

Reconstruction de scènes urbaines à l'aide de fusion de données de type GPS SIG et Vidéo

Gael Sourimant

PhD Thesis Université de Rennes 1 - 2007

Ce thèse de doctorat concerne la reconstruction 3D de scènes urbaines en combinant des données SIG et des images vidéos prises au sol conjointement avec des informations de positionnement GPS.

Gael Sourimant signale l'importance des plates-formes tels que Google Earth actuellement. Il pointe également les lacunes de ces systèmes au niveau des textures et de la précision de modélisation des bâtiments urbains, parfois dénués de fenêtres et de portes (les micro-structures).

1 Reconstruction 3D

Une première partie développe les méthodes de reconstruction 3D et les données utilisées. Définir le type de reconstruction 3D permet de préciser les données nécessaires.

Reconstruction à partir de données aériennes. Un avion ou des satellites permettent d'acquérir des images via le radar par exemple. La zone de couverture de ces méthodes est large et ils ont l'avantage des coûts faibles. En revanche, ils manquent de précision quant aux structures et formes des bâtiments.

La reconstruction nécessite deux étapes : déterminer les zones occupées par les bâtiments et calculer leur hauteur.

Pour *distinguer les bâtiments*, un premier modèle repose sur la segmentation de la zone en plus petites régions et d'identifier une similarité de ces régions à partir de critères d'homogénéité. Des données d'élévation DEM peuvent être utilisées pour la détermination des régions. Par les réseaux de neurones, ils peuvent être classifiés.

Un deuxième modèle projette les empreintes des bâtiments sous forme de SIG dans des photos aériennes obliques. Il faut ensuite une mise en correspondance entre les éléments de deux images d'une même zone pour trouver le contour des bâtiments.

Un dernier modèle consiste à détecter des formes rectangulaires des bâtiments à partir de la projection de lignes extraites d'images radar sur celles d'une image optique classique, la détection de Canny-Deriche est citée comme exemple. Une autre alternative est de considérer les bâtiments comme un ensemble de coins à déterminer. Des segments extraits de l'image radar sont utilisés car les points d'intersections de ces segments servent à déterminer les coins composant les bâtiments.

Pour *calculer la hauteur* des bâtiments, les régions des images segmentées antérieurement permettent de déterminer les polygones des toits. En appliquant la stéréovision, les hauteurs du modèle 3D sont déterminées. Une autre méthode implique la mise en correspondance des données extraites de l'image avec des modèles synthétiques (sous forme CSG : Constructive Solid Geometry) et permet d'avoir des formes de bâtiment complexes.

Reconstruction 3D à partir de données au sol. Les coordonnées, la texture, les formes et éléments des bâtiments sont acquises par des photographies. L'acquisition au sol est précise mais nécessite des moyens considérables.

Il y a deux méthodes de reconstruction : celle reposant uniquement sur les images, ensuite celle qui utilise des modèles pour faciliter la reconstruction.

Une caméra fournit les données sur les images capturées pour l'*approche sans a priori*. La pose, le calibrage et les prises de vue de la caméra sont enregistrés. Un algorithme distingue les façades des fausses facades et détermine les textures des bâtiments. Les structures des villes loin d'être polyédriques sont plus complexes. Tsy modélise la surface des villes par des ondelettes.

Pour la *reconstruction basée modèles*, le logiciel *Façade* qui fait intervenir une personne au processus de placement des primitives, de définition de contraintes du modèle sur la scène. C'est une combinaison de conception assisté par ordinateur et de stéréovision. Un algorithme rassemble les bords du modèle. Les images de la caméra étant projeté sur le modèle choisi par l'utilisateur.

Une autre méthode plus incrémentale déduit un nuage de point de la scène à partir de plusieurs images. Ces points déterminent les plans principaux du bâtiment. Les portes et fenêtres sont ensuite détectées comme des régions d'intérêt par l'observation de la variation de gradient par rapport à la façade. Ces nuages de points permettent aussi de raffiner les zones de micro-structure en donnant des informations sur leur profondeur.

Il est également possible d'établir les plans principaux des bâtiments par l'estimation des points de fuite. Le raffinage du modèle est le fruit le résultat d'un appariement avec des modèles 3D prédéfinis de portes et fenêtres.

