Université Jean Monnet
Saint-Étienne

Ecole Doctorale Sciences Sociales

EVS-ISTHME
composante Université
Jean Monnet Saint-Étienne
EVS
Environnement
Ville Société
cnrs
UMR 5600
Université de Lyon

Doctorat en Géographie : Information géographique et applications

Présenté et soutenu le 31 août 2012 par

Youssef ATTIA

Interfaçage de bases de données photographiques et géographiques par appariement de lignes

Thèse dirigée par :

- Thierry JOLIVEAU — Professeur, Université Jean Monnet
- Eric FAVIER — Maître de conférences, ENISE

Jury :

- Alain BRETTO — Professeur, Université de Caen, rapporteur
- Guillaume MOREAU — Professeur, Ecole Centrale de Nantes, rapporteur
- Bruno TELLEZ — Maître de conférences, université Lyon 1, examinateur

EVS - ISTHME

Doctorat en géographie : Information géographique et applications

Présenté et soutenu le 31 août 2012 par

Youssef ATTIA

Interfaçage de bases de données photographiques et géographiques par appariement de lignes

Thèse dirigée par :

- Thierry JOLIVEAU — Professeur, Université Jean Monnet
- Eric FAVIER — Maître de conférences, Ecole Nationale d'Ingénieurs de Saint-Etienne

Jury :

- Alain BRETTO — Professeur, Université de Caen, rapporteur
- Guillaume MOREAU — Professeur, Ecole Centrale de Nantes, rapporteur
- Bruno TELLEZ — Maître de conférences, Université Lyon 1, examinateur

Cette thèse a été soutenue par une allocation doctorale de recherche de la Région Rhône-Alpes.

Remerciements

À l'issue de ce travail gourmand en temps et efforts, je souhaite adresser mes remerciements à ceux qui méritent beaucoup plus qu'une simple reconnaissance.

Je commence à adresser mes remerciements à tous les membres de mon centre de recherche ISTHME anciennement CRENAM.

Mes collègues au bureau des doctorants ne peuvent être qu'au début de mes remerciements, je m'adresse à Yohann, Céline, Sandra, Wassim, Farid, Olivier et Mohamed. Ils ont été à mes cotés dans les différentes phases de mes trois années de thèse.

Toute ma gratitude à Thierry qui pu diriger avec tant de patience et de courtoisie mes travaux. Merci Eric pour tes conseils et encouragements. J'espère que tu retrouvas rapidement la forme et toute ton énergie.

J'aimerais souligner aussi le rôle de mes deux co-directeurs de thèse qui m'ont appris tout au long de ce travail de recherche à être autonome et persévérant.

Mes remerciements vont aussi aux autres membres de ISTHME et notamment Carole, Catherine, POM et Renaud qui ont fait en sorte que le déroulement de mon travail soit le plus facile et agréable que possible.

Je tient à remercier en particulier POM et Renaud pour leurs assistance, amabilité et bienveillance.

Je remercie mes amis et amies qui m'ont soutenu moralement dans les phases critiques du travail et pendant la rédaction de ce manuscrit.

Mon frère mérite aussi d'être remercié pour ses conseils et aide précieuse pendant les moments forts de ma thèse.

Ma petite famille a contribué avec amour et dévouement au bon déroulement de mon doctorat, milles bises à ma chérie et à mon bébé, mon petit bout de chou qui remplit ma vie de bonheur depuis quelques mois.

Merci Maman, pour ton soutien moral, quand tu es loin et surtout lors de ton court séjour chez moi à l'occasion de l'arrivée au monde de mon petit Ahmed.

Je tiens aussi à souligner le plaisir que j'ai eu à travailler au sein de ISTHME, et j'en remercie encore une fois tous les membres sans exception.

Pour clôturer, il me tient à cœur de dédier cette thèse à Papa et Maman en expression de ma profonde reconnaissance.

Résumé

Interfacer les photographies et les bases de données géographiques est une démarche utile pour les architectes, les historiens, les géographes mais aussi pour le grand public. Affecter une position géographique à une photographie permet en effet de décrire son contenu potentiel en s'appuyant sur les informations contenues dans une base de données géographique.

L'objectif de la thèse est de proposer une méthode permettant de localiser automatiquement une photographie urbaine en la comparant avec des images de synthèse en 3D générées de manière systématique à partir d'un modèle virtuel des bâtiments. Le principe est que la photographie partage avec les images des caractéristiques géométriques qui permettent de les rapprocher et donc de déduire la localisation de la photographie à partir de la position connue de l'image.

La méthode utilisée pour retrouver les images de synthèse correspondantes est un appariement entre les lignes présentes dans la photographie et les lignes détectées dans les images de synthèse par la transformation de Hough. Cet appariement est suivi par une analyse statistique permettant de proposer une localisation probable avec une valeur d'approximation associée.

Malgré les obstacles présents dans les photographies, cette approche utilisant uniquement les lignes est une solution simple et potentiellement efficace pour le positionnement de photographies. Plusieurs scénarios d'usage sont proposés : la géolocalisation d'une image, la validation du positionnement d'une image localisée et l'utilisation de photographies pour repérer des changements dans le paysage urbain.

Mots clés : SIG, photo, appariement de lignes, lignes de Hough, reconstitution 3D, étiquettes, ville

Title : Interfacing a photographic database and a geographical database by line matching

Abstract :

Interfacing photographs with geographic databases is a useful approach not only for architects, historians, geographers but also for a general audience. By assigning a geographical position to a picture it becomes possible to describe its potential content based on the information contained in the geographic database.

The aim of the thesis is to propose a method to locate automatically an urban photograph by comparing this photograph with 3D images generated systematically from a virtual model of the buildings. The principle is to put in relation a photograph with the synthetic images that are sharing analog geometric characteristics and therefore to deduce the location of the photograph from the known position of the image.

In order to find synthetic images related to a photograph the method used is a matching between detected lines in the photograph and the images based on Hough transform. This matching is followed by a statistical analysis to propose a probable location with an estimation of accuracy.

Despite the obstacles present in the photographs, this approach using only lines is simple and potentially effective for positioning photographs in urban space. Several scenarios of use are proposed : first location of an image, validation of the position of an image broadly located and use of photographs to identify change in the urban landscape.

Keywords : GIS, photo, matching lines, Hough lines, 3D reconstitution, tag, city

Unité de recherche : *ISTHME* (Image Société Territoire Homme Mémoire Environnement)
EVS - ISTHME UMR 5600 CNRS
Adresse : 6, Rue Basse des Rives 42023 Saint-Etienne cedex 02

Sommaire

Introduction **11**

Cadre de la thèse 11

Logique de l'exposé 12

1 Géomatique, infographie et fusion de données **17**

1.1 Notions de géomatique 18

1.2 Notions en traitement d'image 31

1.3 Processus de la fusion de données 39

1.4 Conclusion 43

2 Couplage de BD photo et géographiques **47**

2.1 Méthodes de couplage dans la littérature 49

2.2 Méthodes utilisées dans le projet MapYourPicture 55

2.3 Conclusion 62

3 Localisation de photographies par appariement de lignes **67**

3.1 Présentation globale 67

3.2 La démarche d'interfaçage par appariement de lignes 69

3.3 Choix technologiques et méthodologies 77

3.4 Traitement d'image sur les photographies 86

3.5 Conclusion 89

4 Validation du concept **93**

4.1 Méthode d'analyse des données 93

4.2 Données utilisées 95

4.3 Résultats 104

4.4 Limites de la méthodologie 115

4.5 Conclusion 118

5 Scénarios d'utilisation fondés sur la distance **121**
1er scénario 121
2ème scénario 124
3ème scénario 131
Conclusion 132

Conclusion **135**
Bilan 135
Limites de la méthode et perspectives d'amélioration 136

Annexes **141**

A Base de données photographiques **141**

B Appareils utilisés pour la prise des photographies **143**

C Description des données utilisées sous AcrScene **147**

D Tableau des faux positifs **149**

Glossaire et acronymes **151**

Table des matières **153**

Table des figures **157**

Liste des tableaux **161**

Bibliographie **165**

Introduction

Introduction

On rentre maintenant de vacances avec en plus d'innombrables souvenirs des dizaines voire des centaines de photographies, que l'on stocke dans des bases de données personnelles de plus en plus volumineuses. Dans le monde professionnel, les architectes, les historiens, les journalistes et des centres de documentation spécialisés collectent et complètent en permanence de gigantesques bases de données photographiques. Qu'elles soient personnelles ou professionnelles ces bases de données photographiques posent des problèmes d'organisation, de stockage et d'interrogation de plus en plus complexes et stratégiques.

La localisation géographique est devenue une clé centrale pour organiser et retrouver des photographies. Une innovation technologique telle que l'intégration d'appareils photographiques dans les téléphones équipés de systèmes de positionnement et de GPS dans les appareils photographiques et les caméras est une illustration de l'importance grandissante de cette dimension géographique dans l'archivage et l'interrogation des bases de données photographiques, qui est devenue récemment une question importante de recherche et développement. Les plateformes web qui sont de plus en plus systématiquement utilisées pour stocker et partager des photographies offrent elles aussi systématiquement des fonctions de géolocalisation, soit à partir des informations stockées automatiquement dans l'image soit par une localisation manuelle sur différents types de cartes.

Cadre de la thèse

Dans le but d'améliorer et d'automatiser ce processus de description de la dimension géographique des bases de données cartographiques, Thierry Joliveau (UMR EVS - ISTHME et Université Jean Monnet)[1] et Eric Favier (laboratoire DIPI - Diagnostic et Imagerie des Procédés Industriels et École Nationale d'Ingénieurs de Saint-Étienne) ont soumis à l'appel d'offres "Cible" de la région Rhône-Alpes un projet intitulé "MapYourPictures!" dont l'objectif était de proposer des méthodes

1. ISTHME est le résultat de la fusion depuis le 1er janvier 2011 du CRENAM (Centre de REcherche sur l'ENvironnement et l'AMénagement) et de l'IERP (Institut des Études Régionales et des Patrimoines). La structure initiale qui m'a accueilli est le CRENAM.

automatiques pour géorenseigner et géolocaliser et des photographies prises en milieux urbains. C'est ce projet qui a financé pour 3 ans ce travail de thèse, commencé au mois de novembre 2008. Ce projet impliquait donc principalement Thierry Joliveau pour les questions de géomatique et Eric Favier sur le volet informatique et image. Sanjay Rana, chercheur de l'Université de Londres spécialisé dans les calculs de visibilité s'est joint au projet dans sa première année.

La réalisation des objectifs du projet passait par une nécessaire intégration de trois domaines scientifiques : la géomatique et les systèmes d'information géographique d'une part, l'imagerie et l'infographie d'autre part, qu'il s'agissait d'associer au moyen de méthodes de fusion de données. Le projet demandait de maîtriser les méthodes et les outils de gestion des données géographiques, la manipulation d'appareils de mesures de positionnement, divers types d'appareils photographiques et plusieurs outils informatiques.

