

Mini-projet IFT786



Reconstruction d'une ville en 3D à partir
d'images photo : essais sur Sherbrooke

par *Éric Beaudry*

<http://planiart.usherb.ca/~eric/ift786/>

Session été 2005

Plan de la présentation

- ⇒ Motivation de la reconstruction 3D de scènes urbaines
- ⇒ L'état de l'art
- ⇒ Approche choisie
- ⇒ Travail effectué
- ⇒ Résultats
- ⇒ Conclusion et discussion

Motivation de la reconstruction 3D de scènes

- ⇒ Réalité augmentée : présentation de projets de construction avec des images et animations virtuelles
- ⇒ Étude spatiale : étudier la transmission des signaux micro-onde dans une ville
- ⇒ Promotion d'un événement d'envergure internationale (ex.: Montréal 2005 – Championats FINA, les JO, ...)

État de l'art

- ⇒ Approches très variées
 - Sources de données
 - Degré d'automatisation
 - Degré de précision

Principe de base

- ⇒ Décomposable en sous-problèmes :
 - (1) Localisation de la base des bâtiments
 - (2) Relief du terrain
 - (3) Hauteur des bâtiments
 - (4) Forme des bâtiments (principalement les toitures)
 - (5) Réalisme (couleur / texture)

Localisation des bâtiments

➔ Plan de cadastre

- Emplacement précis de la base des bâtiments (polygone 2D)
- Généralement à jour et fidèle à la réalité
- Annotations : type de bâtiment, nombre d'étages, règlements, etc.
- Format électronique nécessaire (dans SIG)
- Données manquantes :
 - Hauteur
 - Forme des murs et type de toiture
 - Couleur et type de revêtement

Localisation des bâtiments (2)

- ⇒ Imagerie aérienne ou satellitaire [3, 7]
 - Segmentation d'images multispectrales en tenant compte des caractéristiques géométriques (forme rectangulaire)
 - Besoin d'images à haute-résolution (~1m)
 - Sensible aux obstacles (arbres cachant une partie d'une maison)

Localisation des bâtiments (3)

➔ Analyse du MNE et MNT

- Le modèle numérique d'élévation (MNE) est l'élévation réelle d'une scène (inclus arbres et bâtiments)
- Le modèle numérique de terrain (MNT) est la modélisation des courbes de niveau du sol sans ce qui a au dessus.
- Différente: $MNT - MNE =$ tout ce que dépasse le sol (arbres + bâtiments)
- Pour discriminer les arbres des bâtiments : analyse de forme et/ou images multispectrales[7]
- Remarque: le MNT est souvent déduit d'un MNE!

Relief du Terrain (1)

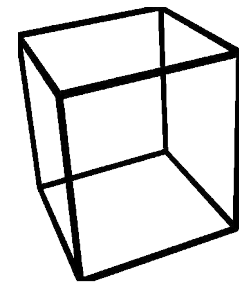
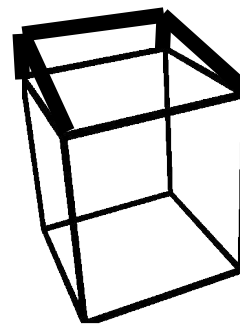
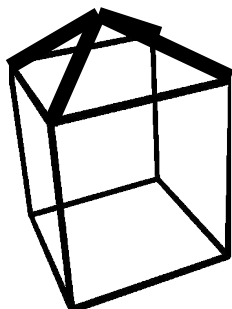
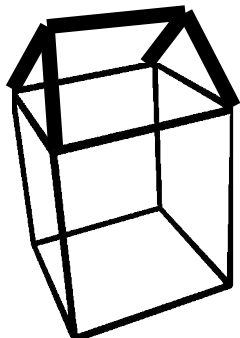
- ⇒ Représentation: modèle numérique de terrain (MNT)
 - Image géoréférencée : chaque pixel représente l'élévation du sol
- ⇒ Plusieurs approches possibles
 - Dériver le MNT à partir d'un MNE
 - Stéréovision sur paires d'images aériennes [9]
 - Utilisation d'un laser sur un avion pour obtenir la profondeur [7, 2]

Hauteur de bâtiment

- ➔ $H = MNE - MNT$ (si d'une bonne résolution d'élévation) [2]
- ➔ L'utilisation de photos prises au sol : afficher la scène 3D (plan de cadastre superposé au MNT) sous le même angle que la photo afin d'éditer manuellement les hauteurs [4]
- ➔ Analyse des ombrages : la hauteur des bâtiments est proportionnel à la longueur des ombres projetées au sol [10]

Forme des bâtiment : toiture

- ➔ La toiture est la principale élément de forme à reconstruire
- ➔ Utilise généralement des primitives simples pour modéliser la forme du bâtiment
- ➔ Rechercher des surfaces planes dans les MNE ou nuages de points 2.5D/3D



Détection automatique d'un groupe de toitures

- ➔ Décomposition de la base d'un bâtiment en formes simples (ex.: rectangles)

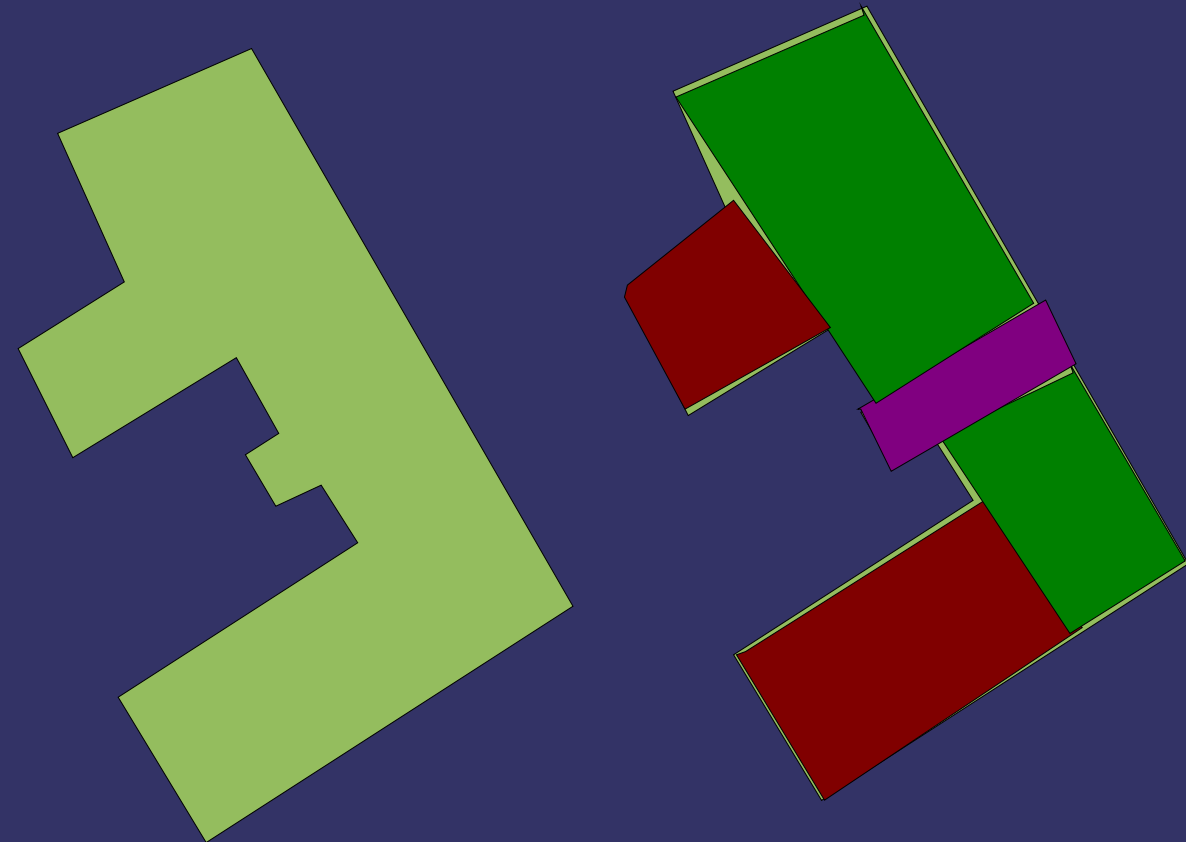
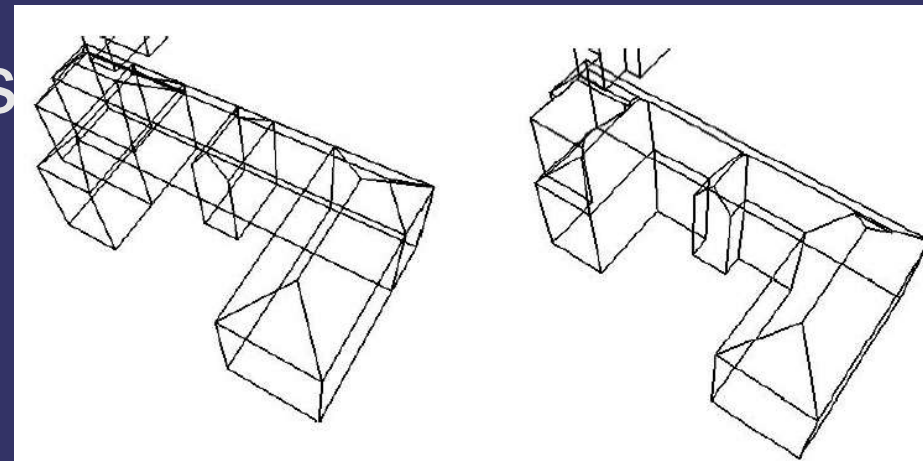


