre de corrélation. Elle a été d'abord basée sur une méthode de diffusion anisotrope [PM90] qui utilise à la fois le MNE calculé par corrélation et une orthoimage calculée à partir de ce MNE. Nous avons ensuite ajouté des processus permettant soit d'accélérer la diffusion, soit de traiter le MNE avant la diffusion pour commencer celle-ci dans un état plus proche du résultat voulu. La méthode peut fonctionner soit sur un MNE et l'orthoimage associée, soit sur une image de disparité et l'image de référence correspondante.

Les résultats montrent, sur un cas difficile (Manhattan), une nette amélioration de la qualité de restitution des contours de bâtiments et des superstructures. Le MNE résultat peut ensuite soit être utilisé tel quel, soit vectorisé manuellement ou automatiquement [WF95].

2 Analyse du problème

2.1 Cas étudié et généralisation

La méthode de production de MNE urbain à ISTAR [GGR94] consiste à effectuer des prises de vues photogrammétriques avec des forts recouvrements (80% le long de la trace et 80% entre traces), pour obtenir des images qui, une fois numérisées, auront une résolution entre 30 et 60cm. On calcule ensuite un MNE à partir de chaque paire stéréoscopique, puis les MNE sont fusionnés pour éviter au maximum les problèmes de parties cachées. Quelques zones non renseignées subsistent (parties cachées, erreurs de corrélation, ...). L'exemple montré figure 1 a été obtenu sur le quartier de Manhattan à New York (on voit un morceau du Madison Square Garden en bas à droite), à partir d'images à

40cm de résolution.