Logique de l'exposé

Ce manuscrit est organisé en cinq chapitres. Dans le premier, nous dressons un panorama des trois domaines scientifiques que nous venons de mentionner de manière à permettre aux spécialistes de chacun de ces domaines de s'imprégner des autres et de présenter les méthodes et outils utilisés dans ce travail. Nous présentons quelques notions de géomatique et les différentes technologies de positionnement existantes, puis les notions de base en traitement d'image avec comme exemples quelques opérations simples d'analyse d'image, et nous terminons par un exposé rapide sur différentes méthodes de fusion de données en relation avec notre sujet d'étude.

Le deuxième chapitre expose les travaux de la littérature dans le domaine du couplage des bases de données géographiques et photographiques. Il se focalise sur les travaux engagés dans ce sens par notre équipe de recherche. La plateforme créée pour démontrer l'intérêt des travaux déjà réalisés est décrite en détails ensuite.

Le troisième chapitre rapporte la méthode que nous proposons d'appliquer : un interfaçage par appariement de lignes. Elle permet de coupler deux images en utilisant un nouveau concept d'appariement. La méthode est retracée en détail avec la présentation de chaque processus et donnée utilisés. Nous concluons ce chapitre par un descriptif de l'algorithme que nous avons conçu pour procéder à cet appariement.

La validation du concept est l'objet du quatrième chapitre qui présente les données utilisées, les outils manipulés, les expérimentations faites et les interprétations des résultats obtenus.

Le dernier chapitre décrit les utilisations pratiques de notre solution en mobilisant les informations de distance à un semis régulier d'images de synthèse en complément des résultats de l'appariement. Cela prend la forme de trois scénarios d'utilisation

illustrés par des exemples et des analyses spatiales et statistiques.

Le chapitre final tire les principales conclusions et propose des perspectives de poursuite du travail. Pour information, les sources des figures sont citées dans les légendes. Dans le cas où aucune source n'apparaît, la figure est une production personnelle.

Chapitre 1 : Géomatique, infographie et fusion de données

Chapitre 1

Géomatique, infographie et fusion de données

Le couplage de bases de données géographiques et de bases de données photographiques nécessite le rapprochement entre deux domaines distincts mais liés. Une photo capture des objets. Ces objets possèdent une position géographique et sont de plus en plus systématiquement intégrés dans une base de données.

Dans notre contexte de travail nous avons plusieurs sources de données :
- les données photographiques prises par des appareils photo.
- les localisations géographiques des points de prise de vue de ces photographies.
- les angles de rotation sur les trois directions de rotation dans l'espace de l'appareil photographique.
- les données géographiques de l'espace urbain étudié (rues, bâtiments...).

La multitude des sources de données nous incite à réaliser une fusion de toutes ces données pour arriver à décrire le contenu photographique.

Nous présentons dans le premier chapitre les trois disciplines mises en œuvre dans ce travail : géomatique, infographie et fusion de données

La première est la géomatique dont nous rappelons d'abord quelques notions puis listons les technologies localisant un individu ou un objet. Ensuite, nous exposons les techniques de géolocalisation utilisant le GPS, technologie la plus connue et la plus utilisée tant par les professionnels que par le grand public. Enfin nous abordons les autres technologies en commençant par les systèmes en extérieur (outdoor) comme la triangulation des lignes GSM et en terminant par les techniques en intérieur (indoor).

La position géographique ne suffit pas dans certains cas d'utilisation. Il est parfois nécessaire de mesurer l'orientation autour des trois axes de rotation dans l'espace pour donner un positionnement plus précis (orientation, direction...).

Tout au long de cette partie on s'intéresse aux différentes technologies utilisées jusqu'à présent dans la géolocalisation. Le niveau de détail accordé à chaque tech-

nique ou technologie est en rapport avec son utilisation ou les expérimentations effectuées au cours des trois années de la thèse.

Le second domaine est l'imagerie qui présente les techniques de traitement d'image adaptées à l'analyse d'images urbaines. Nous présentons quelques notions d'analyse d'image et de segmentation.

Le troisième domaine est la fusion de données. Nous commençons par la présentation du processus général de la fusion de données. Ensuite nous nous concentrons sur la fusion de données géographiques en précisant les différentes approches pour apparier des bases de données géographiques. Puis quelques exemples de fusion de données géographiques sont présentées brièvement. Enfin nous présentons l'appariement de graphes et l'appariement de lignes qui sont au cœur de la thèse.

1.1 Notions de géomatique

Définition

Géomatique combine "Géo"qui veut dire terre, et "matique" qui vient d'informatique. Il recouvre le traitement automatique de l'information géographique.

Selon le département des sciences géomatiques à l'Université Laval, le mot géomatique regroupe "l'ensemble des connaissances et technologies nécessaires à la production et au traitement des données numériques décrivant le territoire, ses ressources ou tout autre objet ou phénomène ayant une position géographique." [LAVAL 07]. C'est une science qui s'est organisée autour des Systèmes d'Information Géographique (SIG).

Il existe de nombreuses définitions des Systèmes d'Information Géographique [Bruslé 07]. Un SIG est défini par la FICCDC (Federal Interagency Coordinating Committee on Digital Cartography) comme "un système informatique de matériels, de logiciel et de processus conçu pour permettre : la collecte, la gestion, la manipulation, l'analyse, la modélisation et l'affichage de données à référence spatiale afin de résoudre des problèmes complexes d'aménagement et de gestion." [Denègre 96].

Les SIG ont accompagné l'évolution des outils informatiques depuis les années soixante [Tupin 06, Muller 06]. Les SIG permettent de gérer un ensemble de données dans l'espace qui s'agisse d'un bâtiment, d'une zone ou de l'évolution spatiale ou temporelle d'un phénomène.

Actuellement les domaines d'application des SIG sont variés [Habert 00]. De la gestion du cadastre à la gouvernance des infrastructures en passant par la localisation des clients, les SIG sont mises en œvre dans le marketing, la protection civile, la planification urbaine, les études de déplacement de populations animales...

Un SIG a cinq grandes fonctions [Denègre 96, IAAT 03] souvent appelées les "5A" des SIG :

1. Abstraction : un SIG suppose une certaine modélisation du monde réel. Il permet d'abord de rendre compte de l'abstraction de cette réalité.
2. Acquisition : l'acquisition des données sert à alimenter le SIG en données. Il s'agit de saisir ou d'importer les formes des objets géographiques puis d'indiquer leurs attributs et relations avec les autres objets.
   Dans notre travail ce sont les photographies qui alimentent la base de données géographique et les bâtiments.
3. Archivage : l'archivage des données se traduit par le transfert des données saisies sur l'espace de travail vers un emplacement de stockage et d'archivage dans des bases de données.
4. Analyse : un SIG propose des fonctions d'analyse des données qui permettent de donner une réponse exacte aux questions que l'on se pose (longueur totale d'un réseau routier dans une zone, nombre d'habitants profitant d'une étendue d'eau, nombre de bâtiments dans une zone...).
   Nous avons utilisé des descripteurs d'analyse spatiale afin d'aider à localiser les photographies.
5. Affichage : un SIG permet d'afficher des résultats. La carte demeure la meilleure façon de percevoir les relations spatiales. Les cartes sont produites automatiquement avec tout ce que l'utilisateur demande comme information.
   Nous avons produit dans la thèse des cartes mais aussi des géovisualisations 3D.

1.1.1 Bases de données géographiques

Sources de données géographiques

Les producteurs de cartes en France sont principalement l'Institut Géographique National (IGN), la Direction générale des Impôts (DGI), le Service hydrographique et océanographique de la marine (SHOM), le CNES... [Ecobichon 94].

D'autres organismes créent aussi leurs propres données géographiques comme les collectivités territoriales, les villes, les conseils régionaux. Enfin on peu citer les grandes entreprises qui gèrent des réseaux de canalisations, des stations ou une flotte de véhicules comme EDF-GDF, France Telecom, Veolia, Lyonnaise des eaux...

Au niveau mondial plusieurs compagnies fournissent des cartes géographiques pour le grand public comme Google, Yahoo, Microsoft,... [Laurini 10]. Les bases de données urbaines que nous utilisons dans ce travail sont issues de l'IGN.

Organisation des données

L'information dans un SIG est organisée classiquement en couches qui regroupent les informations de type analogue.

Il existe deux catégories de couches : les couches raster et les couches vectorielles. Une couche raster est assimilée à une image ou un ensemble de pixels (picture element) [Lim 08, Joel 91]. Le mode raster s'applique principalement au traitement d'images satellitaires ou de photos aériennes et à la télédétection.

La résolution est la taille que représente un pixel sur le terrain [Magdelaine 11]. Les images prises par les satellites SPOT ont une résolution de 10 mètres en mode panchromatique (noir et banc).

Dans une couche en mode vecteur les objets sont spatialement définis par une séquence de coordonnées décrivant leur contour [Joel 91]. C'est une structure économe en volume de stockage.

Une base de données SIG en mode vecteur est constituée de trois types d'objets géométriques :

- Le point : un objet ponctuel peut représenter une ville, un arbre, un centre... selon l'échelle de saisie.
- La ligne : les objets linéaires peuvent être présentés par un arc ou par une simple ligne. Une rivière, une route ou un chemin de fer sont les candidats idéaux à cette géométrie.
- Le polygone : certains objets sont représentés d'une façon surfacique. Un polygone représente la forme et la localisation d'objets homogènes, comme des lacs, des communes, des forêts...

Un fleuve peut être représenté par une ligne ou une surface en fonction de l'échelle de saisie.

A chaque entité spatiale sont associées dans un SIG des caractéristiques qualitatives et quantitatives [Christine 10]. Ces spécifications sont de type alphanumérique et sont archivées dans une table d'attributs.

Ces informations ont pour rôle de décrire l'objet en précisant par exemple son nom, son libellé, sa date de création ou toute autre information pertinente. A un bâtiment on peut attribuer les informations suivantes : identifiant, nombre d'étages, hauteur, superficie... Toutes ces informations sont stockées dans une structure appelée table attributaire.

Ces données sont actuellement de moins en moins organisées en couches mais stockées dans des bases de données spatiales contenant les trois types d'objets géométriques avec leurs attributs. Cette forme permet de stocker et requêter des données reliées à des objets référencés géographiquement. La fouille de données spatiale dans une structure de bases de données relationnelle est beaucoup plus performante que celle en format de couches [Aufaure 00].

Topologie

Il existe deux méthodes d'organisation des données en mode vecteur [Pumain 03]. La structure dite "spaghetti" ou filaire consiste à enregistrer chaque structure indépendamment les unes des autres. Chaque segment a deux sommets qui ont deux coordonnées chacun. Un polygone est constitué par les coordonnées de chaque sommet. Ce modèle est simple mais ne rend pas compte des relations géométriques entre les objets.

Cette méthode d'organisation des données est susceptible de créer des anomalies dans la représentation des objets : polygones non fermés, chevauchements ou intersections entre polygones adjacents.