Figure tirée de [2]

Détection automatique d'un groupe de toitures (2)

- ➔ Utilisation des moindres carrés pour trouver le type de toit qui correspond le mieux au MNE [2]
- ➔ Segmentation des toiture (région prise dans le plan de cadastre sur les images) [2]
- ➔ Fusion (union) des primitives de solides obtenues [2]



Figures tirées de [2]

Reconstruction à partir d'une séquence vidéo

- ➔ D'autres approches complètement différentes
- ➔ Exemple : à partir d'une séquence vidéo[5]
 - Aucune information sur les paramètres de la caméra
 - Aucun points de référence
 - Auto-calibrage
 - Reconstruction à une échelle prêt
 - Suivit de « feature » dans les séquences d'images
 - Détection de lignes et d'angles
 - Obtenir automatiquement la position relative des caméras

Choix de l'approche (1)

- ⇒ Site choisi : Sherbrooke (campus ouest UdeS)
- ⇒ Évaluation des données disponibles
 - Plan de cadastre de Sherbrooke :
 - Disponible en Shapefile (difficile à lire (parser) ?)
 - Données non superposables (problème géoréférencement)
 - MNT : disponible dans les BDTQ (format ArcInfo *.e00) à faible résolution planaire (10 à 20m) et d'élévation (5 à 10m)
 - MNE : (aucun)

Choix de l'approche (2)

- ⇒ Évaluation des données disponibles (suite)
 - Images satellitaires de Sherbrooke
 - Ikonos: 2002 et 2004 (panchromatique à 0.8m et multispectrale à 3.2m)
 - SPOT (faible résolution)
 - Images aériennes
 - Photos aériennes de la Photocartothèque québécoises : Estrie disponible au complet
 - Photos à échelle 1/15000
 - Photos 1/15000 les plus récentes : juillet 1998

Choix de l'approche (3)

⇒ Méthodologie choisie

- Générer un MNE à partir d'une paire de photos aériennes
- Tenter de détecter automatiquement les bâtiments à partir du MNE
- Si nécessaire, créer un «pseudo» plan de cadastre manuellement sur une petite zone
- Faire une première attribution des hauteurs en utilisant le MNE
- Raffiner les hauteurs à l'aide de photos prises au sol

Acquisition des images aériennes

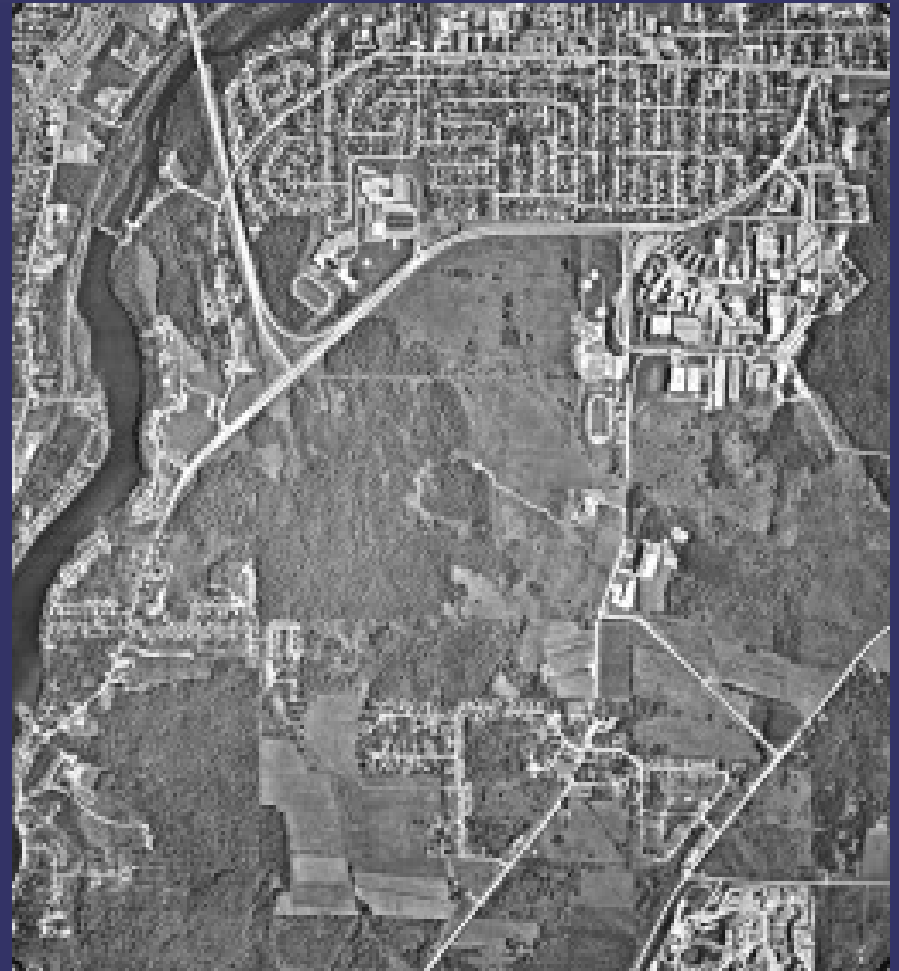
- ➔ Photos en format imprimé seulement
- ➔ Numérisation des images
- ➔ Correction de l'alignement des marques par une rotation, un cisaillement et un découpage