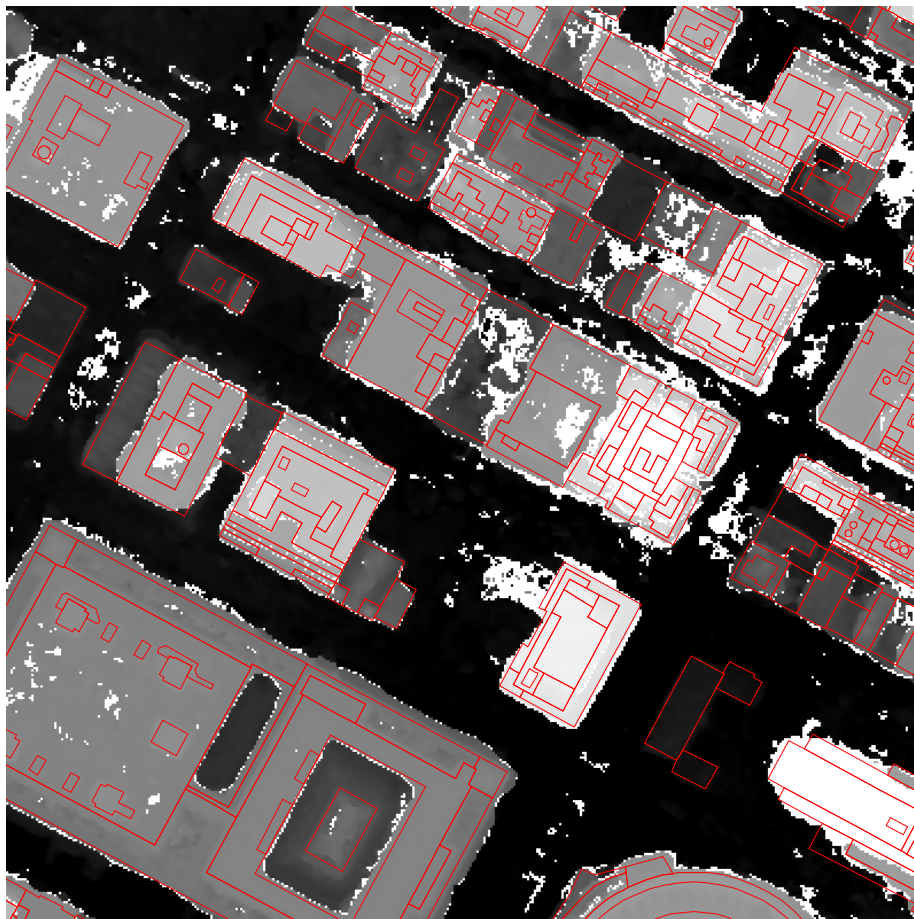


FIGURE 1 – Le MNE brut obtenu par corrélation puis fusion de plusieurs paires stéréo de photographies aériennes : en blanc les points sans information, en rouge les contours 3D réels des bâtiments extraits manuellement à partir des mêmes images.

On remplit ensuite les zones non renseignées du MNE par interpolation contrainte par un critère de régularité, puis on calcule une orthoimage en utilisant toutes les images disponibles, pour éviter les problèmes de faces cachées (Figure 2).

Le but de la méthode proposée ici est d'améliorer la qualité du MNE ainsi calculé en utilisant à la fois le MNE et l'orthoimage. Cependant, cette méthode peut être généralisée à la situation simple où on dispose seulement d'une paire stéréoscopique : le MNE est alors remplacé par la carte de disparité, et l'orthoimage par l'image de référence de la carte de disparité.

Sur un cas comme Manhattan, on comprend bien l'avantage de la stéréoscopie par corrélation : sans utiliser des images à très haute résolution, on a pu calculer un MNE correct, sans faire d'hypothèse sur la forme des bâtiments.

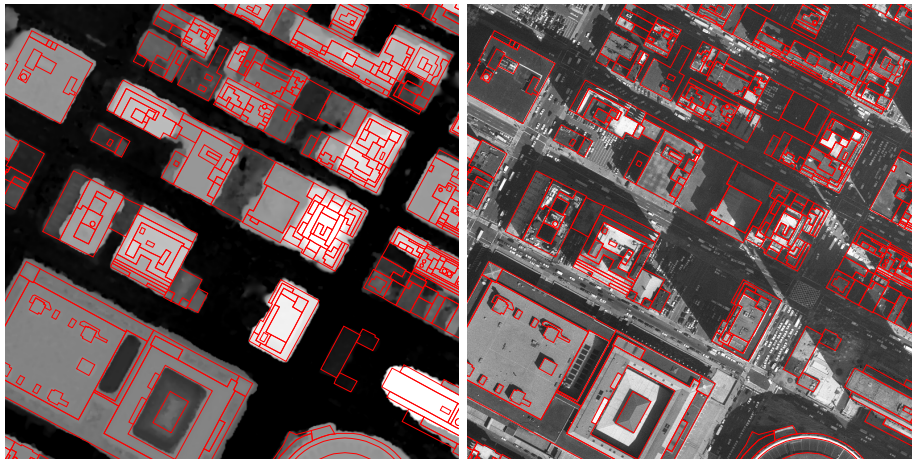


FIGURE 2 – Le MNE complété par une méthode de régularisation anisotrope (gauche), et l’orthoimage calculée à partir de ce MNE (droite).

2.2 Défauts de la corrélation

Pour la fabrication de MNE urbain à partir de prises de vues aériennes, le principal défaut est une erreur sur le positionnement en planimétrie des points 3D. Cette erreur est due au fait que la stéréoscopie par corrélation ne donne pas la disparité du point situé au centre de la fenêtre [Bus97, FW98] : si une zone de la fenêtre de corrélation est plus texturée ou plus contrastée que le reste, c’est plutôt la disparité de cette partie-là qu’on obtiendra. On peut donc avoir une erreur d’estimation planimétrique allant jusqu’à la moitié de la taille de la fenêtre de corrélation, dans les directions x et y de l’image (qui correspondent aux axes cartographiques X et Y dans notre cas).