La deuxième structure, dite topologique, résout ce problème. Les classes d'entités sont organisées suivant des relations entre ces entités. Chaque arc est caractérisé par un nœud de départ, un nœud d'arrivée et un sens. La structure topologique élimine donc toute intersection de lignes (les nœuds représentent les intersections) et assure la fermeture de tous les polygones.

1.1.1.1 Géoréférencement

Une localisation à la surface de la terre se fait au moyen de système de coordonnées géographique en longitude et latitude. Elle s'exprime aussi sous la forme d'une représentation plane en coordonnées projetés [Rollet 05].

Du fait que la terre n'est pas une sphère, mais un géoïde irrégulier difficile à utiliser dans les calculs, les scientifiques préfèrent utiliser un ellipsoïde pour simplifier la modélisation du globe terrestre.

Afin de préserver l'ellipsoïde de la terre en le projetant sur un plan, les géomaticiens et géographes utilisent plusieurs types de projection dont les principales sont : la projection cylindrique (suivant un cylindre), la projection conique et la projection azimutale. En France, le système de projection choisi est le Lambert, c'est une projection de type conique (voir figure 1.1 [Sigea 11]).

Depuis 2009 le système officiel utilisé en France est le RGF93 (Réseau Géodésique français calculé pour l'année 1993) [Certu 08a, Mougel 03, Legifrance 06]. Ce dernier utilise une projection conique conforme Lambert appliquée à la France entière [Certu 08c] (Lambert93) et 9 projections appelés coniques conformes [Certu 08b].

Les données de positionnement que nous manipulons sont projetées conformément au système RGF93.

1.1.1.2 Outils SIG

Les outils informatiques utilisés dans les SIG se sont diversifiés [Sig-france 11, Klipfel 07, Sorel 10]. On trouve des simples visionneuses gratuitement en téléchar-

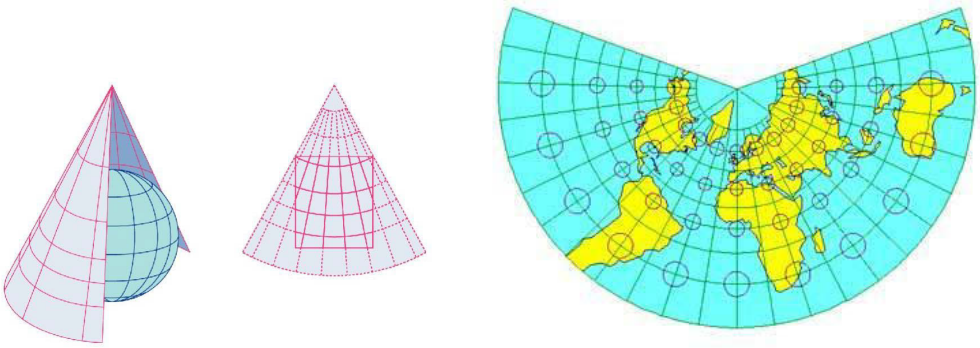


FIGURE 1.1 – Projection conique Lambert.(Source : Sigea(Systèmes d'Information Géographique pour l'Enseignement Agricole))

gement libre sur internet. Plusieurs petites plateformes open source existent aussi. Dans le monde du libre, de nombreux applications existent dont la plus ancienne et la plus complète est GRASS GIS (Geographic Resources Analysis Support System) qui est le fruit des recherches des ingénieurs de l'armée américaine dans les années 80. C'est le seul logiciel SIG complet et véritablement libre.

QGIS (Quantum GIS) [QGIS 11] est une application libre et autonome, couramment utilisée et qui peut se connecter à GRASS.

Les produits SIG commerciaux sont aussi variés et vont du plus simple au plus complexe.

En France un des SIG les plus utilisés est MapInfo de Pitney Bowes Business Insight (PBBI) qui propose une vaste liste de fonctionnalités en restant un outil proche des applications de bureautique.

Un produit français très utilisé est GéoConcept [Geoconcept 11] qui date de 1990.

Un des leaders historiques au niveau mondial est ArcGIS d'ESRI (Environmental Systems Research Institute). Ce produit contient une panoplie d'applications interconnectées : ArcView pour l'analyse et le traitement des cartes 2D, ArcScene pour les cartes 3D... Il existe également des adaptations de ces applications pour les mobiles (ArcGIS mobile, ArcPad et ArcGIS Desktop mobile). ESRI offre des fonctions permettant de se rapprocher de l'environnement bureautique standard. La simplicité de l'interface [Sig-france 11] et la richesse en fonctionnalités ont fait d'ArcGIS le logiciel SIG le plus vendu au monde.

Ces atouts ainsi que sa disponibilité au sein de notre structure d'accueil nous ont conduit à utiliser la suite ArcGIS dans nos calculs, représentations et gestion de données 2D et 3D durant cette thèse. L'extension 'Spatial Analyst' nous a permis d'effectuer nos statistiques spatiales du chapitre 5.

Tous ces produits proposent des modules de gestion des données en trois dimensions et des versions nomades pour les Pocket PC les Smartphones.

1.1.2 Les technologies de positionnement

1.1.2.1 Positionnement Outdoor (en extérieur)

Le premier système de positionnent satellitaire global est le TRANSIT en 1964 de l'U.S NAVY. Il utilisait l'effet doppler pour donner une position toutes les heures en moyenne et avec une précision au sol d'un kilomètre [Soulier 73].

En 1994, le département de la défense des États-Unis lance le système GPS qui est le seul système global de navigation par satellites entièrement opérationnel [EDUCNET 11].

A l'heure actuelle, la plupart des grandes nations lancent leurs propres satellites pour couvrir de manière autonome leur territoire. La première fut la Russie, suivie de l'Europe puis de la Chine avec son système Compass.

L'Inde a décidé en 2006 de lancer son propre système de positionnement régional IRNSS (Indian Regional Navigarional Satellite System).

Le Japon a commencé la mise en place de trois satellites qui couvrent le Japon et ses environs. Ces satellites permettront une augmentation de l'information donnée par le système GPS. Ce projet Japonais est baptisé QZSS (Quasi-Zenith Satellite System).

Le système SBAS (Satellite-Based Augmentation System) est formé actuellement par quatre systèmes d'amélioration de signal GPS. Chaque région du globe (les États-Unis, l'Europe [Correia 06], l'Inde et le Japan) a lancé des satellites géostationnaires qui complètent les satellites du système GPS afin de donner une meilleure précision.

Afin d'avoir une indépendance par rapport au système de positionnement par satellite américain l'Europe s'est décidée depuis 2001 à créer son propre système de positionnement par satellites. Les deux premiers satellites de la constellation ont été lancés le 21 Octobre 2011 [LEMONDE 11].

GPS

Le système de géolocalisation mondial le plus répandu est le Global Positioning System communément abrégé par "GPS".

Ce projet nait dans l'US Air Force en 1963. Quatre ans plus tard le premier satellite est mis en orbite. En 1978 le système est déclaré opérationnel et le Ministère américain des transports est désormais le premier utilisateur civil de ce système. D'autres utilisations à des fins civiles sont autorisées avec une précision dégradée de 100m et contrôlée par le ministère de défense américaine.

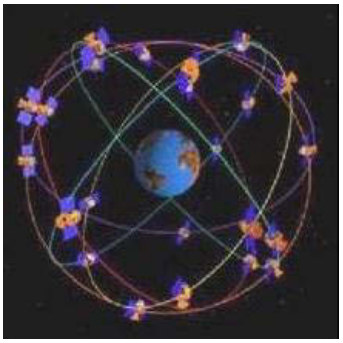
Bill Clinton lève le 1er mai 2000 [Clinton 00] la réduction de précision imposée par les militaires américains, ce qui permet au monde d'avoir une géolocalisation dix fois plus précise. Le GPS est toujours géré par le département de la défense des

États-Unis.

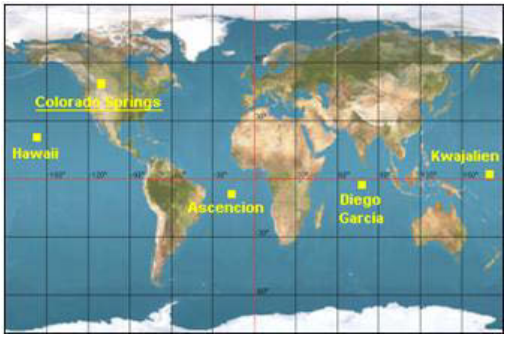
Actuellement les 24 satellites américains baptisés NAVSTAR (Navigation Satellite Timing And Ranging) gravitent autour du globe terrestre selon 6 orbites à 20 200 km d'altitude [Lechner 00, Mayer 11]. Ces orbites sont schématisées dans la figure 1.2a.

Cette disposition permet d'avoir en tout endroit à 99.9 % du temps au moins 4 satellites visibles. Un satellite est dit visible si l'angle d'inclinaison qu'il fait avec l'horizontale est supérieur à 5 °.

Il existe cinq stations de surveillance qui forment le segment de contrôle illustrées dans la figure 1.2b). Ce dernier assure le bon fonctionnement du système. Ces stations sont gérées par une station maîtresse située aux États-Unis. Elle envoie les signaux de correction aux satellites après réception des signaux des autres stations de contrôle.



(a) Orbites des satellites



(b) Les 5 stations terrestres de contrôle et de commande

FIGURE 1.2 – Système GPS. (Sources : centre national d'études spatiales (CNES) et [EDUCNET 11])

Le principe de fonctionnement est simple. Chaque satellite émet des ondes électromagnétiques qui se propagent à la vitesse de la lumière. La mesure du temps mis par le signal pour parvenir au récepteur détermine la distance entre le satellite et le récepteur. Ces temps sont de l'ordre de 67 millisecondes pour un satellite se trouvant au zénith. Cette durée est calculée en discernant la différence entre le temps d'émission et le temps de réception du signal.

La synchronisation des horloges des satellites et des récepteurs est un élément fondamental pour assurer le bon fonctionnement du système GPS. Les erreurs sont d'origines diverses. Certaines sont naturelles et d'autres sont techniques. Le signal est retardé par l'ionosphère, couche large de quelques centaines de kilomètres de l'atmosphère terrestre. Ce retard est en partie modélisé et corrigé par les récepteurs GPS terrestres. Les couches atmosphériques non modélisées engendrent une erreur de 4 m [Correia 06].

Les erreurs d'origine technique sont principalement la dérive de l'horloge du satellite, les bruits au niveau du récepteur et la dérive de l'horloge du récepteur.

Toutefois la précision varie selon la géométrie des satellites émettant le signal vers le récepteur. Le coefficient d'affaiblissement de la précision ou GDOP (Geometry Dilution Of Precision) est l'indicateur de cette source d'erreur. Les spécialistes en positionnement par satellite ont défini quatre facteurs de précision permettant de donner une approximation de l'erreur de positionnement par GPS.

1. Le PDOP (Position dilution Of Precision) est l'indicateur de précision en trois dimensions.
2. Le HDOP (Horizontal Dilution Of Precision) est l'indicateur de précision en deux dimensions ce qui est traduit par les deux valeurs de longitude et latitude.
3. Le Vertical Dilution Of Precision (VDOP) informe de l'erreur sur l'altitude.
4. Le TDOP (Time Dilution Of Precision) détermine la précision de l'heure.