Images



hmq98-131-181



hmq98-131-180

Images captées le 19 juillet 1998.
Taille: 10036x10008 pixels

Hypothèses pour stéréovision

⇒ Paramètres intrinsèques

- Parfaitement connus : rapport de calibrage disponible

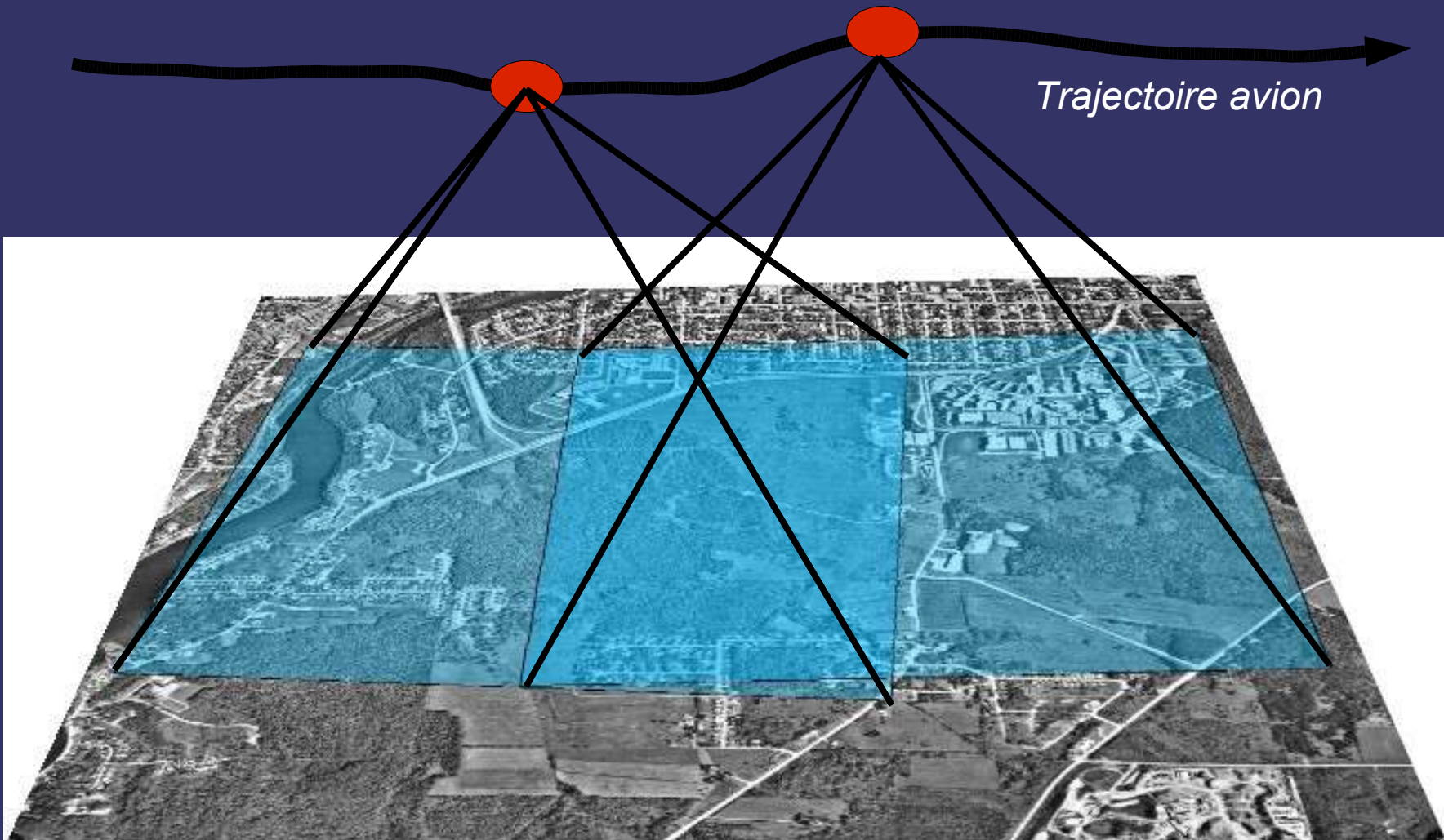
⇒ Paramètres extrinsèques

● Grossièrement connus

- Altitude = 2724m (8936')
- Positions : $p1=(-71.91^\circ, 45.37^\circ)$ $p2=(-71.93^\circ, 45.37^\circ)$
- Directions de regard de la caméra pour les photos
 - Proches de la perpendiculaire du sol
 - Presque parallèles
- Axe des X de l'image presque aligné avec axe ouest-est

Paramètres extrinsèques (1)

→ L'avion subit des perturbations



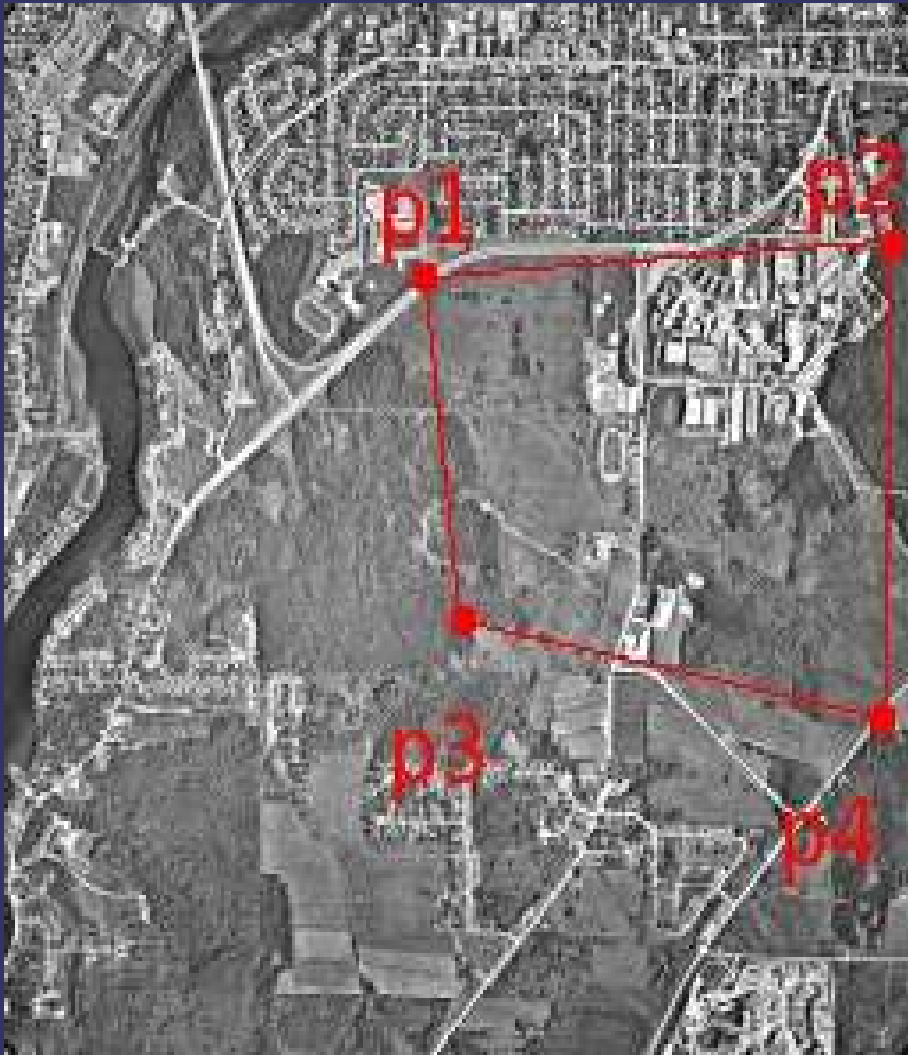
Paramètres extrinsèques (2)

⇒ Solutions possibles

- Utiliser l'algorithme à 8 points [12,] ?
- Corriger et rectifier l'image de droite par rapport à l'image de gauche (« normalized images pairs ») [8]
- Transformations : rotation + perspective

Rectification (1)

→ Trouver 4 points de correspondance



Rectification (2)