Ce défaut se manifeste plus particulièrement aux contours des bâtiments. En effet, ce sont des discontinuités de profondeur, et les deux côtés de la discontinuité (sol et toit) ont souvent des contrastes très différents, dus à la différence d’éclairage (ombres portées au sol). Le plus souvent, donc, les bâtiments vont « baver », surtout du côté des ombres, et leur taille sera plus large que la taille réelle (Figure 3).

Un autre défaut souvent observé est que dans les zones d’ombres, où la dynamique est faible, la corrélation ne donne aucun résultat. Si la rue est étroite ou que les bâtiments sont hauts, on peut n’avoir presque aucune information au niveau de la rue (en haut à droite de la Figure 1).

2.3 Solutions

La solution la plus courante consiste à utiliser les fenêtres adaptatives [KO94]. Dans cette méthode, la taille de la fenêtre de corrélation diminue près des contours des images, et augmente aux endroits de faible contraste. Cette méthode fonctionne plutôt bien en milieu urbain [Lot96], mais est plutôt complexe à mettre en œuvre, ce qui nous a poussé à étudier vers une autre solution.

Cette solution est basée sur l’observation suivante : les zones que la fenêtre

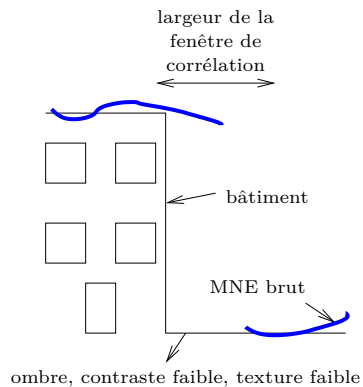


FIGURE 3 – La stéréo par corrélation accroche plus aux endroits contrastés (le plus souvent les toits, qui ne sont pas à l’ombre). La largeur de la zone affectée peut atteindre la taille de la fenêtre de corrélation rapportée au sol (prises de vues aériennes).

de corrélation a «accroché» correspondent aux zones fortement contrastées de l’orthoimage. On doit donc pouvoir chercher à corriger le MNE, en utilisant l’orthoimage, en modifiant les valeurs de disparité qui se situent dans des zones peu contrastées proches de zones contrastées. Ces zones à corriger sont de largeur au plus égale à la moitié de la taille de la fenêtre de corrélation rapportée au sol.

3 Approche générale

La première approche est basée sur une technique de diffusion anisotrope [PM90], qui agit sur le MNE, et dont le coefficient de conduction est calculé à partir de l’orthoimage. Cette approche est générale et peut s’appliquer à tout type de MNE (même non urbain), voire même à une paire stéréoscopique quelconque avec la carte de disparité associée.

On commence par filtrer le MNE brut, contenant déjà des zones non renseignées dues aux erreurs de corrélation, pour enlever les valeurs aberrantes, puis on remplit les zones non renseignées en tenant compte du contraste local de l’orthoimage. On applique ensuite un nombre fini d’itérations de diffusion anisotrope.

3.1 Image de contraste

Le remplissage et la diffusion sont basés sur une image représentant le contraste local de l’image de référence (l’orthoimage). Cette mesure doit avoir une précision de localisation meilleure que la taille de la fenêtre de corrélation, puisqu’on cherche à déterminer quels sont les endroits qui ont pu «accrocher» la corrélation à l’intérieur-même de cette fenêtre.

Le problème est évidemment de trouver une bonne mesure du contraste local.

On peut par exemple proposer :

$$\overline{(I - \bar{I})^2} \quad (1)$$

où I est l'orthoimage, et le surlignage représente la convolution par un noyau de lissage de petite dimension (par ex. une gaussienne avec $\sigma = 1.5$ pixel). Les meilleurs résultats ont cependant été obtenus avec une classique extraction de contours (nous avons utilisé le filtre de Kirsch [BB82]), sans doute grâce au fait que les zones contrastées sont, en milieu urbain, le plus souvent des contours. Si l'image de référence contient surtout des zones contrastées dues à des textures (une image SPOT ou des photographies aériennes en milieu rural ou résidentiel par ex.), on préférera une mesure similaire à 1.

