

Extraction de contenu photographique en utilisant les systèmes d'information géographique(SIG)

Youssef Attia¹, Thierry Joliveau¹, Eric Favier² et Benoit Delaup

1 : CRENAM, Université Jean Monnet, 6 rue Basse des Rives 42023 Saint Etienne- France.

2 : ENISE, 58, rue Jean PAROT - 42023 ST-ETIENNE CEDEX 2.

Contact : youssef.attia@univ-st-etienne.fr

Résumé

L'annotation des photos avec des informations qualitative est une tâche longue et ennuyeuse. Une des manières optimales d'organiser des photos est de les annoter en précisant le contenu et la position des objets dans la photo. Ces objets sont les éléments qui sont observables à partir de la position de la caméra. Notre projet consiste à évaluer les possibilités de relier le plus automatiquement possible des bases de données géographiques et des photographies. Ce travail vise à mettre en œuvre un concept qui effectue l'extraction automatique de contenus de photos urbaines puis lie automatiquement le contenu de l'image avec une base de données géographiques.

Abstract

Photo annotation with qualitative captions is a timeconsuming and annoying task. A supreme way to organize photos is to have annotation of their content and the map position of objects in the photo. These objects are the elements that are observable from the camera point of view. Our project consists in evaluating the possibilities of linking geographic and photographic databases. This work implement a concept that aims to an automatic extraction of urban photo contents then to an automatic linkage of the picture itself with its content and geographic databases.

Mots-clés : SIG, photographie, reconstitution 3D, cartes

Keywords: GIS, photography, 3D environment reconstitution, maps

1. Introduction

Les appareils ayant une puce GPS intégré se répandent de plus en plus ce qui permet d'avoir un grand nombre de photos avec une localisation associée. De plus l'intérêt de positionner géographiquement les photos se développe rapidement. Pour ce faire, la localisation de photos en ligne est présente à titre gratuit (Google Maps, Microsoft Virtual Earth ou Flickr) et montre que le monde cherche à géolocaliser ses photos.

La géolocalisation des photographies trouve un intérêt direct dans trois domaines. A un niveau général et dans un contexte d'investigation plutôt scientifique, c'est un moyen de décrire l'environnement visible et d'organiser des vues photographiques en fonction de l'espace qu'elles décrivent. Cela facilite les visualisation hybrides, cartographique et photographique, d'un espace donné et la construction d'itinéraires liant les photos entre elles par les fragments d'espace qu'elles ont en commun. Dans un usage professionnel, l'exploitation des vastes collections de photographies existantes dans les musées, les collections ou les photothèques de multiples services techniques nécessitent la géolocalisation des vues anciennes, qui sont en voie de numérisation. Mais ces tâches de localisation sont fastidieuses et nécessitent de nouvelles méthodes d'appui pour les accélérer. Enfin, la multiplication des collections de photographie en ligne et l'apparition d'appareils permettant une localisation automatique des photographies ouvre de nouveaux usages personnels des individus souhaitant garder la trace des espaces visités. Des enjeux très importants apparaissent liés au classement de ces collections de photographie par l'intermédiaire de production automatique de géotags.

Le présent document est organisé comme suit : le point 2 décrit les dispositifs et les techniques décrites dans la littérature servant à ajouter des métadonnées spatiales aux photos. Le point 3 propose un algorithme pour l'extraction de contenu photo. Le point 4 présente l'outil créé dans le but de valider le concept. Enfin, le point 5 conclut et présente les futurs travaux.

2. Travaux antérieurs

James Hays et Alexei A. EFROS [6] ont visé à évaluer l'information géographique à partir d'une seule image. Ils ont construit une base de données d'images géolocalisées et ont réalisé une comparaison entre la photo en question avec tout les éléments de cette base et en utilisant une combinaison de plusieurs descripteurs. Ce travail utilise la technique de l'association. Elle considère que l'on peut mettre la photo dans une collection d'images ayant chacune une position géographique connue et qu'elle est prise non loin des photos de cette collection. Les résultats ne sont acceptables que dans le cas des photos touristiques courantes. P. Ghadirian et I.D. Bishop [5] ont intégré de la réalité augmentée dans le paysage afin de donner une reconstitution en 3D des images. Ils présentent une vidéo d'une vue panoramique qui donne la position de la caméra sur une carte.

Le projet Snap2Play [2] essaye de diriger des joueurs à l'aide d'indications en réalité augmentée affichées sur un téléphone portable. La connaissance préalable du terrain ainsi que les capteurs reliés au téléphone permettent de guider l'utilisateur à retrouver des points préalablement géolocalisés.