- ⇒ Estimer la rotation
 - Comparer les orientations des droites P1-P2 et P3-P4 dans les deux images
- ⇒ Étirer l'image de droite dans le sens de la hauteur seulement pour faire correspondre les lignes
 - Utilisation d'un « warp » perspective qui fait correspondre un quadrilatère source vers un autre
 - *Réalisé avec librairie JAI de Sun Microsystems (Java Advanced Imaging)*

Rectification (3)

⇒ Pseudo-code

```
# angle image gauche
langle = (atan2(p2l.x-p1l.x, p2l.y-p1l.y) +atan2(p4l.x-p3l.x, p4l.y-p3l.y)) / 2
# angle image droite
rangle = (atan2(p2r.x-p1r.x, p2r.y-p1r.y) +atan2(p4r.x-p3r.x, p4r.y-p3r.y)) / 2
# première transformation : une rotation
t = rotation(rangle - langle)
# transformer les points de l'image de droites
p1r = t(p1r)
p2r = t(p2r)
p3r = t(p3r)
p4r = t(p4r)
# Ajuster les Y de l'image droite à celle de l'image gauche
t *= perspective( p1r -> Point(p1r.x, p1l.y),
                 p2r -> Point(p2r.x, p2l.y),
                 p3r -> Point(p3r.x, p3l.y),
                 p4r -> Point(p4r.x, p4l.y))
imageresult = t(imagesource)
```

Rectification (4)



Image de droite brute



Image de droite rectifiée

Paramètres extrinsèques (3)

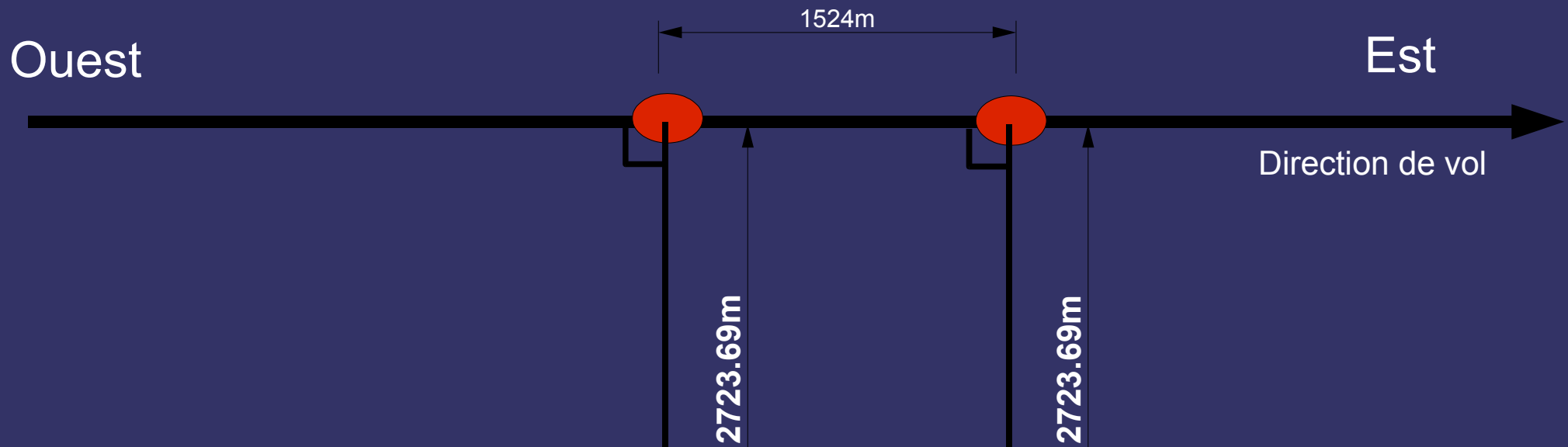
⇒ Nouvelles hypothèses

- Les directions de regard sont parfaitement parallèles
- Les positions de la caméra sont à la même altitude
- Le déplacement de la caméra s'est fait dans la direction ouest-est qui coïncide parfaitement avec l'axe de X de l'image
- Géométrie épipolaire : point \rightarrow ligne correspondante

Paramètres extrinsèques (4)

- ⇒ (suite)
- ⇒ Estimer le déplacement t de la caméra
 - Mesurer la distance entre les pixels au centre des 2 images
 - Utilisation d'une image géoréférencée Ikonos pour obtenir les coordonnées de ces 2 points
 - (Bonne alternative à se déplacer sur le terrain avec un GPS à la main!)
 - $t=1524m$

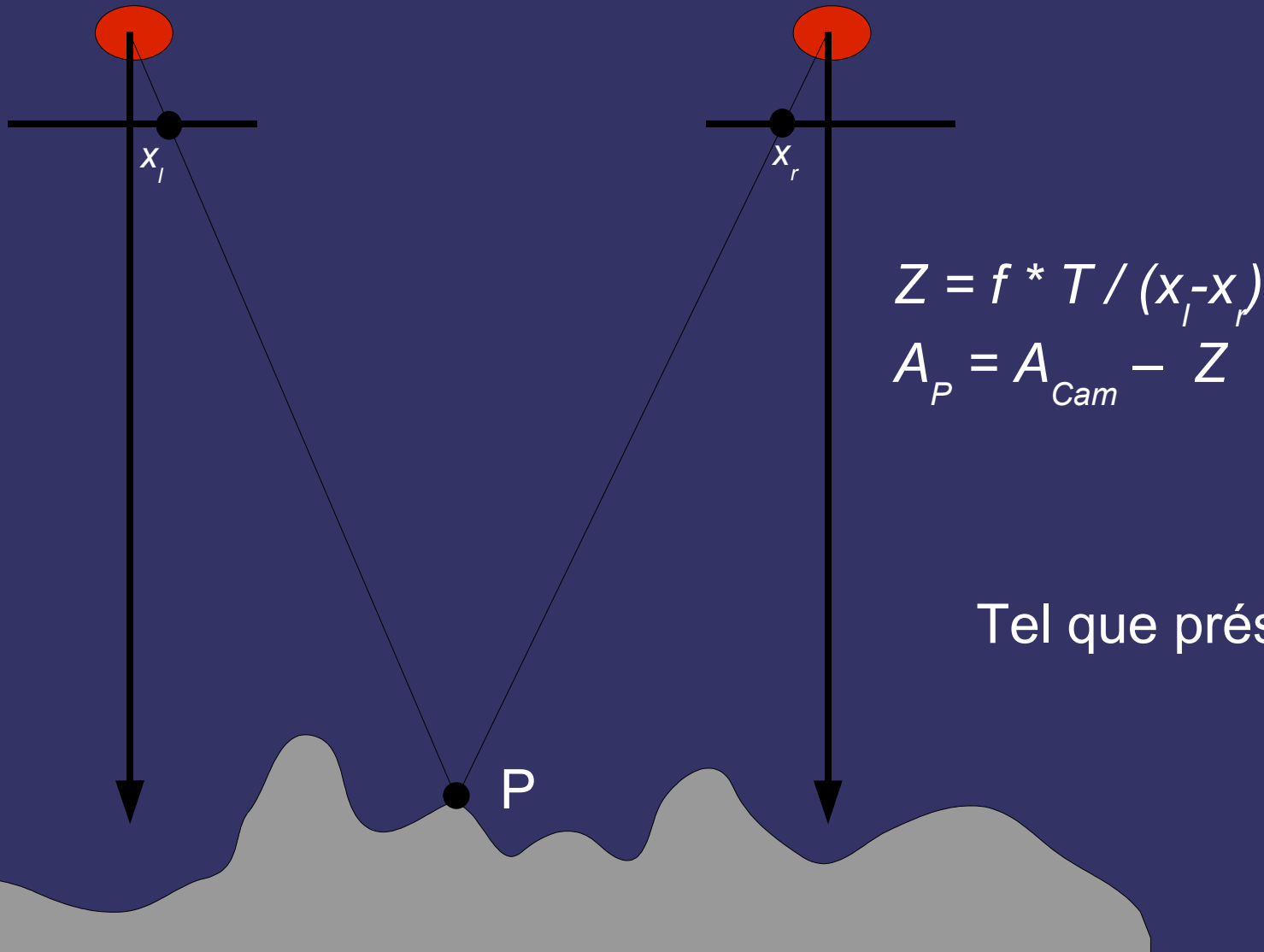
Nouveau modèle



Calcul simple de l'altitude

image
gauche

image
droite



$$Z = f * T / (x_l - x_r)$$
$$A_P = A_{Cam} - Z$$

Tel que présenté dans [12]