Lors des opérations de remplissage et de diffusion, un seuil est utilisé (κ , le même pour les deux opérations) qui correspond à la limite, dans l'image de contraste, entre les zones contrastées et les zones peu contrastées (ombres).

3.2 Filtrage

L'élimination des valeurs aberrantes du MNE aurait pu être réalisée par des opérations morphologiques (en éliminant les petits îlots dont l'altitude est très différente du voisinage), mais cette opération éliminerait également des valeurs valides et cohérentes mais éparses. Nous avons donc préféré une méthode basée sur des histogrammes d'élévation.

L'image est divisée en tuiles (par ex. $10\text{m} \times 10\text{m}$), et pour chaque tuile on calcule un histogramme de pas en élévation p fixé (par ex. 1m) de l'élévation du MNE de cette tuile et de ses 8 voisines. Pour chaque valeur de l'histogramme, si la surface associée à cette valeur est inférieure à un seuil (par ex. 4m^2), tous les points concernés sont marqués comme «invalides».

Le traitement par tuiles paraît peu naturel, et les paramètres plutôt abstraits, mais la méthode s'est montrée rapide et efficace sur des paysages urbains variés.

3.3 Remplissage

La deuxième phase consiste à remplir les valeurs «invalides», dues à des erreurs de corrélation ou résultant du filtrage, par des valeurs probables. Ce remplissage consiste en une succession de dilatations basées sur l'image de contraste : un pixel du MNE dont le contraste est inférieur à un seuil κ est rempli par la médiane de ses voisins valides dont le contraste est inférieur à κ . Le seuil κ est progressivement augmenté, jusqu'à ce que tout le MNE soit rempli de valeurs valides.

Cette méthode de remplissage a pour effet de remplir les zones invalides de préférence à partir des zones de contrastes équivalent. En effet, un pixel situé à côté d'un pixel plus contrasté et valide et qui est effectivement à la même hauteur a peu de chance d'être invalide. Il est donc raisonnable de faire hériter aux pixels invalides l'altitude de pixels moins contrastés. Les pixels invalides qui correspondent à des maxima locaux de contraste ne peuvent être remplis par ce processus, et on leur affecte l'altitude médiane de leurs voisins valides.

À la fin de cette étape, les zones aberrantes du MNE ont été filtrées et les zones non renseignées remplies par des valeurs plausibles, mais toutes les

parties considérées comme valides n’ont pas été modifiées. L’étape suivante (la diffusion) consiste à modifier légèrement ces valeurs valides pour tenir compte des effets dus à la taille de la fenêtre de corrélation.

3.4 Diffusion

La diffusion consiste à observer l’évolution dans le temps d’une grandeur scalaire (en général l’intensité) définie dans un espace de dimension n (2 pour une image), soumise à une loi d’évolution similaire à l’équation de la chaleur. La diffusion anisotrope [PM90] consiste à ajouter un coefficient de conduction calculé à partir des dérivées de l’image, de manière à ce que la diffusion conserve les contours présents dans l’image initiale.

Dans notre cas, nous avons à notre disposition deux grandeurs scalaires : l’altitude du MNE et l’intensité de l’orthoimage. Nous allons faire diffuser l’image d’altitude, en utilisant l’image d’intensité (plus précisément de contraste) pour calculer le coefficient de conduction. Ainsi, à un endroit où l’image est très contrastée, la donnée d’altitude est en général fiable, et elle ne doit pas être modifiée. Par contre, si il y a à une distance inférieure à la taille de la fenêtre de corrélation une zone peu contrastée, elle aura probablement été masquée par la zone contrastée lors de la stéréoscopie par corrélation, et son élévation doit donc être corrigée à partir de l’élévation des zones peu contrastées de son voisinage.