Ces facteurs sont traduits en une précision donnée en mètres. À chaque unité de HDOP correspond une imprécision de 5.5 mètres. Dans la pratique chaque récepteur prend en considération un ou deux coefficients d'affaiblissement de la précision.

Deux phénomènes s'ajoutent aux éléments qui minimisent la précision de la localisation satellitaire. Le phénomène de multi-trajets (figure 1.3) résulte de la réflexion des signaux sur la surface de la mer pour les bateaux, par les arbres et les bâtiments (surtout s'ils sont vitrés) sur la terre ferme. Les réflexions parasites génèrent 1.4 m d'imprécision.

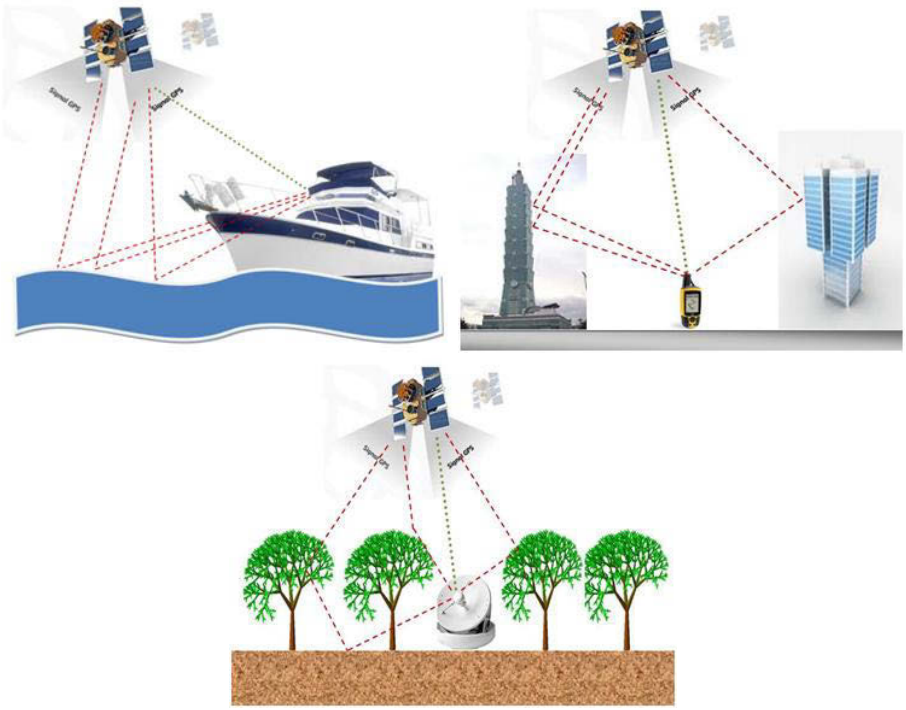


FIGURE 1.3 – Perturbation du signal par multi-trajets. (Source : production personnelle inspirée de [Ray 09])

Le deuxième phénomène est l'atténuation du signal à cause d'arbres ou de bâtiments par exemple ou de tout autre type d'obstacle se trouvant entre les satellites et le récepteur.

Ces incertitudes de positionnement peuvent s'aggraver en cas de dysfonctionnement des satellites ou simplement d'arrêt simple pour maintenance ou panne. Ces incidents se produisent souvent mais sont corrigés usuellement dans un délai variant entre trois et six heures. De plus les gestionnaires du système GPS ont prévu quelques satellites de secours qui gravitent déjà sur chacune des orbites et qui ne sont utilisés qu'en cas de besoin.

GPS différentiel

Le GPS permet d'avoir une précision suffisante dans la majorité des utilisations. Cependant une précision plus importante est requise dans certains domaines. Les gardes-côtes américains furent les premiers à en avoir le besoin lors du guidage d'approche de bateaux aux ports [Rudiger 11]. Ce besoin a fait naître le DGPS (Differential Global Positioning System) qui permet de corriger les erreurs du GPS en s'aidant de stations de référence au sol géolocalisées avec une haute précision.

Un abonnement auprès des stations de référence est nécessaire. En France, l'IGN offre gratuitement l'accès à quelques stations fixes appelé RGP (Réseau GPS Permanent).

Dans ce travail de thèse nous cherchons à se localiser en ville, ce qui rend utile l'utilisation du GPS différentiel.

Nous avons utilisé le GPS différentiel Trimble GeoXT. Une sortie de terrain se prépare en paramétrant l'application "GPS Pathfinder" tout en précisant le système de projection, en chargeant la carte en tant qu'arrière plan et en transférant les données sur le GPS.

L'application TerraSync installée sur le GPS permet ensuite de collecter des données de type point, ligne ou surface. Une fois les données collectées, c'est le tour de la phase de post traitement. Il est nécessaire de rebrancher le GPS à un ordinateur afin d'extraire les données et de les corriger.

En chargeant du site *http ://rgp.ign.fr/DONNEES/diffusion/* le fichier de la station la plus proche et surtout en veillant qu'elle a été bien disponible au moment du relevé. La correction avec ce type de dispositif permet d'avoir des précisions inframétriques, précision inatteignable avec les prises de positionnement GPS originales.

AGPS

L'Assisted GPS repose sur des récepteurs GPS très simplifiés et assistés par un serveur connecté aux réseaux téléphoniques mobiles. Le serveur assiste le récepteur GPS en lui transmettant la liste des satellites visibles *à priori*. Le récepteur capte

tous les signaux (même les plus faibles) et les transmet via le réseau au serveur qui fait les calculs nécessaires pour déterminer la meilleure estimation de la position.

Ce système améliore considérablement la rapidité de localisation dans un milieu urbain. Cependant il n'a pas d'effet sur la précision de géolocalisation.

1.1.2.2 Positionnement Indoor (en intérieur)

Les signaux GPS sont incapables de traverser les murs de bâtiments. Dans les milieux intérieurs (indoor) d'autres dispositifs permettent une localisation : GSM, WiFi, RFiD...

La combinaison de ces moyens est souvent la meilleure solution à adopter [Becker 08, Nagel 10].

GSM

L'utilisation des antennes GSM (Global System for Mobile Communications) pour la géolocalisation est possible par le moyen de plusieurs méthodes. La technique la plus répandue actuellement est l'identification de la cellule de service de la téléphonie. Elle ne nécessite pas de configuration particulière contrairement à la triangulation par exemple.

Une connexion au réseau internet est nécessaire pour définir la position. En effet chaque borne émettant des ondes GSM (EDGE, UMTS...) possède une position et un identifiant unique. Un rayon autour de cette borne est donné comme positionnement. Ce rayon varie selon la présence d'autres bornes et la force du signal reçu par le téléphone.

L'utilisation d'un terminal GSM pour la géolocalisation offre une faible précision qui varie de 100 mètres à plusieurs kilomètres en milieu rural.

WiFi

Le Wireless Fidelity est un réseau qui permet à l'origine de relier des appareils informatiques au sein d'un réseau sans fil. Le WiFi utilise des signaux avec des larges bandes de fréquence pour assurer un haut débit. Durant ces dernières années, cette technologie a servi aussi à se positionner [Cisco 06]. La force du signal reçu par un émetteur WiFi dont la position exacte est connue permet de calculer la distance à l'émetteur. La triangulation nécessite d'avoir plusieurs émetteurs à proximité et de connaître la position de ceux-ci.

Le WiFi Positionning system (WPS) est un des premiers systèmes réalisés aux États-Unis en 2005 permettant de donner un positionnement géographique avec une précision de 20 à 40 mètres.

En France la première application mobile de géolocalisation a eu lieu dans le

centre commercial du quartier de Paris-La Défense en février 2011 [Amiot 12]. La cartographie des quatre étages et l'installation de bornes WiFi supplémentaires ont permis d'avoir une application mobile affichant avec précision la position du porteur du téléphone ainsi que toutes les possibilités qui s'offrent à lui aux alentours (ascenseurs, magasins, promotions, escaliers...).

RFID

Le Radio Frequency Identification [Cisco 06] offre initialement une géolocalisation à un instant t d'un objet en contact direct avec un capteur X géolocalisé avec précision.

Un badge, une carte ou une puce intégrée dans un objet permet d'avoir à un instant précis la position de cette puce. Dans la pratique un récepteur RFID capte la présence d'une puce RFID à une distance de quelques centimètres.

Actuellement, le développement des RFID actifs permet une localisation en temps réel dans un environnent plus large allant jusqu'a 100 mètres.

Conclusion

D'autres techniques de positionnement existent dans le commerce tel que le Bluetooth, le rayonnement par infrarouge ou ultrason et tout autre type de capteur capable d'émettre ou de recevoir un signal [Ray 09, Delot 06].

Chaque technique a ses points forts et ses faiblesses. Les GPS sont incapables de fournir des informations de positionnement à l'intérieur des bâtiments ou dans un milieu urbain très dense [Deveryware 11]. Ils sont aussi très lents lors de l'amorçage. On ne peut pas se servir des bornes Wifi en zone rurale faute de précision. La dépendance au réseau GSM se traduit par la nécessité de trouver une couverture 3G ou GPRS (General Packet Radio Service) au moins pour avoir accès à internet, ce qui n'est pas possible partout.

Notons que la solution WiMAX (Worldwide Interoperability for Microwave Access) [CETE 06] permet d'offrir une connexion haut-débit dans les endroits isolés. Mais cette technologie donne une localisation dont la précision est trop faible pour la quasi-totalité des utilisations.

En conclusion, la solution est de combiner ces techniques dans des systèmes hybrides. C'est celle qu'ont adoptée les constructeurs des nouveaux téléphones "intelligents" (Smartphones) qui offrent des appareils permettant une géolocalisation hybride en optant toujours pour la technique la plus adaptée selon la situation.

Ces appareils utilisent les satellites GPS s'ils sont à l'extérieur des bâtiments, les réseaux GSM en cas d'absence de signal GPS et proposent un positionnement en utilisant la connexion aux réseaux WiFi aux alentours pour géolocaliser le terminal.

Malgré un coût très élevé, on déploie de plus en plus une infrastructure de ré-

péteurs de signal GPS à l'intérieur des bâtiments pour avoir la continuité du signal GPS.

1.1.3 Techniques de mesure d'orientation

Un positionnement exact est une information de plus en plus recherchée mais un autre paramètre important est l'orientation. Dans le domaine de la photographie par exemple, l'orientation de la caméra en plus de sa position est utile latéralement pour repérer l'édifice visé et verticalement car cela influe sur la part du sol dans la photographie.

L'information sur la direction est donnée par une simple boussole. L'orientation autour des axes autres que l'axe vertical est fournie par des centrales à inertie. Des capteurs permettant ces calculs d'inclinaison sur les 3 axes possibles de rotation, sont actuellement implantés dans divers appareils mobiles comme les Smartphones, les télécommandes de télévision ou les manettes de jeux vidéo.