Algorithme de stéréovision (1)

- ⇒ Mise en correspondance : corrélation
- ⇒ Premier essais
 - Très lent
 - Très bruité
- ⇒ Correctifs appliqués
 - Le relief du terrain varie peu localement --> borner la portion de la ligne où chercher dans l'image de droite (chercher dans le segment $[-15m, 15m]$ (danger: désynchronisation)
 - Utiliser une fonction d'énergie pour ignorer les régions trop homogène
 - Réduction de la taille des images : 2509x2502

Algorithme de stéréovision (2)

T : déplacement de la caméra dans le sens ouest-est

f : distance focale

l : largeur du plan image

altitude[*, *] = Indéfinie

Pour tout (x, y) de l'image de gauche

altitude_min = Min{ Min{altitude(x-10 à x-1, y)}, Min{altitude(x, y-10 à y-1)} }

altitude_max = Max{ Max{altitude(x-10 à x-1, y)}, Max{altitude(x, y-10 à y-1)} }

altitude_min -= 15

altitude_max += 15

scene_min = camera_gauche.convertir_en_coor_scene(x, y, altitude_min)

scene_max = camera_gauche.convertir_en_coor_scene(x, y, altitude_max)

rx1 = camera_droite.convertir_en_coor_image(scene_min)

rx2 = camera_droite.convertir_en_coor_image(scene_max)

mx = trouver_correspondance(rx1, y-1, rx2, y+1)

if(energy>seuil)

altitude[x,y] = f * T / (mx-x) * l

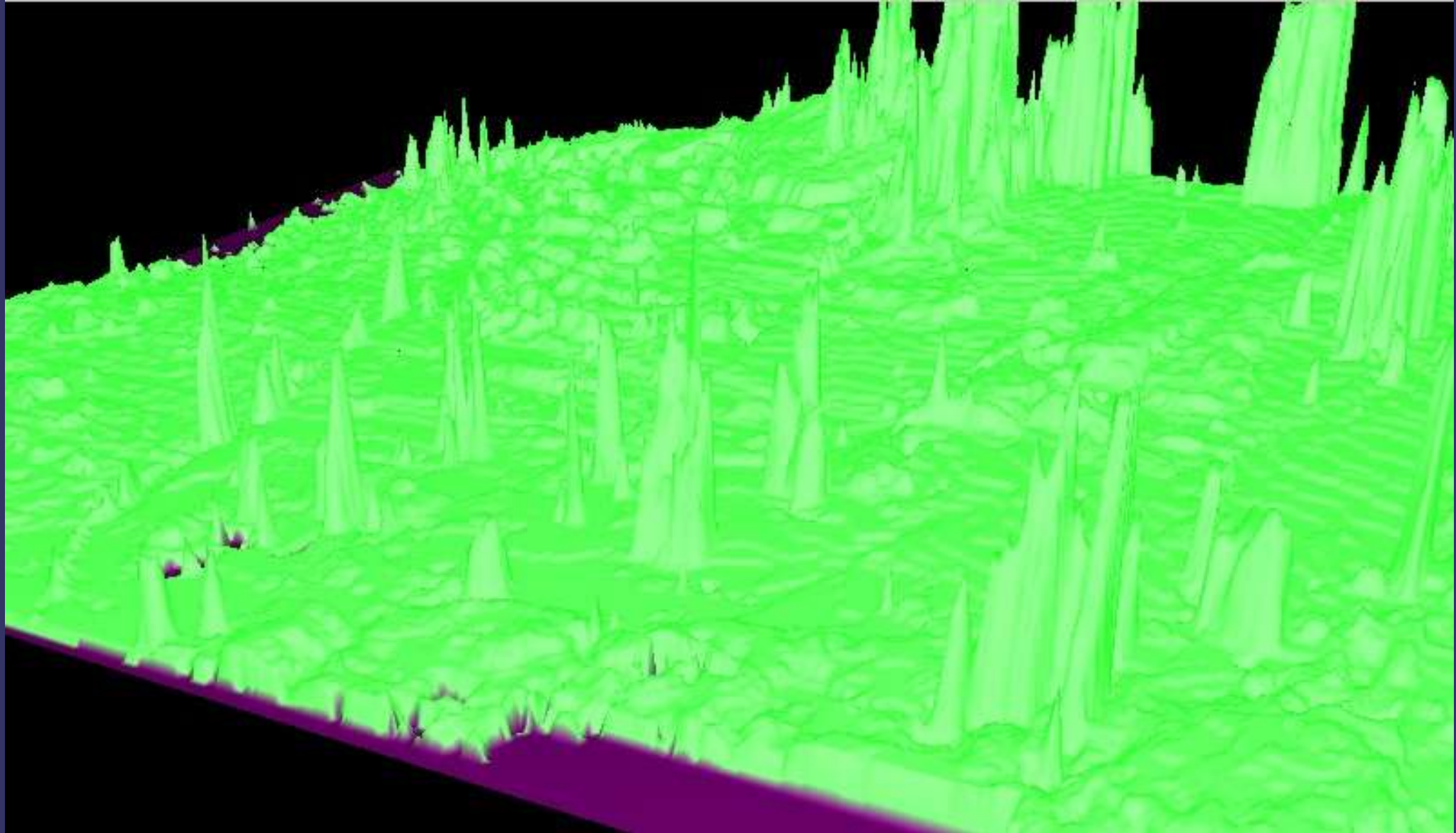
else

altitude[x,y] = dernière_valeur(altitude[,])

MNE résultant (fenêtre = 19x19)