L’évolution de l’élévation d’un point entre les dates t et $t + 1$ s’écrit alors :

$$Z(t + 1) = Z(t) + \lambda \sum_{v \in V} k_v(Z_v(t) - Z(t)) \quad (2)$$

où λ est le coefficient de diffusion (0,24 dans nos expériences), k est une fonction positive décroissante du contraste, avec $k(0) = 1$ (par ex. $k(x) = e^{-\frac{x^2}{\sigma^2}}$ ou $k(x) = \frac{1}{1 + \frac{x^2}{\sigma^2}}$), et V est l’ensemble des 4 ou 8 voisins du point.

En appliquant cette formule de diffusion, on note de nettes améliorations du MNE, notamment au niveau des contours des bâtiments, des superstructures, et de la végétation qui sont beaucoup plus francs. Cependant, la méthode et les résultats souffrent de plusieurs défauts en milieu urbain :

- le nombre d’itérations nécessaires pour résorber les erreurs dues aux immeubles les plus hauts est important (plus de 200) ;
- à cause du grand nombre d’itérations, les erreurs du MNE ont une zone d’influence dans l’image qui va bien au-delà de la taille de la fenêtre de corrélation ;
- si un immeuble n’a pas un contour partout contrasté, l’élévation «fuit» par les ouvertures et son altitude diminue ;
- de même, les débordements des immeubles peuvent, par diffusion, relever le niveau du sol de plusieurs mètres.

Ces défauts disparaissent lorsqu’on applique la méthode à un MNE non urbain ou de plus faible résolution (fait par exemple à partir d’images SPOT). Nous avons donc cherché un processus à appliquer avant la diffusion, qui permette d’en diminuer le nombre d’itérations.

4 Améliorations spécifiques au MNE urbain

4.1 Segmentation et érosion

Le défaut le plus important à éliminer dans le MNE urbain obtenu par corrélation est le débordement des toits des immeubles sur les zones peu contrastées, qui sont le plus souvent à une altitude inférieure (par ex. des ombres). Si on arrive à corriger ce défaut asymétrique, la phase de diffusion pourra s'effectuer plus facilement, et en un nombre d'itérations largement réduit. On va donc essayer de délimiter ces régions pour les éliminer. Le problème est de ne pas détruire de bâtiments ou de superstructures lors de cette opération : pour beaucoup d'applications, notamment les télécoms, on préfère avoir des bâtiments un peu trop gros plutôt que de voir disparaître les plus petits.

La première phase consiste à effectuer une segmentation du MNE en régions. On commence par enlever du MNE tous les points pour lesquels le contraste est supérieur au seuil κ . On segmente ensuite les points restants en zones homogènes en altitudes (par ex. la différence en élévation de deux points voisins doit être inférieure à 1m).

Dans la deuxième phase, on commence par déterminer quelles régions sont en conflit. La région A attaque la région B si toutes ces conditions sont réunies :

- elles ont une frontière commune ;
- le rapport de leurs cardinaux est supérieur à un seuil c , $\frac{\text{card}(A)}{\text{card}(B)} > c$ ($0,5 < c < 1,5$ dans nos expériences) ;
- l'altitude moyenne de A est inférieure à celle de B (les erreurs se produisant très rarement dans l'autre sens).

Ensuite, pour chaque région attaquée, on effectue un certain nombre d'érosions (correspondant à la moitié de la taille de la fenêtre de corrélation) en ne considérant que cette région et l'attaquant de plus grand cardinal. Ici, une érosion consiste à attribuer à un pixel l'élévation minimum de ses voisins.

Ces opérations permettent de s'approcher sensiblement du résultat voulu : S'il reste quelques erreurs, l'étape de diffusion permettra de les résorber ou au moins de les atténuer. Une quinzaine d'itérations de diffusion suffit alors à corriger les derniers défauts du MNE.