Cette section présente ces instruments de mesure d'orientation à partir d'appareils manipulés au cours de mes travaux de thèse.

1.1.3.1 Boussole

Un des plus anciens instruments de navigation, la boussole a été inventée par les chinois et apportée en occident par les marchands arabes [Ampère 32]. Elle reste l'outil incontournable depuis des milliers d'années.

Les boussoles électroniques fonctionnent à l'aide de capteurs de champs magnétique. Elles coûtent beaucoup plus chère que les boussoles à aiguille et nécessitent un calibrage [Smith 11]. Cependant elles sont plus précises, de plus petite taille et fournissent les données en format numérique.

1.1.3.2 Centrale à inertie

Une centrale à inertie ou centrale inertielle est un instrument de mesure de haute précision servant à la navigation. Elle est composée d'un ensemble d'accéléromètres (capteurs d'accélération), de gyroscopes (capteurs de position angulaire) et de gyromètres (capteurs de vitesse angulaire) [Navigation 98]. Cet instrument fournit 3 positions et 3 vitesses autour des 3 axes de rotations.

Au début de ma thèse j'ai passé une journée en tant que visiteur à l'école navale en Bretagne. L'institut de Recherche de l'école Navale (IRENav) utilise des centrales inertielles pour construire des systèmes de transports intelligents. Leurs choix de centrale d'inertie était orienté vers la MTx de la marque XSENS [1]. Elle fournit des

1. En manipulant le MTx, j'ai pu constater que c'est un petit instrument (38x53x21 mm) qui ne pèse que 30 g. Il se connecte via USB aux ordinateurs ce qui offre une grande facilité pour créer

vitesses de rotation et des positions autours des 3 axes avec une imprécision toujours inférieure à 2 degrés.

La société InterSens vend un instrument baptisé NavChip capable de transmettre la position au sein d'un bâtiment en lui donnant en entrée uniquement la position initiale. Comme d'autres domaines technologiques, l'armée est le moteur majeur pour le développement de ces techniques de positionnement. Des puces minuscules (0,5 cm) sont commercialisées par la firme InvenSense à des prix très bas. Une quarantaine d'euros suffit pour avoir une puce donnant l'accélération sur les 3 axes de rotations [InvenSens 11].

L'utilisation de centrale inertielle permet un positionnement précis dans tous les environnements (couvert ou pas). Le LORIA à Nancy [Aron 06] a travaillé sur le couplage entre l'utilisation d'une centrale inertielle et le traitement de vidéo pour localiser des objets sur une scène en réalité augmentée.

La société française Sagem construit des gyroscopes avec toutes les classes de précisons [Sagem 11]. Ces équipements sont largement utilisés dans les appareils navigant dans l'air et en mer y compris dans les sous-marins.

Le niveau de précision de ce type de dispositifs s'est avéré inutile dans le cadre de notre travail.

1.1.3.3 Systèmes hybrides

L'avancement de la technologie a entraîné une miniaturisation des instruments de mesure. De ce fait on trouve des boussoles intégrées dans des appareils photos, dans des Smartphones et même dans des montres. Les appareils photographiques sont ainsi de plus en plus souvent équipés de GPS et de boussoles électroniques. Notons que le premier appareil photographique numérique équipé à la fois d'un GPS et d'une boussole fut le Ricoh Caplio 500SE (voir figure 1.4). Au sein de notre unité de recherche nous avons acquis cet appareil.

Il permet un confort lors de la prise de vue en ajoutant au sein même du fichier de la photo les coordonnées GPS, la précision de cette position, l'altitude, la vitesse, l'orientation donnée par la boussole... Il est possible de récupérer ces informations dans l'extension EXIF du fichier de la photo.

D'autres appareils photos peuvent comporter des dispositifs de localisation soit en les plaçant sur la griffe du flash (JOBO photoGPS) soit par une simple synchronisation *ex post*. Ils sont fondées sur la synchronisation temporelle de l'appareil photographie et du GPS.

Les nouveaux Smartphones contiennent désormais une multitude d'instruments de mesures. Boussole, gyroscope, accéléromètre, puce GPS, appareil photo sont les

des applications. Cette centrale inertielle vaut dans les environs de 2000 €, beaucoup moins que les modèles incluant en plus une puce GPS qui se vendent au double de ce prix. Ces dispositifs sont capables d'émettre leurs données de positionnement 100 fois en une seconde.



FIGURE 1.4 – L'appareil Ricoh Caplio S500SE équipé de puce GPS et d'une boussole numérique. (Source : documentation de l'appareil)

exemples qui deviennent de plus en plus courants dans un iPhone, un Smartphone HTC, Samsung ou Motorola.

L'accéléromètre "LIS302DL" équipe les iPhones et quelques autres Smartphones actuels. Petit (3 X 5 X 0.9 mm), il offre une précision comparable aux accéléromètres présentés dans la section précédente.

Un des Smartphones du marché au début de l'année 2009 était le HTC Hero, acquis lui aussi par le centre de recherche. Cet appareil permet de téléphoner mais aussi de prendre des photographies, de préciser une position sur une carte à l'aide de la puce GPS embarquée. En cas de mauvaise réception ou d'absence totale du signal GPS, la position est donnée par un rayon autour de la borne WiFi à laquelle l'appareil est connecté. A défaut de cette information, il fournit la position sous forme d'un rayon autour de l'antenne d'émission GSM. Le HTC Hero permet aussi de donner l'angle de rotation sur les 3 axes de rotation dans l'espace.

1.2 Notions en traitement d'image

La vision assistée par ordinateur comprend trois volets :

1. L'acquisition des images : c'est le processus de prise de l'image et le point de départ de la vision par ordinateur. Ce volet s'intéresse au choix du type de capteur : infrarouge, ultraviolet, optique...

   Il s'intéresse aussi au positionnement et au paramétrage du capteur afin d'assurer le meilleur résultat exploitable.

2. Le traitement d'image permet d'analyser l'image acquise. Il permet d'enlever du bruit, de segmenter, de filtrer, de modifier le contraste...

3. L'analyse de l'image est l'interprétation du résultat de l'image traitée. Cette phase permet de retrouver un objet dans une image, de distinguer des formes...

La vision par ordinateur intervient de nos jours dans de nombreux domaines [Manzanera 09] : surveillance vidéo, médecine, imagerie satellite, contrôle qualité

dans les usines, robotique mobile...

La vision assistée par ordinateur s'est toujours inspirée de la vision humaine. Cependant, l'œil utilise un processus trop complexe [Nixon 02] pour l'imiter intégralement.

La vision humaine est susceptible d'être trompée par plusieurs phénomènes comme la différence de contraste ou toute autre illusion optique [Nixon 02]. La vision par ordinateur a l'avantage de ne pas tomber dans ces phénomènes d'illusion. Mieux encore, elle rend visible ce qui est invisible pour l'œil humain. Les travaux menés à Télécom Paris Tech [Rossant 09] démontrent qu'une maladie au niveau de la rétine peut être diagnostiquée par traitement de l'image de la rétine. Le diagnostic de cette maladie se fait avant que les médecins spécialistes ne soient capables de la repérer.

Les caméras thermiques informent sur la chaleur des objets pris en photo. Cette information permet dans le domaine industriel de devancer la perception humaine. Ce sont des exemples parmi d'autres démontrant que la vision par ordinateur peut supplanter la vision humaine.

Ce chapitre propose une présentation générale du traitement d'image en se focalisant sur celle-ci et les opérations qu'elle peut subir. Ensuite sont définis les principes et les techniques utilisés pour la segmentation. La dernière partie s'attarde sur les méthodologies utilisées pour extraire les lignes d'une photo.

1.2.1 L'image numérique

Une image numérique est un signal bidimensionnel [Nasr 87]. Elle est définie par le nombre de pixels qui la constituent en largeur et en longueur. Chaque pixel possède un niveau de gris ou une couleur. Cette représentation de l'image offre la possibilité d'effectuer plusieurs opérations mathématiques sur la matrice des pixels d'une image.

L'image numérique est stockée habituellement dans des fichiers compressés. La compression permet de minimiser la taille des fichiers qui contiennent l'image en modifiant la structure de stockage de l'information. Le gain d'espace de stockage se fait aussi par élimination des détails présents sur la photographie et qui ne sont pas ou sont peu perceptibles par l'œil.

Le format de compression le plus utilisé est le JPEG (Joint Photographic Experts Group) [Rabbani 02]. Il permet d'avoir un espace minimal de stockage avec la possibilité d'effectuer quelques opérations sur les images au niveau matériel exécutées par des cartes électroniques [Manzanera 09].

1.2.2 Opérations sur les images

L'analyse d'images peut se faire à deux niveaux :

- Le bas niveau concerne les opérations sur l'image pour produire une autre image en utilisant les opérateurs bas niveau (filtrage, variation du contraste, modification de la résolution, manipulation de l'éclairage...).
- Les opérateurs haut niveau permettent une reconnaissance de forme ou de région. A ce niveau, on procède à l'interprétation de la scène.

Le pont entre ces deux classes est la segmentation. La segmentation permet de faire des modifications sur une image pour favoriser l'analyse d'image à haut niveau [Lilly 93].

On distingue trois types d'opération [Rosenfeld 82] qu'une image numérique peut subir :

1. Opérations ponctuelles : il s'agit de modifier la valeur d'un pixel sans considérer son voisinage. Une opération ponctuelle remplace la valeur du pixel par le résultat d'une opération ou en effectuant une addition ou un produit avec la valeur du pixel dans une autre image et dans la même position.
2. Opérations locales : la valeur d'un pixel dépendra de son voisinage [Cocquerez 85]. Les effets les plus courants dans ce mode d'opération sont l'élimination de bruit, la détection de contours et l'allègement de l'effet de flou.
3. Opérations géométriques : les pixels en sortie dépendront de la position d'autres pixels dans l'image. Il s'agit des transformations géométriques que subit l'image comme la translation, la rotation, le changement d'échelle...

   Ces transformations permettent de donner une nouvelle image ressemblant à l'originale avec des modifications de configuration géométrique [Lin 01].

1.2.3 Les filtres

Les opérations locales sont effectuées habituellement par l'application d'un filtre sur l'image. Un filtre est un opérateur qui va changer les valeurs des pixels d'une image. On distingue principalement trois types de filtres.

1. Les filtres de convolution permettent de diminuer le bruit en atténuant les détails de l'image. Ces filtres sont appelés passe-bas car ce sont des filtres laissant passer les basses fréquences en éliminant les hautes fréquences.
2. Le filtre passe-haut accentue les contours et les détails. Il a pour conséquence aussi d'accentuer le bruit.
3. Le filtre passe-bande élimine certaines fréquences d'apparition de pixels indésirables.