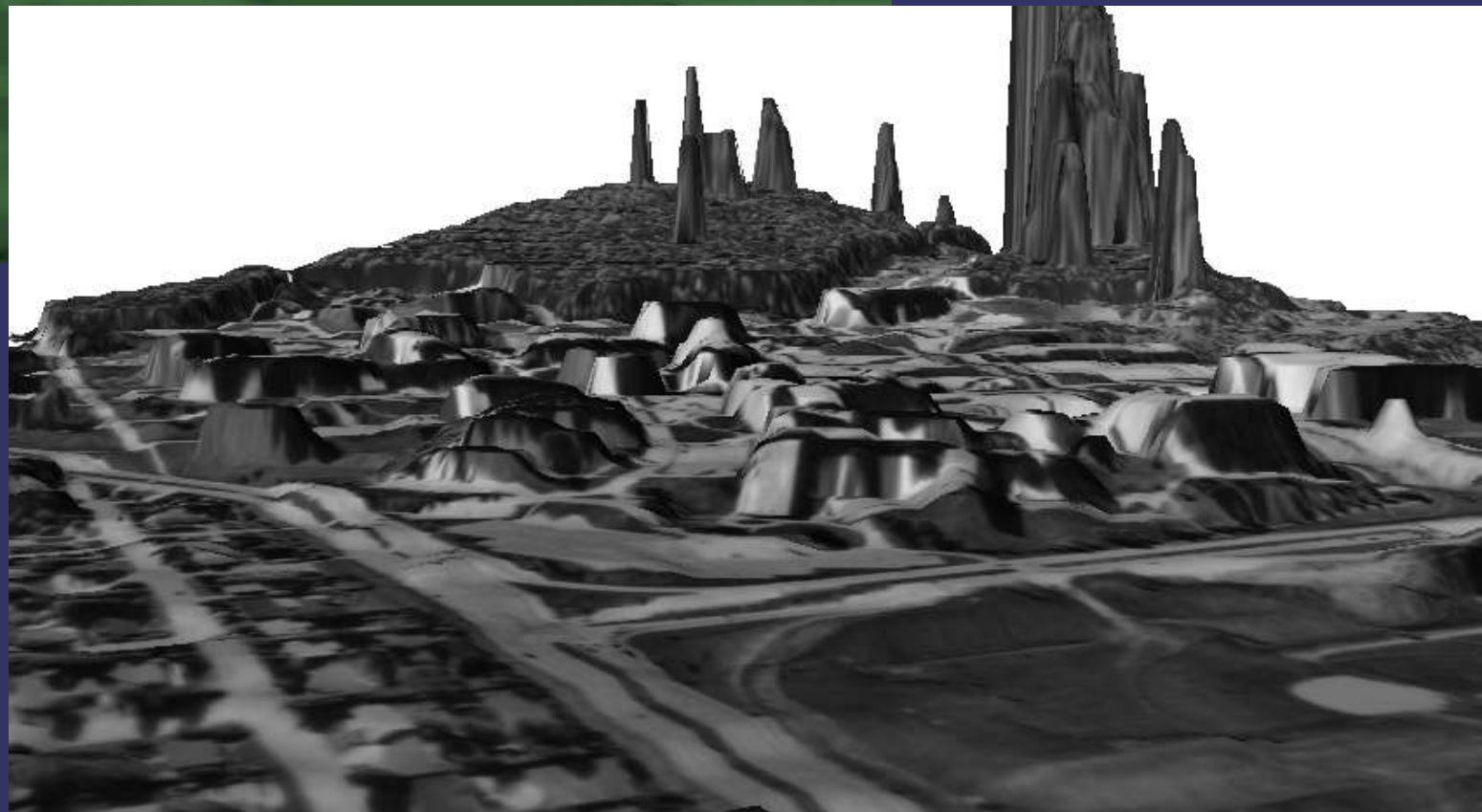
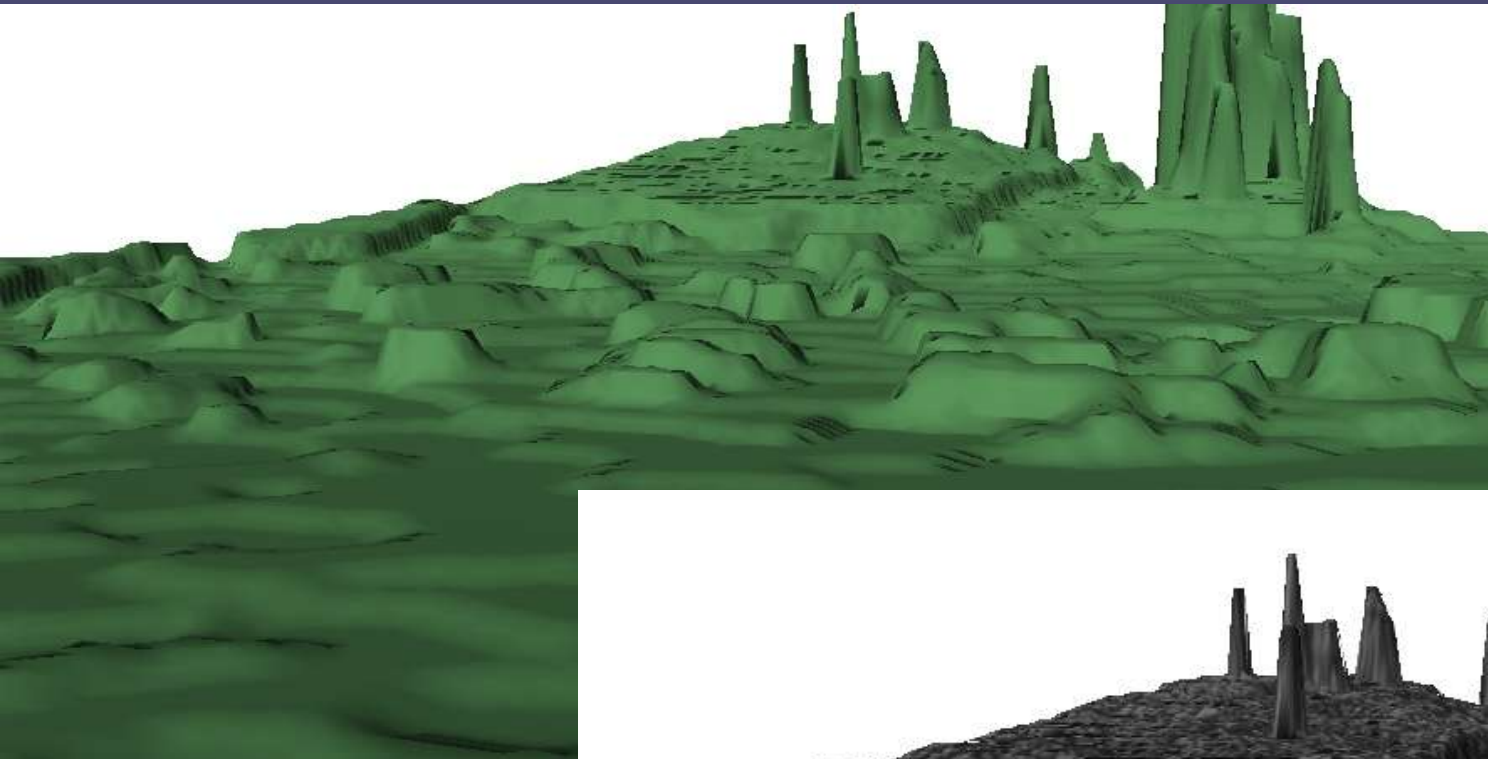
MNE Résultant en 3D