4.2 Méthode en deux passes

La méthode en deux passes consiste à prendre le résultat de l'application des processus de segmentation et érosion puis diffusion, et à le comparer au MNE avant ces opérations (c.-à-d. après le remplissage). Tous les pixels dont la variation est supérieure à un seuil donné sont marqués comme faux dans le MNE initial, et on effectue une série d'érosions sur ces pixels. On effectue ensuite quelques itérations de diffusion anisotrope, principalement pour améliorer la qualité des contours d'immeubles. Cette deuxième passe permet de corriger quelques erreurs qui auraient été commises pendant la première passe, notamment lors de la phase de segmentation/érosion, mais ne corrige pas tous les défauts du MNE brut.

5 Résultats

Pendant la phase de développement de la méthode, des tests ont été effectués sur un MNE de la presqu'île de Manhattan, et sur la ville de Fribourg en Allemagne. Malgré des paysages urbains complètement différents, la méthode a été aussi efficace sur les deux villes, le seul paramètre à régler étant le seuil κ sur l'image de contraste. En pratique, le MNE est découpé en tuiles d'environ 500×500 pixels, avec une marge de recouvrement, la méthode est appliquée sur chaque tuile, puis les tuiles sont assemblées. Les résultats sur Manhattan sont montrés Figures 4 et 5 pour la méthode en une passe et en deux passes. Une vue tridimensionnelle (Figure 6) permet de mieux apprécier les effets de la méthode, notamment sur les contours d'immeubles complexes.

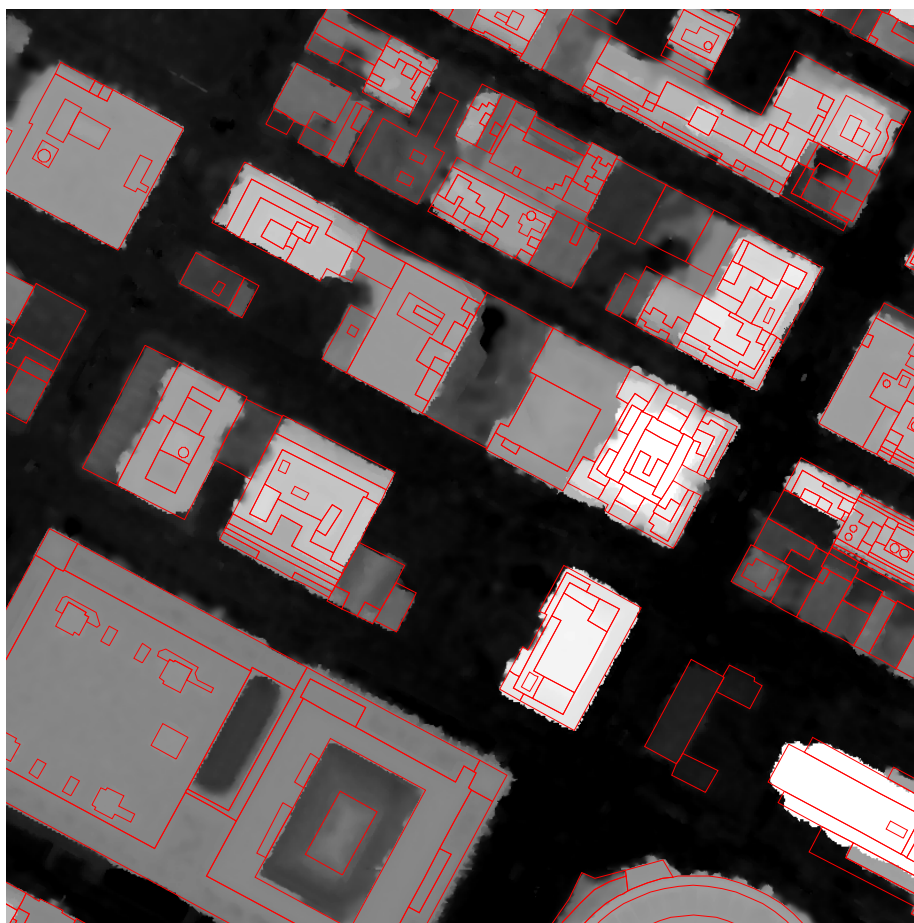


FIGURE 4 – Le MNE après traitement : filtrage, remplissage, régionnement et érosions, diffusion.