Un filtre de moyenne par exemple (voir figure 1.5) est un filtre passe-bas. Il permet de diminuer le bruit mais de créer aussi du flou sur l'image. C'est un opérateur considéré de bas niveau puisqu'il agit directement sur l'image pour en fournir une autre.

$$\frac{1}{9} \times \begin{array}{|c|c|c|}\hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline 1 & 1 & 1 \\ \hline \end{array} \Rightarrow \begin{array}{|c|c|c|c|c|c|}\hline 5 & 77 & 23 & 5 & 14 & 12 \\ \hline 14 & 44 & 86 & 1 & 12 & 2 \\ \hline 05 & 15 & 169 & 2 & 27 & 60 \\ \hline 147 & 36 & 52 & 196 & 141 & 146 \\ \hline 35 & 14 & 75 & 7 & 69 & 36 \\ \hline \end{array}$$

FIGURE 1.5 – Mise en place d'un filtre sur une image.

Concrètement on place le centre du filtre (ou masque) sur chaque pixel pour calculer une nouvelle valeur de ce pixel. La nouvelle valeur d'un pixel est la somme des produits des valeurs des pixels de l'image et du filtre. Dans notre exemple chaque pixel prendra la valeur moyenne des 8 pixels qui l'entourent dans les 4 directions en plus de sa valeur ancienne. Cette opération est connue sous l'appellation de convolution numérique.

En mettant par exemple le filtre sur le pixel de la colonne 4 et la ligne 3, le pixel en question prendra la valeur moyenne (76) au lieu de la valeur ancienne 2.

En manipulant les valeurs des coefficients et les tailles des filtres, on déclenche un traitement particulier sur une image. On trouve un énorme catalogue de filtres permettant de filtrer des éléments particuliers. A titre d'exemple, si on applique un filtre médian, on remplace la valeur d'un pixel par la valeur médiane (par rapport à l'entourage de ce dernier) et ceci pour chaque pixel de l'image. Ce filtre fournit l'effet poivre et sel s'il est de taille 3 X 3, et élimine le bruit s'il est de taille 5X5 ou plus.

La distinction de contours par exemple passe par des filtres divers comme celui de Sobel, Robinson, Roberts, Prewitt ou de Kirsh. Il est recommandé dans certains cas de passer plusieurs fois le même filtre (ou plusieurs filtres différents) pour avoir les résultats souhaités.

$$\frac{4}{(\alpha+1)} \times \begin{array}{|c|c|c|}\hline \frac{\alpha}{4} & \frac{1-\alpha}{4} & \frac{\alpha}{4} \\ \hline \frac{1-\alpha}{4} & -1 & \frac{1-\alpha}{4} \\ \hline \frac{\alpha}{4} & \frac{1-\alpha}{4} & \frac{\alpha}{4} \\ \hline \end{array}$$

FIGURE 1.6 – Filtre de Laplace.

A titre d'exemple, le filtre passe-haut de Laplace (figure 1.6), permet de détecter les zones de haute fréquence. Il s'agit des zones de grand contraste : les contours.

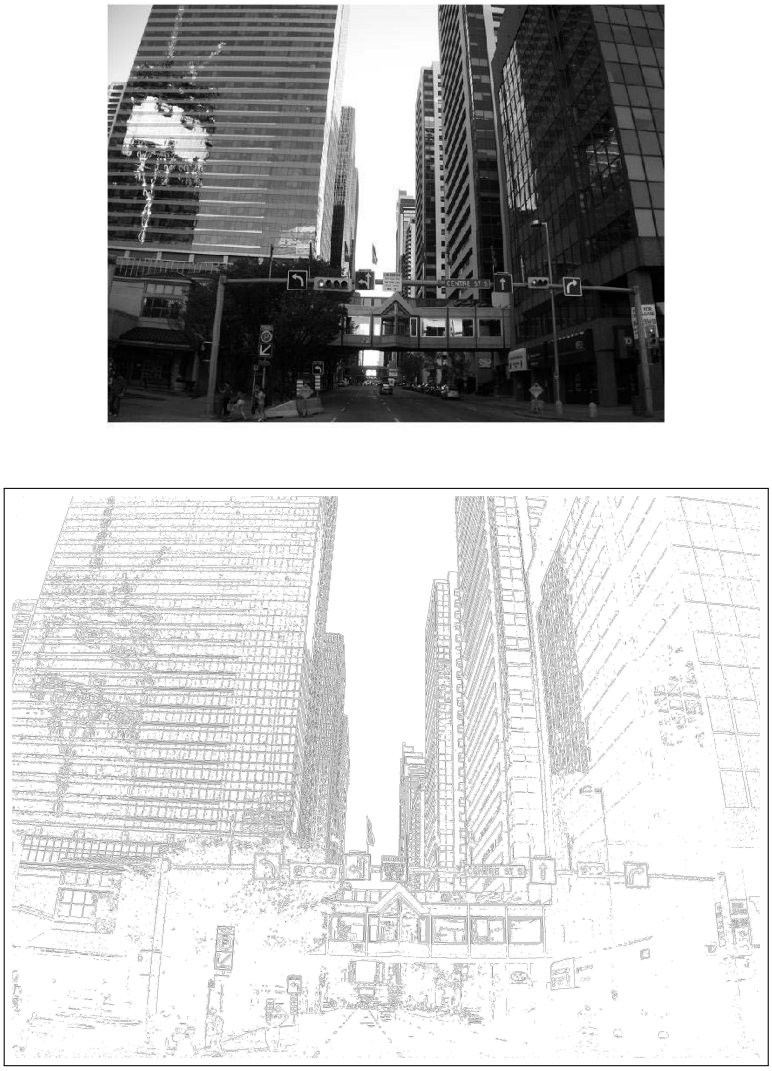


FIGURE 1.7 – Utilisation du filtre de Laplace pour détecter les contours. (Source : photo personnelle | rendu par Matlab)

En utilisant la valeur $\alpha = 0.2$, la valeur par défaut utilisée par Matlab, nous obtenons le filtre suivant :

0.1667	0.6667	0.1667
0.6667	−3.3333	0.6667
0.1667	0.6667	0.1667

Une image prise en scène urbaine est présentée dans la figure 1.7 avec l'illustration de l'effet du filtre de Laplace sur cette photographie.

1.2.4 Segmentation

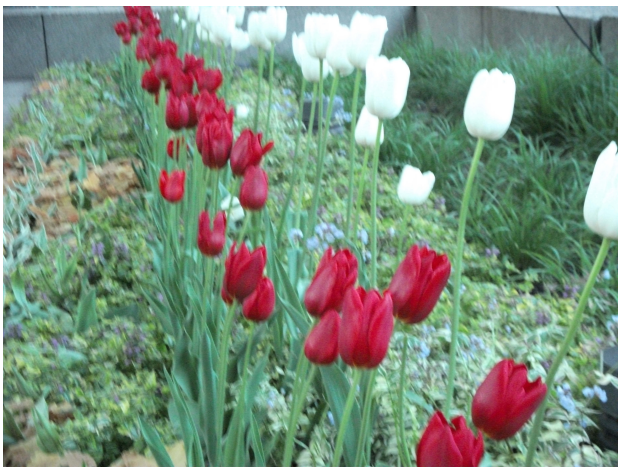
La segmentation [Baillie 03, Wladyslaw 94] est un processus de découpage d'une image en régions homogènes selon un critère donné. Ce critère peut être la couleur, la texture, le degré d'éclairage...

La segmentation est aussi une étape primordiale pour l'extraction d'informations qualitatives de l'image [Cheng 01]. C'est la segmentation [Lilly 93] qui permet de fournir les éléments nécessaires pour les descripteurs haut niveau.

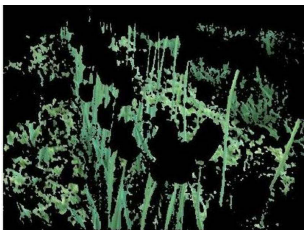
Le découpage en zones ou la segmentation d'effectue selon trois grandes classes d'algorithme :

1. La segmentation basée sur les pixels est la première famille des algorithmes de segmentation. L'image est considérée comme une matrice de pixels. L'information utilisée dans ce cas est la couleur (ou niveau de gris). Ce type de segmentation se base sur l'histogramme. Un histogramme [Nixon 02] est un graphique affichant l'intensité individuelle des niveaux de luminosité dans une image. Pour chaque pic de l'histogramme une zone homogène est définie. La segmentation basée sur les pixels se fait également par le regroupement des zones partageant des intervalles de valeurs.
2. Le second type d'algorithme de segmentation est basé sur les régions. Il s'agit de l'une des deux méthodes suivantes [Wladyslaw 94] :
   (a) La première consiste à choisir des zones homogènes puis à étendre chaque zone avec son entourage s'il se rapproche des caractéristiques de ces zones voisines.

   Une segmentation selon la couleur est illustrée dans la figure 1.8. Nous distinguons à titre d'exemple les zones vertes, les zones rouges et les zones blanches. Chacune des zones ou segments est regroupée dans une image à part.
   (b) La seconde méthode consiste à diviser l'image d'une façon non uniforme puis à essayer de regrouper les zones qui se rapprochent pour former de



(a) Photographie



(b) Segment vert



(c) Segment rouge



(d) Segment blanc

FIGURE 1.8 – Segmentation d'une image par la couleur.

grandes zones homogènes. Cette technique est connue sous l'appellation "split and merge".

3. La troisième catégorie s'intéresse aux contours des objets visibles dans l'image. Des filtres détecteurs de contours sont appliqués pour trouver les zones de contraste élevé qui représentent les contours. En d'autres termes, la segmentation par contours permet de délimiter chaque région de l'image par sa frontière [Bergounioux 09].

1.2.5 Détection de contours

Les ingénieurs en informatique et en électronique essayent de retrouver les mécaniques qu'utilise l'œil humain afin de les reproduire. Certains aspects de la vision humaine en tant que modèle sont recopiés fidèlement. D'autres aspects sont encore des mystères pour les scientifiques.

La médecine légale et la biométrie [Nixon 02], ont détaillé à l'aide des biologistes et des psychologues la perception des objets par l'homme. Ces analyses ainsi que l'observation des perceptions des illusions optiques ont montrés que l'œil utilise les contours des objets pour reconnaître le monde visuel [Hirsch 95]. L'œil est capable

de détecter un triangle blanc en voyant seulement le tiers qui est tracé [Kaiser 09, Rosenfeld 82]. Il est capable de reconnaitre un cube en absence de la majorité du traçage comme l'illustre la figure 1.9. Ce sont les lignes imaginaires que nous somme capable d'apercevoir et d'interpréter avec notre cerveau.

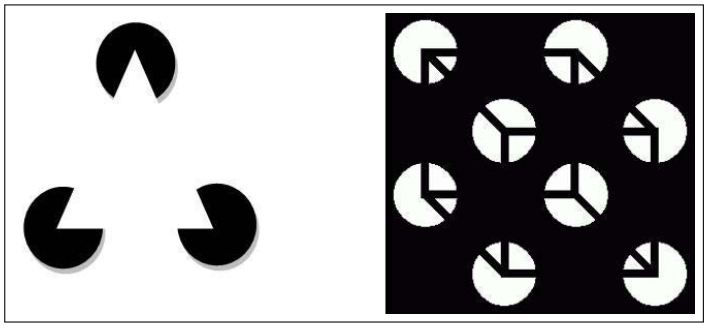


FIGURE 1.9 – Vision des contours par l'œil. (Source :[Kaiser 09, Petitesexperiences 11])

Détection de lignes

La détection des contours des objets est l'étape primordiale pour retrouver des objets simples. Les objets simples dans une image sont les droites et les cercles.