Filtrer le MNE

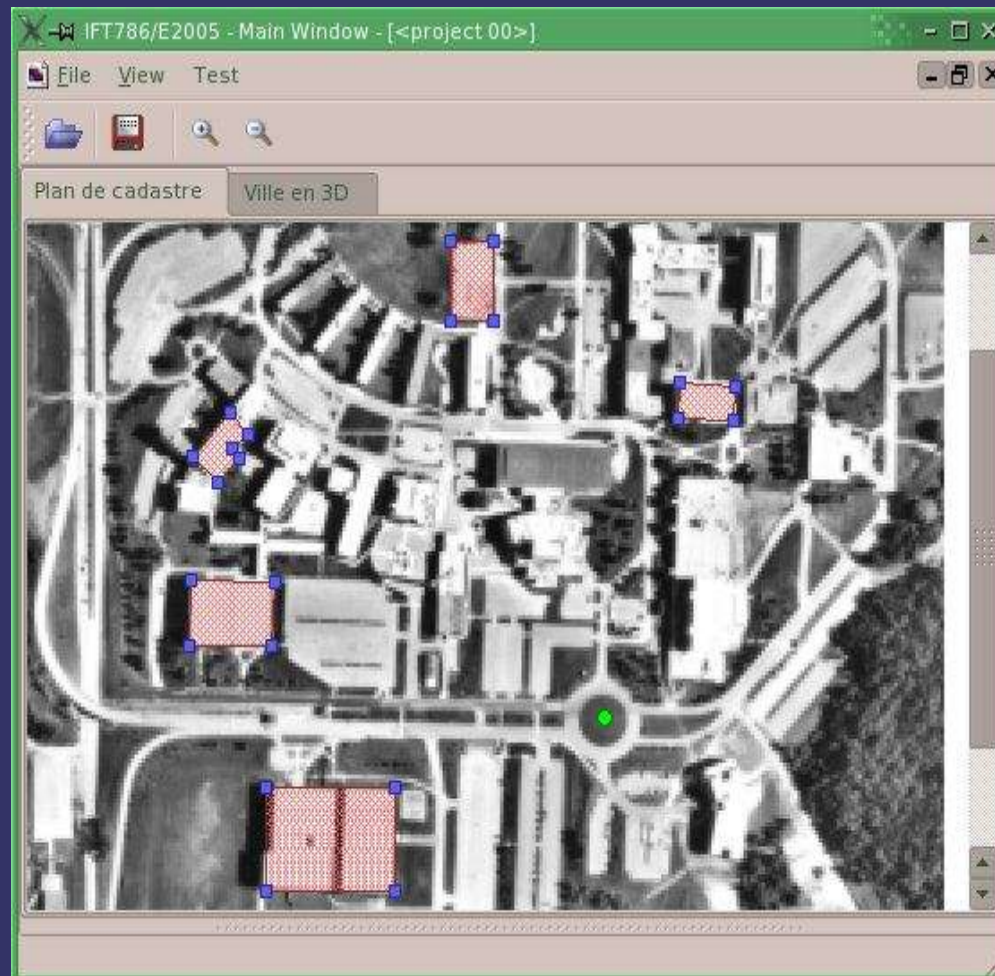
1. Filtre médian sur un grand voisinage pour éliminer les valeurs extrêmes
2. Filtre gaussien pour adoucir les variations
3. Découpage de la zone d'intérêt (UdeS)
4. Réduction d'échelle 25% (256x256)

MNE filtré



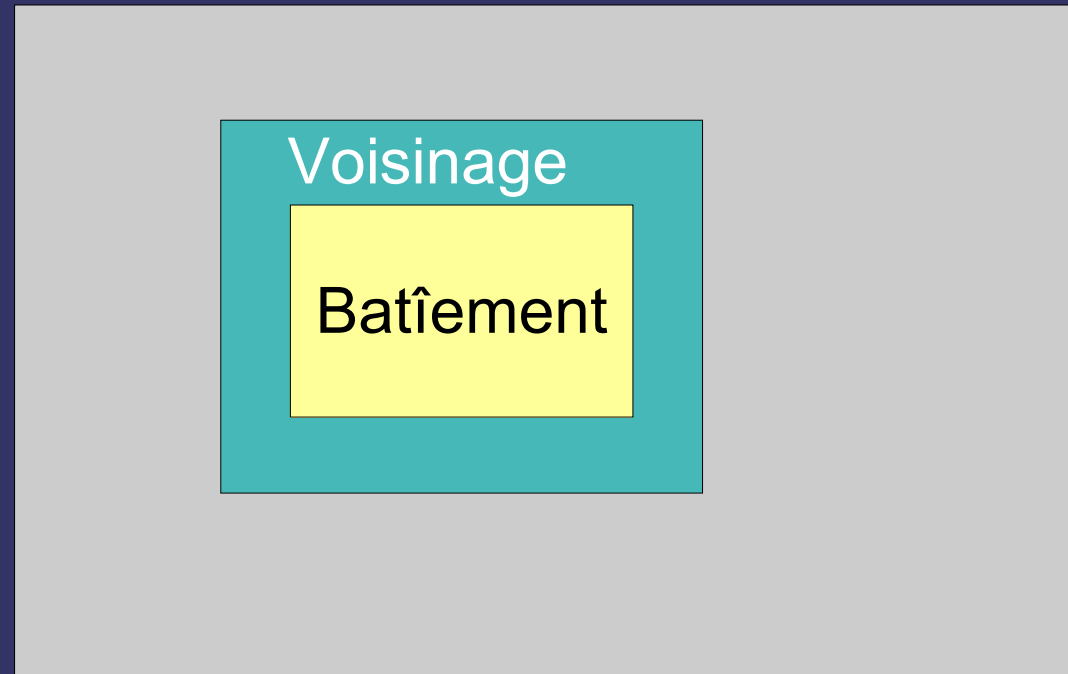
Édition de la carte (plan cadastre)

- ➔ Méthode simple : dessiner des polygones sur une image aérienne



Première passe pour le calcul des hauteurs

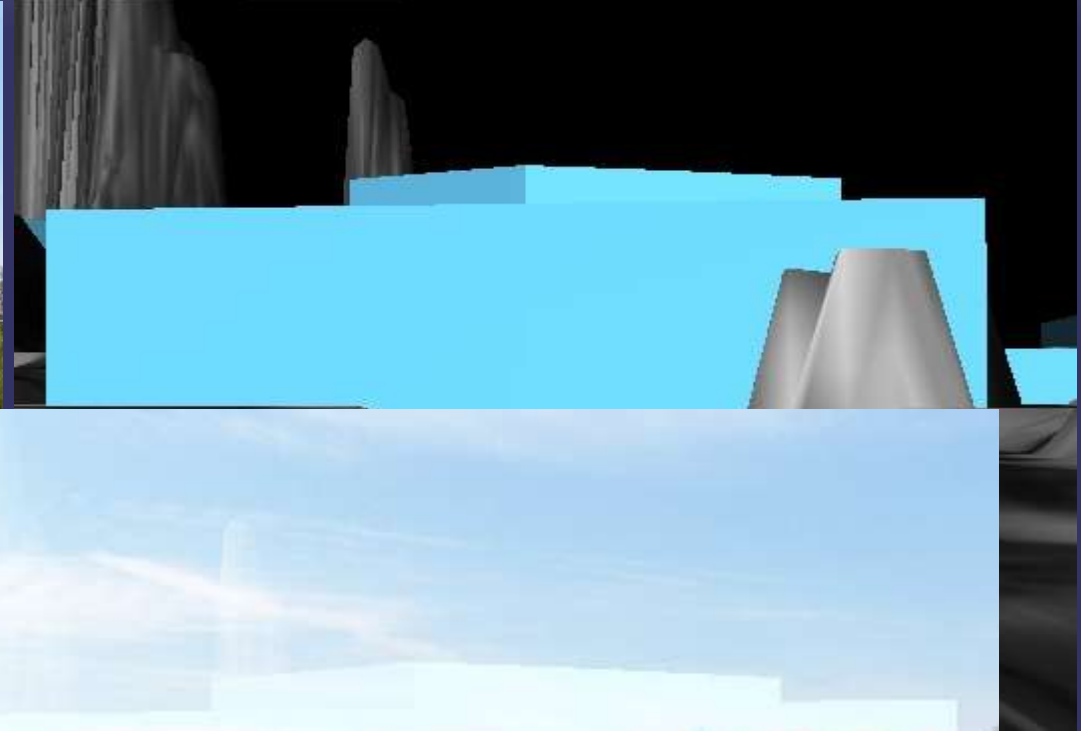
- ➔ Utilisation du MNE pour déterminer l'élévation de la base et la hauteur des bâtiments
 - Élévation = $\text{Min} \{ \text{Altitude}[\text{Voisinage}] \}$
 - Hauteur = $\text{Max} \{ \text{Altitude}[\text{Intérieur}] \} - \text{Élévation}$



Partie manuelle : photos aux sol (1)

- ➔ Prendre des photos au sol
- ➔ Superposer l'image au modèle 3D courant : ajuster les paramètres de la caméra virtuelle aux mêmes que la photo
- ➔ Retrouver la position et l'orientation de la caméra (avec 3 points connus : pas 100% fonctionnel, voir démo.) [4]
- ➔ Éditer les hauteurs en faisant correspondre la photo à la scène virtuelle 3D