Certains dispositifs offrent un certain confort de géolocalisation comme les Tablet PC qui peuvent être connecté a une antenne GPS et un logiciel dédié comme est le cas de MAP IT [4]. Ce type de dispositifs permet de cartographier, de décrire et d'attribuer de nombreux types d'annotations aux photos.

2.1. Les recherches de Boer

AD Boer [1] a proposé une solution pour ajouter des labels sur des photos géoréférencées. Les données utilisées au cours de ce travail sont des données vectorielles 2D, l'ajout de la composante

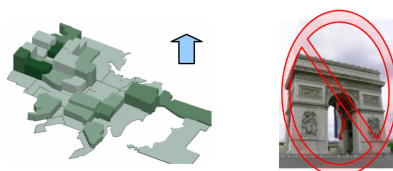


FIGURE 1 – Extrusion d'une carte 2D.

hauteur rend les données de la forme 2,5D, une simple augmentation de l'image 2D nous permet d'avoir des formes 3D. Ce modèle diffère d'un modèle 3D puisque c'est le résultat d'une élévation de la carte, de sorte que nous ne pouvons pas définir un bâtiment comme l'Arc de Triomphe (voir



FIGURE 2 – Modèle 3D, identification d'objets et ajout de labels.

figure 1). Cependant c'est utilisable est suffisant pour nos objectifs. Son travail s'est basé sur les possibilités offertes par Viewshed sur ArcView, le logiciel SIG de ESRI. La figure 2 présente à gauche la reconstitution 3D de l'environnement traité, au centre, la deuxième étape qui est la reconnaissance des bâtiment en comparant le rendu donné à partir des mêmes paramètres de la prise de photo, finalement à droite une illustration de l'ajout de labels des bâtiments qui ont de l'importance sur la photo.

2.2. Le calcul d'Isovist

Rana et Joliveau ont créé en 2007 un outil permettant de calculer une intersection entre l'espace visible à partir d'un point donné et les bâtiments susceptibles d'être pris en photo au moyen d'un calcul d'isovist. En effet, la notion d'isovist a été introduite par Benedikt [3] comme un outil de modélisation de la perception spatiale de l'homme dans le contexte de la conception architecturale et, plus tard, pour la perception de l'espace. Rana, qui travaille sur les isovist et Joliveau qui cherche à définir l'espace dans la photographie ont travaillé ensemble [7] et ont présenté une solution reposant sur le technique de lancer de rayons. Pour le calcul de la zone ouverte par la photographie et pour trouver les objets correspondants, il est possible d'envoyer des rayons panoramiques autour du point de vue et de déterminer ensuite les objets qui croisent ces rayons. Le calcul d'isovist envoie des rayons sur 360°, cependant on peut se limiter à l'angle couvert par la caméra. La procédure pour identifier les caractéristiques urbaines visible dans une photographie se fait en deux étapes. La première consiste à générer l'isovist relatif à la position de la caméra.

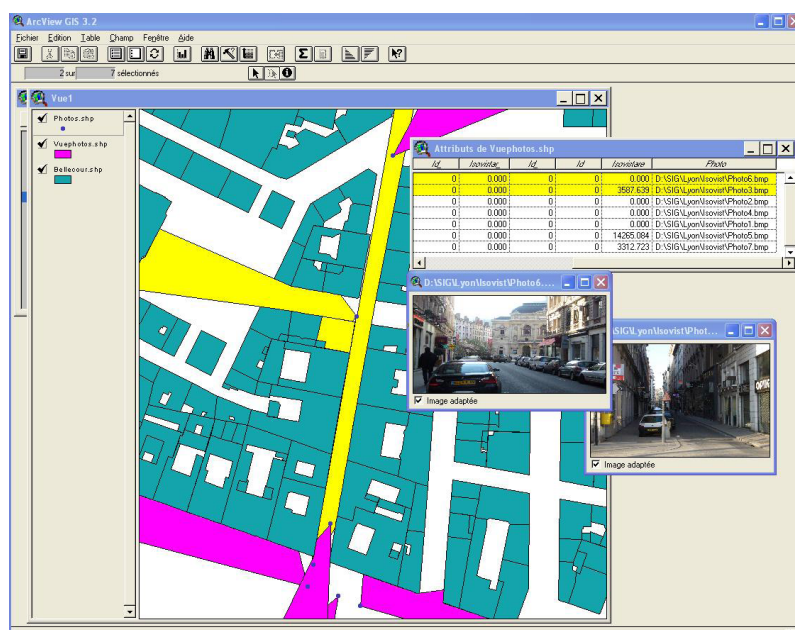


FIGURE 3 – Isovisits et requêtes spatiales.