En se basant sur ces contours détectés, on détecte des objets plus complexes comme les coniques ou plus composés comme une voiture, un animal...

La détection des lignes [Nozick 10] est une problématique toujours d'actualité dans le domaine de la recherche en traitement d'image et vision par ordinateur. Une ligne fait un ou plusieurs pixels de large. Les lignes ne sont que rarement tout à fait droites et recoupent fréquemment d'autres objets de l'image.

Il existe plusieurs filtres locaux de détection de contours. Ce sont principalement les filtres de Sobel, de Robinson, de Roberts, de Prewitt et de Kirsh.

Nous illustrons par la figure 1.10 la détection des lignes par le filtre de Robinson et la transformation de Laplace-Gauss.

La méthode la plus récente des techniques de détection de lignes ou de segments sur une photographie est l'LSD [von Gioi 10] qui permet une détection rapide et robuste au bruit.

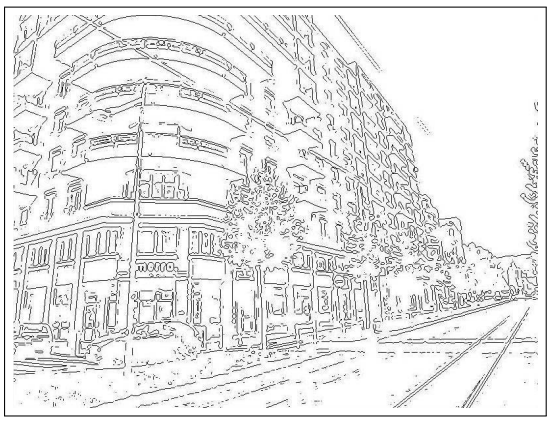
Des transformations globales permettent de distinguer les contours sur une image. Ces transformations sont principalement : la transformation de Fourrier, la transformation de Laplace et sa dérivée Laplace-Gauss. La transformation de Hough permet une reconnaissance de formes dans une image, elle est détaillée dans le chapitre 3. Il s'agit de la transformation adoptée pour analyser nos images dans ce travail de thèse.



(a) Photographie



(b) Filtre de Robinson



(c) Transformation Laplace-Gauss

FIGURE 1.10 – Détection de lignes avec le filtre de Robinson et la transformation Laplace-Gauss.

1.3 Processus de la fusion de données

L'interaction entre le domaine de l'image photographique et le domaine géographique fait partie du monde de la fusion de données.

La fusion de données est définie comme "un procédé traitant de l'association, de la corrélation et de la combinaison des données et de l'information provenant de sources uniques ou multiples afin d'aboutir à des estimées affinées des positions et des identités, à une évaluation complète en temps utile des situations et des menaces, et de leurs importances." [Grandin 06, Liggins 08].

C'est une discipline qui monte en puissance et concerne tous les domaines d'application : les services de sécurité, les collectivités locales, la logistique... qui profitent de la fusion de donnés pour prendre leurs décision.

Le modèle de fusion de données proposé en 1985 par le JDL (U.S. Joint Directors of Laboratories Data Fusion Group) décrit trois fonctions principales. Nous les illustrons dans la figure 1.11.

Les données différentes venant des sources différentes sont tout d'abord alignées.

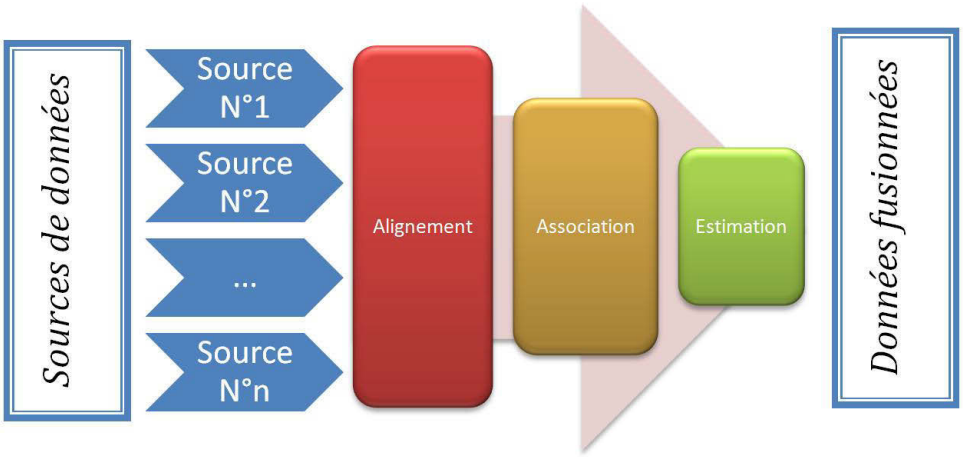


FIGURE 1.11 – Fonctions de la fusion de données. (Source : Y. Attia à partir de [Grandin 06] et [Liggins 08])

Cela consiste à uniformiser les informations et à les transformer en une représentation commune où elles peuvent être comparées.

L'association passe par la génération d'hypothèses d'appariement, de leur évaluation puis de la sélection du meilleur de ces appariements.

L'estimation est la fonction qui définit la combinaison résultante de la fusion.

Un exemple simple de fusion de données peut être trouvé dans le travail de S. Li et J. Gong [Li 06] qui ont créé un flux d'informations en fusionnant des données venant de plusieurs plateformes web de cartographie ainsi que de divers moteurs de recherche sur le web.

1.3.1 Appariement de données géographiques

L'appariement de données géographiques [Ruas 02] consiste à établir des liens de correspondance entre des ensembles d'entités géographiques. Ces ensembles symbolisent les mêmes phénomènes du monde réel [Olteanu1 08] dans des représentations différentes.

Il existe trois approches d'appariement de données géographiques [Ruas 02] :

1. Appariement sémantique ou attributaire : il est basé sur la notion d'identifiant unique pour chaque élément de la base de données géographique. L'effort de bien modéliser et nommer les différentes entités est essentiel pour permettre par exemple d'apparier un itinéraire avec une route. En absence d'un tel système fiable de description de données, l'appariement ne se fait qu'a travers la géométrie ou les relations topologiques.
2. Appariement géométrique : il s'effectue en trois phases. La première est la sé-

lection des candidats susceptible d'être appariés. Ensuite une phase de mesures s'entame. Il faut calculer les distances entre les éléments sélectionnés.

Notons que le calcul de la distance entre deux entités géométrique (polygone ou ligne) peut adopter plusieurs approches [Olteanu1 08] plus au moins complexes. Le calcul le plus simple est celui de la distance entre deux points. La distance entre des lignes ou polygones peut se faire par le calcul de la distance moyenne, la distance de Hausdorff, la distance de Fréchet...

3. Appariement topologique : il se base sur les relations de voisinage et de connexion qui existent entre les objets géographiques. Il se décompose en appariements de graphe et de réseaux. L'appariement de graphes examine les relations entre les nœuds. L'appariement des réseaux est une simplification de l'appariement des graphes en se basant plus sur les arcs que sur les nœuds.

Les chercheurs fusionnent des données géographiques différentes (LIDAR, photos aériennes...) avec d'autres données (vidéos, caméra laser, photographies...) pour construire un modèle 3D riche avec ces données [Haala 05, Sourimant 07].

Regrouper des cartes géographiques différentes est une fusion qui permet de fournir des informations très diversifiées pour les utilisateurs [Usery 05, Chen 09, Michel Roux 92].

Dans notre travail nous utilisons deux types d'appariements. L'appariement de graphes est un moyen de plus en plus utilisé pour faire correspondre des images, des molécules, des réseaux... Il fait l'objet du premier appariement étudié.

Le second type d'appariement est un sous ensemble de l'appariement de graphes, c'est l'appariement de lignes.

1.3.2 Appariement de graphes

La mise en correspondance de deux graphes se traduit par une recherche des isomorphismes de sous-graphes avec ou sans tolérance d'erreurs selon le domaine d'application [Jouili 08, Zampelli 06].

Une autre approche est le calcul de distance entre ces deux graphes [Sorlin 06, Sorlin 05]. C'est cette distance qui permet de déclarer si deux graphes sont semblables.

Actuellement, l'appariement de graphes permet de fusionner de nombreux types de données. La seule contrainte est de bien modéliser les données sous forme de graphes.

L'analyse d'images et la reconnaissance de formes ont bénéficié largement de l'appariement de graphes. Le graphe offre une représentation compacte, structurée, complète et facile à manipuler [Tupin 06]. Chaque élément ou partie détecté dans une image est représenté sous forme d'un nœud et les relations entre ces derniers sont modélisées par des arcs [Ramel 10, Jolion 05].

L'analyse des images pour les modéliser sous forme de graphe permet de nombreuses utilisations pour la reconnaissance d'images [Theoharatos 06, Boeres 04, Haugeard 10].

La modélisation des images microscopiques des diatomées (micro-algues unicellulaires) sous forme de graphe a permis à R. Ambauen, S. Fischer et Horst Bunke d'identifier les diatomées par appariement de graphes [Ambauen 03].

Des symboles graphiques linéaires (symboles de composants électroniques, signalisation...) sont modélisés en graphe afin de faire une reconnaissance automatique. Le laboratoire d'Informatique de l'université de Tours est une référence dans la modélisation des symboles en graphes [Luqman 10, Qureshi 06].

Une équipe chinoise a modélisé le visage humain en graphe pour proposer une méthode de reconnaissance faciale à travers l'appariement de graphes [Wan 09].

L'utilisation de l'appariement de graphes dans la géographie est aussi prometteur. L'appariement topologique nécessite avant tout de mettre la géométrie des entités géographiques sous forme de graphe. Cette opération dépend de la nature des données géographiques. Une route par exemple est modélisée par un arc et une intersection de routes est logiquement considérée comme un nœud d'un graphe. La forme géométrique d'un bâtiment peut se modéliser sous forme de sous-graphe [Dunkars 03]. Le polygone représentatif de l'empreinte d'un bâtiment au sol est modélisé par un arc correspondant à chaque coté et un nœud représentatif de chaque sommet de ce polygone.

L'appariement de graphes est utilisé en géographie pour retrouver des données en changeant d'échelle. Ce changement d'échelle affecte considérablement le visuel d'une couche de données. Thomas Devogele, dans sa thèse [Devogele 97] a proposé une méthode pour retrouver par exemple un nœud routier (carrefour, échangeur...) dans une couche de données où ce nœud routier est modélisé en tant qu'intersection simple entre deux arcs.

Ce type d'appariement permet aussi à Steffen Volz de coupler plusieurs types de données géographiques pour retrouver par exemple une route dans une autre couche de données géographiques [Volz 05]. Une route est soit appariée avec un chemin, soit avec une rue, soit avec une piste ou un passage.