Partie manuelle : photos au sol (2)



Partie manuelle : *photos aux sol (3)*

⇒ Problèmes

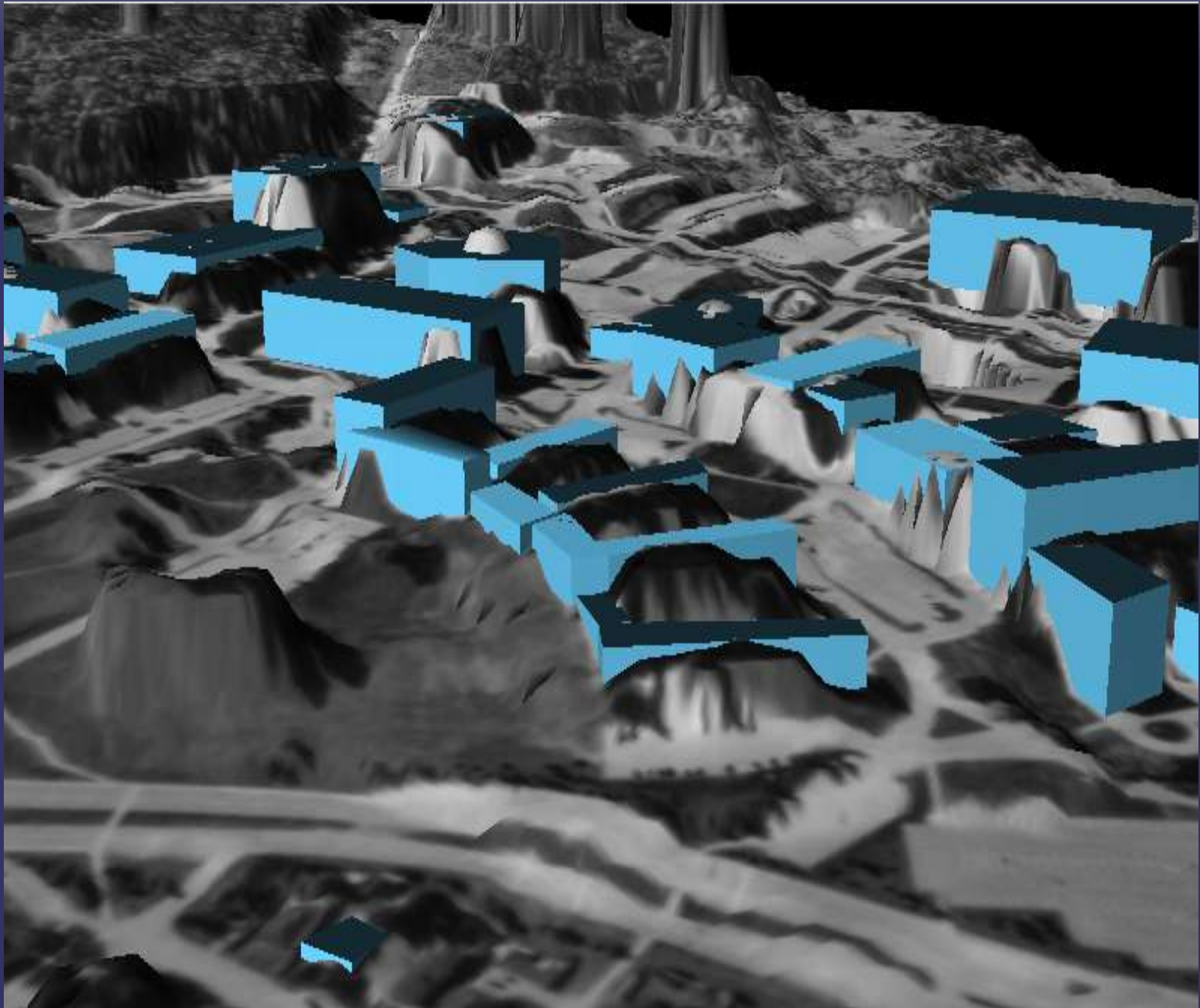
- Mon MNT n'est pas bien filtré (en réalité, c'est encore un MNE)
 - Difficile à voir les points au sol en raison des obstacles
 - MNT bruité : pas très robuste pour fixer des points de repère
 - Algorithme pour retrouver la position de la caméra à l'aide de 3 points n'est pas implanté
- ⇒ Il est tout de même possible de faire quelques ajustements des hauteurs

Partie manuelle : photos aux sol (4)

- Le pavillon central clairement trop haut.
- Le centre culturel est aussi trop élevé par rapport au pavillon central : on ne le voit pas sous l'angle de la photo.



Résultat



Conclusion

- ➔ La détermination automatique des hauteurs par stéréovision est un bon début
- ➔ Fonctionne bien dans les zones à grande surface (le campus est dégagé, faible densité de bâtiment)
- ➔ On est capable de reconnaître l'UdeS en regardant les forme 3D des bâtiments reconstruits (relief et hauteur)

Conclusion (2)

⇒ Points à améliorer

- Améliorer l'implémentation de la stéréovision
 - Utilisation d'une approche « multipass »[8] ou multiéchelles
 - Utilisation d'une meilleure mise en correspondance [8,9]
 - Revenir à une résolution plus fine (10000x10000)
- Mieux filtrer le MNE pour obtenir un bon MNT
- Corriger la géométrie du MNE et de la photos avant de réaliser la carte (enlever distorsions dû au relief)
- Détecter les formes de toitures [2]
- Couleur / texture de bâtiment : on a les photos!

Références

- [1] C. Baillard. 3-D Reconstruction of Urban Scenes from Aerial Stereo Imagery: a Focusing Strategy, 1998
- [2] C. Brenner et N. Haala. Fast Production of virtual reality city models. ...
- [3] R.T. Collins, A.R. Hanson, E.M. Riseman et H. Schultz. Automatic Extraction of Buildings and Terrain from Aerial Images. International workshop on automatic extraction of man-made objects from aerial and space images. April 1997, p169-179.
- [4] C. Chevrier et J.P. Perrin. Interactive 3D reconstruction for urban areas: an image based tool. ??
- [5] O. Faugeras, L. Robert, S. Laveau, G. Csurka, C. Zeller, C. Gauclin et I. Zoghalmi. 3-D Reconstruction of Urban Scenes from Image Sequences. Computer Vision and Image Understanding. Vol 69, no 3, March, p. 292-309, 1998.
- [6] R.C. Gonzalez et R.E. Woods. Digital Image Processing.
- [7] N. Haala et C. Brenner. Extraction of building and trees in urban environments, Journal of Photogrammetry & Remote Sensing 54 (1999) 130-137.
- [8] Y-P. Hung, C-S. Chen, K-C. Hung, Y-S Chen, C-S. Fuh. Multipass hierarchical stereo matching for generation of digital terrain models from aerial images. Machine Vision and Application (1998) 10: 280-191.
- [9] H-Y. Lee, T. Kim, W. Park et H.K. Lee. Extraction of digital elevation models from satellite stereo images through stereo matching based epipolarity and scene geometry. Elsevier, Image and Vision Computing, 21 (2003) 789-796.
- [10] Massalabi, A., D. C. He et G. B. Bénié. Détection et exploitation des zones d'ombre sur les images de très haute résolution spatiale en milieu urbain : application aux données IKONOS de Sherbrooke. AQT et CASI. 25e Symposium canadien sur la télédétection. Montréal, Québec, Canada. 14-17 octobre 2003.
- [11] R. Nevatia, C. Lin et A. Huertas. A System for Building Detection from Aerial Images. 1997
- [12] E. Trucco et A. Verri. Introductory Techniques for 3-D Computer Vision.