La deuxième étape passe par l'exécution de requêtes spatiales pour identifier les objets présents. Plusieurs requêtes sont possibles, comme la recherche de toutes les photos dans la base de données qui sont censés représenter la façade d'un bâtiment particulier. La figure 3 montre les isovists de différents points de vue sur une partie du centre de Lyon. Une liste des bâtiments ou de n'importe quel objet présent dans la base de données présents sur les 2 photos est affichée sous forme de table. Cela permet donc de tagger les photographies grâce aux objets de la base de données.

3. Une nouvelle approche d'extraction d'information

Nous proposons ici une nouvelle méthode qui consiste à combiner une photographie dont on connaît les coordonnées et certains autres paramètres avec les données d'une base de données géographiques pour extraire des informations sur le contenu de la photographie.

L'extraction du contenu des photos se fait en se référant à la position géographique de la prise de vue, la focale (distance qui sépare le foyer de la lentille de son plan principal) et l'azimut (angle horizontal entre la direction d'un objet et le nord géographique). Ces informations sont fournies par la caméra et les équipements connexes. L'idée est de faire une projection prospective dans notre monde 3D pour avoir un même contenu visuel que celui d'une photographie prise dans le monde réel. Pour ce faire, trois étapes sont à suivre. Tout d'abord, nous décrivons nos données : nous utilisons des cartes 2,5D, cela signifie que nous avons une carte 2D et pour chaque point, nous avons une altitude.

Après une préparation des données, nous explorons l'image pour y déterminer les bâtiments et les routes. Nous supposons que les bâtiments sont des polyèdres, mais les espaces ouverts causent un problème lors de l'identification. Les espaces de parcs, espaces libres et les routes ont la propriété d'être vides mais ils sont caractérisés par des limites physiques, tels que les bâtiments. Pour ce faire, nous définissons les places et les rues par les murs des bâtiments qui l'entourent. Cependant, il faut noter que le mur d'un bâtiment appartient à une rue sans que tous les murs de l'immeuble appartiennent à la même rue. L'indexation des routes et des bâtiments est illustrée par la figure 4.

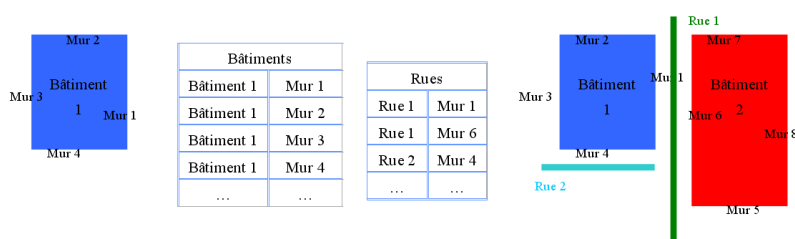


FIGURE 4 – Indexation des bâtiments et des rues.

Les bâtiments sont représentés sous forme de polygones dans la base de données et découpés en segments élémentaires. Ces segments, après avoir été extrudés, représentent les murs de l'environnement. La troisième étape est l'extraction de contenu de l'image. Une fois que la base de données contient nos données, et que l'environnement 3D est construit, nous créons une projection perspective à partir de 3 coordonnées cartésiennes (X, Y et Z), un azimut et la focale de la photo prise en réalité par la caméra. Dans le rendu obtenu, nous codifions chaque identifiant (ID) de bâtiment directement dans sa couleur, et nous ajoutons la distance qui le sépare du point de prise de vue dans la couleur. Un bâtiment très proche sera de couleur plus foncée tandis qu'un bâtiment lointain sera plus clair. De cette façon nous avons la possibilité de récupérer l'identité et la distance de chaque pixel trouvé sur la photo de synthèse produite.

4. Réalisation

L'implémentation est faite en C++ comme langage, en utilisant la bibliothèque OpenGL, pour les traitements 2D et 3D et PostgreSQL pour stocker les données, l'application assure la liste des fonctionnalités suivante :

- Détermination des bâtiments et des espaces sur une photo en connaissant les paramètres géographiques (X, Y, Z, azimut) et la longueur focale de la photo.
- Génération de données géographiques (X, Y, Z, et de l'azimut) et photographiques (distance focale) pour des photos n'ayant pas ces paramètres.
- Gestion du contenu du format de fichier image échangeable (EXIF).