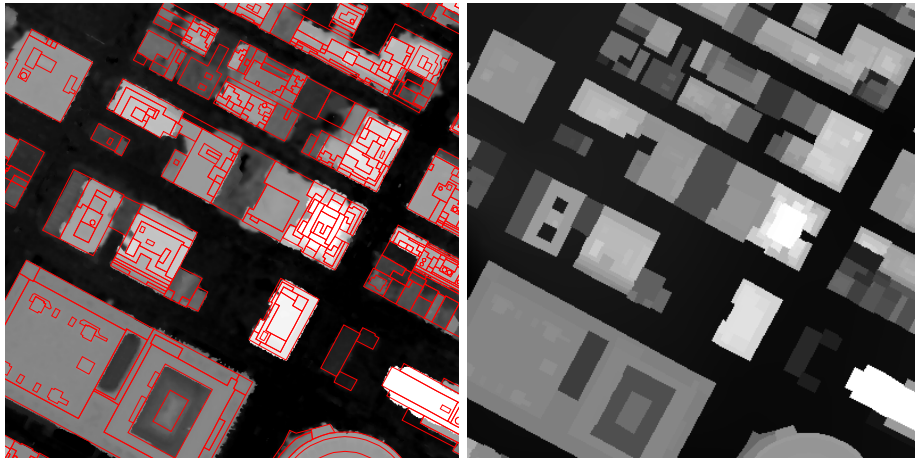


FIGURE 5 – Le MNE après traitement, en utilisant la méthode en deux passes, et le MNE idéal calculé à partir des vecteurs 3D extraits manuellement (droite).