La méthode de Torr et al. représente la scène par des couches, détectées en estimant les nuages de points de la scène en 3D. Le plan principale de la scène en est une, les micro-structures sont des couches superposées. Une couche est constituée par les propriétés de positionnement, la taille, la texture, etc. Elle est associée à un modèle prédéfinies (arche, fenêtre rectangulaire) et regroupée par des hyperparamètres (propriété d'alignement par exemple).

Reconstruction à partir de données hybrides.

Pour profiter des forces des deux approches de reconstruction et avoir des modèles 3D précis à grande échelle. Des méthodes hybrides combinent à la fois les données aériennes et les données au sol. Un premier modèle se base sur les images aériennes pour en dégager les empreintes au sol des bâtiments, avec une intervention manuelle. Des données Lidar permettent ensuite de reconstruire les bâtiments. Des images au sol servent ensuite à définir les textures. La fusion d'images aériennes avec les données au sol et Lidar ne donnent pas de géométries précises pour les façades.

Une autre méthode modélise les scènes urbaines à partir de données acquises au sol par deux lasers et une caméra. Les données brutes sont acquises et une estimation de pose de la caméra est effectuée, les textures des façades sont plaquées automatiquement.

2 Données utilisées

Sourimant introduit son idée de combiner des données aériennes fournies par l'Institut Géographique National (IGN) et des données acquises au sol avec des caméras. Pour cela, il utilise les mesures GPS conjointement à l'acquisition des images au sol. Il présente ces trois types de données en détails pour en dégager les spécificités et limitations.

Un **Sig** (Système d'Information Géographique) est un système donnant une représentation d'un environnement géographique à partir de primitives simples. Les informations fournies par le SIG dépendent de la nature (hydrographie, routes, bâtiments) de l'environnement et sont regroupées par couches. Evidemment, la couche qui nous intéresse est celle des bâtiments. Elle est composée par une liste de bâtiments sous forme de points et représentée par le triplet [empreinte au sol, altitude, hauteur]. Dans le repère spécifique UTM (Universal Transverse Mercator), le bâtiment est décrit par le triplet

[liste de points 2D dans le plan $(X, Y)_{utm}$, altitude a_{utm} , hauteur h_{utm}]. La représentation GIS a des limites : aucun détail géométrique pour les micro-structures, une représentation strictement plane, pas d'information sur la texture. L'esthétique des bâtiments n'est pas modélisée. En conséquence, il est nécessaire d'introduire les images de la vidéo. Elles servent à extraire les textures et les micro-structures par recalage, ces images étant projetées sur le modèle choisi.

Vidéos Les images vidéos sont acquises au sol grâce à une caméra numérique. Sourimant suppose le pire des cas où la caméra n'est pas calibrée et que seules les propriétés données par le constructeur de l'appareil sont disponibles, et ainsi rendre robuste la reconstruction.

GPS Pour mettre en relation les données SIG et les images des vidéos, une information de GPS est ajoutée aux images au sol qui sont acquises conjointement. Le GPS est un système de navigation américain par satellite constitué de 24 satellites gravitant autour de la planète sur 6 orbites circulaires. Les informations requises pendant leur passage permettent de repérer tout point du globe. Cinq stations au sol inséminées sur la terre sont chargées de le piloter. Ces satellites émettent sur deux fréquences pour minimiser les erreurs de localisation dues aux interférences électriques sur le signal envoyé depuis les satellites lors de la traversée des couches protectrices terrestres.

L'auteur détaille ensuite la manière dont le système GPS opère pour repérer un point sur la planète. Notamment, il note l'intersection des sphères satellitaires qui doivent être au nombre de quatre pour fournir des données d'une grande précision en 3D. Il indique que les coordonnées doivent être polaire (latitude, longitude, altitude) en rappelant les règles de transformation entre les coordonnées. Les sources d'erreur du GPS sont les perturbations atmosphériques, les trajets multiples occasionnés par les environnements urbains par réflexion des signaux et la configuration des satellites (évaluée par la DOP : Dilution of Precision).

Pour corriger ces erreurs, les USA introduisent les GPS différentiel qui fait appel à des stations au sol pour réduire les erreurs en effectuant des mesures à partir de points éloignés les uns des autres. L'autre système proposé EGNOS (European Geostationary Navigation Overlay System) joue le même rôle que le GPS différentiel à la différence que EGNOS utilise trois satellites géo-stationnaires pour relayer les informations de corrections au lieu de stations au sol. Les récepteurs GPS récents sont munis d'un canal supplémentaire pour les corrections transmises par EGNOS, en plus du canal captant le GPS normal.