– Estimation du coefficient de présence de chaque objet identifié sur la photo.

Les erreurs de mesures font qu'un recalage d'image est parfois nécessaire. Cela est dû aux incertitudes des coordonnées GPS (jusqu'à 8m) et des autres appareils de mesure en particulier la boussole qui peut aller jusqu'à 14° d'erreur. La figure 5 présente un exemple concret où l'on a en haut à gauche l'image prise par l'appareil, juste en dessous sa position, son angle de vision et sa focale. À droite en haut on trouve le rendu généré, qu'il a fallu réajuster manuellement pour le rendre encore plus proche de la photo initiale. Le rendu recalé est visible au coin bas à droite de la figure.

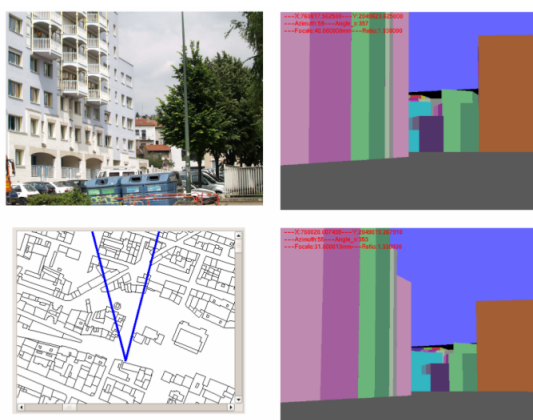


FIGURE 5 – Résultat d'extraction de contenu d'une photo.

5. Conclusion

Il est possible de conclure que cette proposition de la géolocalisation des photos avec une reconstitution en 3D de l'environnement a fait la preuve de son intérêt. Le fait d'avoir 2 photos identiques, une prise avec un appareil photo et une vue générée par rendu, nous permet d'extraire un grand nombre d'informations. On a accès à la liste des bâtiments visibles et à la distance à chaque objet par rapport à l'objectif. Des calculs sont possibles, par exemple le pourcentage de présence qui correspond à la surface de l'objet sur la photo divisé par la distance de ce dernier par rapport au point de prise de vue. Ceci est valable pour tout objet qu'il s'agisse d'une rue, d'un bâtiment, d'une place ou autre.

Nous proposons deux techniques pour accroître la précision des données GPS. La première consiste à mettre en œuvre un algorithme de vision, qui permet de détecter les contours des objets. On fait ensuite la correspondance entre les formes dans la photo et dans le rendu 3D ce qui permet un recalage fidèle à la réalité. Une deuxième technique consiste à déterminer la matrice de projection. Cette dernière permet de retrouver la position exacte du point de prise de vue à partir des éléments pris en photographie. La perspective de projection peut être formalisée à l'aide d'outils mathématiques de la géométrie projective. R. Hartley [6], dans un livre consacré à l'étude de la vision à base de caméra, indique qu'il est possible de déterminer la matrice de projection en ayant la projection (c'est-à-dire la photographie) et la position spatiale de certains éléments.

Bibliographie

1. Arnoud De Boer, Eduardo Dias, et Edward Verbree. *Processing 3D Geo-Information for Augmenting Georeferenced and Oriented Photographs with Text Labels*, pages 351–365. 2008.

2. Tat-Jun Chin, Yilun You, Celine Coutrix, Joo-Hwee Lim, Jean-Pierre Chevallet, et Laurence Nigay. Mobile phone-based mixed reality : the Snap2Play game. *The Visual Computer*, 25(1) :25–37, 2009.
3. Larry S. Davis et Michael L. Benedikt. Computational models of space : Isovists and isovist fields. *Computer Graphics and Image Processing*, 11(1) :49–72, septembre 1979.
4. Mauro De Donatis et Lorenzo Bruciatelli. MAP IT : the GIS software for field mapping with tablet pc. *Computers & Geosciences*, 32(5) :673–680, 2006.
5. Payam Ghadirian et Ian D. Bishop. Integration of augmented reality and GIS : a new approach to realistic landscape visualisation. *Landscape and Urban Planning*, 86(3-4) :226–232, 2008.
6. J. Hays et A.A. Efros. IM2GPS : estimating geographic information from a single image. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, 2008.
7. Thierry Joliveau et Rana Sanjay. Using urban viewsheds for embedding geographical context in photograph databases of urban areas. Maynooth, Ireland, 2008.

Ce travail est soutenu par la Région Rhône-Alpes.