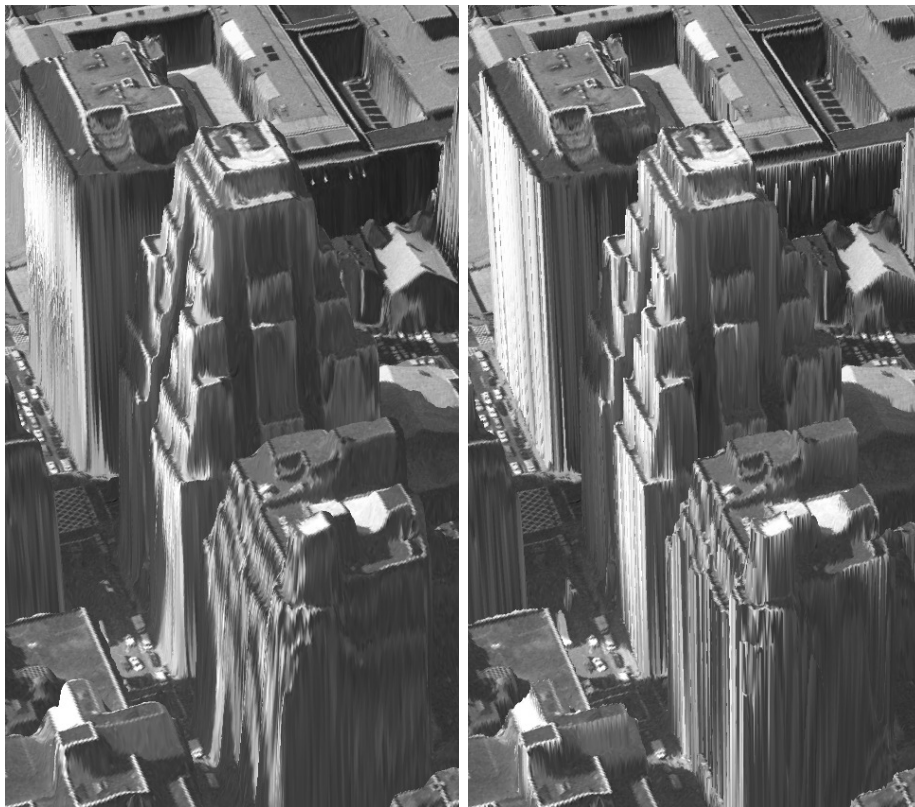


FIGURE 6 – Vues 3D du MNE brut (gauche) et du résultat (droite). Les contours des immeubles et les discontinuités de profondeur sont plus franches et mieux localisées.

6 Conclusion

La stéréoscopie par corrélation possède l'avantage de pouvoir restituer des formes complexes sans connaissance a priori sur le modèle d'objet observé. En milieu urbain, elle est le moyen le plus sûr d'obtenir un MNE, principalement à cause de la diversité des architectures. Malheureusement elle possède aussi quelques défauts, dont celui de faire déborder les toits des immeubles en dehors de leurs limites.

L'approche que nous avons présentée visait à corriger a posteriori les défauts de la corrélation classique. La première idée fut d'appliquer une technique mathématiquement séduisante, la diffusion anisotrope, mais à force d'essais nous avons abouti à un enchaînement d'opérations essentiellement morphologiques, qui tient plus de la recette de cuisine, et où la diffusion est cantonnée à un rôle de finition (*polishing*).

Il reste sans doute beaucoup à faire pour améliorer cette méthode, et l'effort le plus important est à mettre sur la clé de voûte de cette méthode, l'image de contraste. En effet, il faudrait pouvoir trouver un opérateur qui ait la précision d'un détecteur de contours, mais qui puisse aussi avoir une réponse positive sur une texture aléatoire. L'idéal serait d'avoir un opérateur qui soit en plus normalisé pour permettre de s'affranchir de la détermination du seuil κ , et qui donne des zones mieux fermées pour éviter les «fuites» lors de la diffusion.

Références

- [BB82] D.C. BALLARD ET C.M. BROWN, *Computer Vision*, Prentice-Hall, Englewood Cliffs, New Jersey, 1982.
- [Bus97] RALF BUSCHMANN, Efficiency of displacement estimation techniques, *Signal Processing : Image Communication* **10** (1997), 43–61.
- [FW98] LUTZ FALKENHAGEN ET THOMAS WEDI, Improving block-based disparity estimation by considering the non-uniform distribution of the estimation error, in *3D Structure from Multiple Images of Large-Scale Environments, proceedings of SMILE'98 workshop*, R. Koch et L. Van Gool (édité par), pages 93–108, *Lecture Notes in Computer Science* numéro 1506, Springer-Verlag, 1998.
- [GGR94] L. GABET, G. GIRAUDON ET L. RENOARD, Construction automatique de modèle numérique de terrain à haute résolution en zone urbaine, *Bulletin de la Société Française de Photogrammétrie et de Télédétection* **4**,135 (1994).
- [KO94] T. KANADE ET M. OKUTOMI, A stereo matching algorithm with an adaptive window : Theory and experiment, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**,9 (septembre 1994), 920–932.
- [Lot96] JEAN-LUC LOTTI, *Mise en correspondance stéréo par fenêtres adaptatives en imagerie haute résolution*, PhD thesis, Université de Nice - Sophia Antipolis, France, février 1996.
- [PM90] P. PERONA ET J. MALIK, Scale-space and edge detection using anisotropic diffusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**,7 (juillet 1990), 629–639.

- [WF95] UWE WEIDNER ET W. FÖRSTNER, Towards automatic building extraction from high resolution digital elevation models, *ISPRS Journal* **50,4** (1995), 38–49.