1.3.3 Appariement de lignes

Une base de données géographiques vectorielles est composée de points, lignes et polygones. L'appariement de graphes permet de modéliser tous ces éléments.

La structure représentative d'une ligne est plus simple qu'une forme géométrique appartenant à une base de données géographique. De ce fait l'appariement entre deux lignes est une opération plus simple que celle de l'appariement de graphes.

Bouziani et Pouliot ont démontré l'efficacité de l'appariement géométrique d'ob-

jets géographiques en appariant ligne par ligne et segment par segment les données [Bouziani 04].

Une droite est caractérisée dans un repère par l'équation 1.1 à deux paramètres.

$$Y = aX + c \tag{1.1}$$

Si deux droites partagent les coefficients a et c alors ces deux droites sont identiques.

De manière analogue nous définissons la transformation de Hough au chapitre 3 qui permet de représenter une droite sous la forme 1.2 :

$$\rho = X\cos(\theta) + Y\sin(\theta) \tag{1.2}$$

Si deux droites dans l'espace de paramètre ont les mêmes coefficients ρ et θ alors ces droites sont semblables.

L'appariement de ligne est assez simple puisqu'il ne nécessite pas des notions mathématiques complexes, contrairement aux graphes. Il s'agit de comparaisons simples de valeurs numériques.

1.4 Conclusion

Tout au long de ce chapitre, nous avons présenté les aspects techniques et technologiques en rapport avec les travaux au cours de cette thèse.

Nous avons commencé par la présentation des techniques de mesure de localisation géographique et de positionnement. Ensuite la géomatique a fait l'objet d'une section qui insiste sur les données géographiques utilisées pour effectuer notre couplage entre les bases de données géographiques et photographiques.

Les limites des techniques de géolocalisation forment un argument de plus pour utiliser les SIG afin de se positionner plus exactement sur une carte. Cela fait du couplage des données photographique avec les données d'un SIG un nouvel outil de géolocalisation. C'est le but de notre travail et qui sera expliqué dans les chapitres suivants.

La section sur les traitement d'image s'est attardée sur les techniques de segmentation et de détection de lignes.

Nous avons présenté à la fin les différents aspects de la fusion de données en rapport avec notre travail de couplage entre les bases de données géographiques et photographiques.

Chapitre 2 : Couplage de bases de données photographiques et géographiques pour annoter les photographies

Chapitre 2

Couplage de bases de données photographiques et géographiques pour annoter les photographies

La démocratisation des appareils photos numériques a déclenché l'irruption de gigantesques banques de photographies [Attia 09a]. Ce phénomène touche pratiquement toutes les couches de la société.

L'apparition du Web a contribué à la mise en ligne de centaines de millions de photographies. Ces dernières se sont propagées d'abord sur des sites personnels puis, plus récemment, sur des plates-formes contributives et sur les réseaux sociaux. De nos jours, les utilisateurs publient, partagent et commentent leurs photographies sur le Web. Ils annotent ces photographies avec des informations sur les circonstances, le lieu, le contenu ou tout autre élément décrivant ce qu'ils prennent en photo.

La plupart des plates-formes Web de partage des photographies, comme par exemple Panoramio, Zooomr, Lo.ca.lise.us et Flickr, attachent une position géographique à chaque photographie [Torniai 07]. Les logiciels grand public de retouche et classement de photographies tels Picasa de Google ont rapidement proposé aux utilisateurs de localiser manuellement leurs photographies sur les globes virtuels de Google ou Yahoo par exemple. Le succès de ces solutions auprès des internautes, malgré le caractère fastidieux et approximatif des techniques de localisation manuelle, est bien le signe de l'intérêt que suscite actuellement la géolocalisation des photographies.

Annotations de photographies

Face à ces millions de photographies en ligne, le problème principal devient la gestion, le tri et la recherche dans ces bases de données qui se fondent sur la description du contenu de la photographie par le biais de mots clés ou de tags (étiquettes).

Une photographie est caractérisée par divers types d'informations affectées en tant qu'annotations [Liu 09]. Elle peut avoir comme étiquette la date de prise de vue, le nom de la personne, du monument ou de l'évènement pris en photo, la position de la prise de vue...

Le plus souvent, la classification des photographies se fait simplement par rapport à la date de prise de vue [Douglas 08].

La description du contenu d'une photographie est une tâche fastidieuse et très personnelle. L'étiquetage d'une photographie dépend du profil socio-culturel de la personne décrivant le contenu [Van Der Drift 09].

Des travaux récents [Jang 10] proposent une méthodologie pour regrouper les tags choisis par un groupe d'individu prenant en photo un même évènement ou une situation commune. Les tags d'une photographie sont affectés aux autres photos du même groupe.

Ce phénomène d'indexation des documents numériques par l'usager est connu sous le nom de folksonomie [Deuff 06, Jaschke 08]. Cette problématique ne cesse de prendre de l'ampleur avec l'avènement des nouvelles technologies du web qui donnent plus de possibilités d'expression à l'usager. Les internautes n'ont aucune règle ou repère commun pour rédiger les annotations.

Géotag

Une autre forme de classification des photographies a émergé au cours des dernières années, le positionnement géographique des photographies. Un tag qui contient une information géographique associée à la photo et informant sur le lieu de prise de vue, est appelé géotag.

Il faut distinguer deux positions parfois très distantes qu'en attribue à une photographie.

- La position de la prise de vue. C'est l'emplacement géographique de l'appareil photo qui a permis de faire la capture.
- La position de l'élément géographique ou du principal monument pris en photo.

Par exemple, la personne qui photographie la tour Montparnasse (figure 2.1) à Paris en se positionnant au sommet de la Tour Eiffel mettra vraisemblablement en géotag la tour Montparnasse car elle est l'objet pris en photo. Elle pourra aussi géotagger la photographie par l'endroit de prise de vue, la Tour Eiffel.

De multiples compagnies proposent des dispositifs matériels permettant de localiser automatiquement la photographie dès la prise de vue. Plusieurs types de solutions existent. Comme on l'a vu dans le chapitre 1, les appareils photographiques sont de plus en plus souvent équipés de GPS et de boussoles électroniques. Des applications sur Tablet PC [Donatis 06] permettent de cartographier, de décrire et d'attribuer de nombreux types d'annotations aux photos en plus de la position géographique.

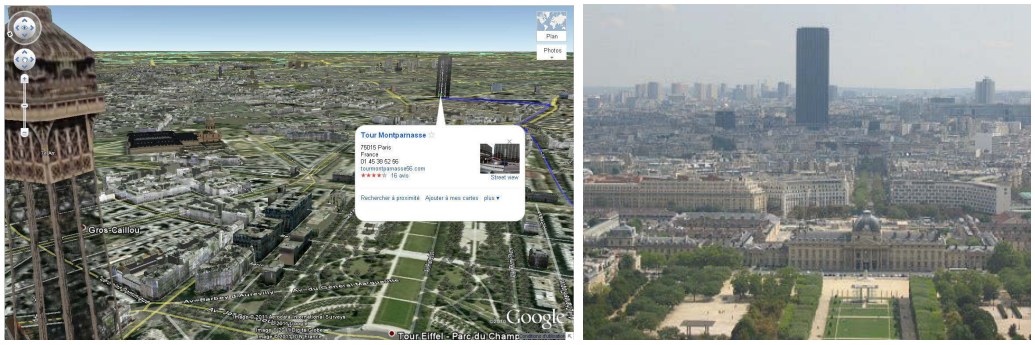


FIGURE 2.1 – La tour Montparnasse vue de la Tour Eiffel. (Sources : Google Earth et 123rf.com)

Formalisme

Le formalisme de ces métadonnées varie selon le domaine d'utilisation. L'imagerie médicale a adopté le format DICOM (Digital Imaging and Communications in Medicine) permettant d'ajouter les commentaires des médecins [Riesmeier 01]. D'autres domaines nécessitent une structure particulière de ces métadonnées [Thomas 05, De La Briere 08].

Un des formats standards normalisés est l'EXIF (Exchangeable Image File Format) qui permet d'ajouter des informations diversifiées [Douglas 08, Wu 09]. C'est le standard adopté par la majorité des constructeurs d'appareils photo.

2.1 Méthodes de couplage dans la littérature

Cette partie est organisée en deux temps. Après une analyse critique des propositions de solutions pour l'annotation existant dans la littérature, nous présentons la méthode et l'algorithme mis au point dans notre équipe de recherche pour extraire le contenu potentiel de la photographie.

Le degré d'implication de l'utilisateur dans la géolocalisation est un critère majeur dans la classification des méthodes de géolocalisation [Attia 09b]. On commence cette section par les méthodes semi-automatiques qui nécessiteraient l'intervention de l'usager dans l'opération de géolocalisation. Les méthodes totalement automatiques pour la géolocalisation sont présentées dans la seconde partie de cette section. Une géolocalisation automatique ne nécessite aucune intervention humaine pour ajouter une information géographique à une photographie.

2.1.1 Méthodes orientées utilisateur

De nombreux sites Web proposent d'ajouter des annotations informant sur le contenu d'une image [Sigurbjornsson 08]. Ces étiquettes informent également sur la position de la photographie en question. La position géographique est une information s'exprimant par un point sur une carte. Dans la plus part de ces sites la géolocalisation de photographies se fait en positionnant avec la souris un pointeur sur une carte indiquant la position de la prise de vue. Nous pouvons citer à titre d'exemple Panoramio et Loc.alise.us (figure 2.2a) qui s'appuient sur les cartes de Google, Flikr sur celles de Yahoo et Locr (figure 2.2b) qui positionne l'image sur les trois globes virtuels les plus connus, notamment Google Maps, Microsoft Bing Maps et Yahoo Maps. La géolocalisation des photographies avec cette méthode manuelle est assez fastidieuse en considérant le nombre toujours croissant de photographies. Plusieurs projets se sont donné comme objectif d'automatiser cette opération. La section suivante en cite quelques uns.



(a) Loc.alise.us (b) Locr

FIGURE 2.2 – Photos géolocalisées manuellement sur le site loc.alize.us et locr.

2.1.2 Méthodes automatiques

Une équipe de chercheurs à l'Université de Maryland a développé une méthode semi-automatique d'annotation d'images [Suh 07]. Elle profite d'une photo géolocalisée pour retrouver les personnes de cette photographie dans d'autres photographies. Elle affecte à toute personne portant les mêmes vêtements dans d'autres photos un même emplacement géographique.

En milieu rural, P. Ghadirian et I.D. Bishop ont fait une reconstitution 3D des images qu'ils prenaient à partir d'une caméra [Ghadirian 08]. Ils aboutissent à une vidéo d'une vue panoramique affichant la position de la caméra sur une carte. Sur la vidéo en sortie, l'information sur l'occupation du sol est ajoutée à l'image réelle du paysage.

James Hays et Alexei A. Efros ont essayé de trouver l'information sur la position géographique à partir d'une seule photo en la comparant avec tous les éléments d'une base de données d'images géolocalisées qu'ils ont construite [Hays 08]. Une association permet de conclure que la photo est prise au même endroit que ses semblables. La comparaison est optimisée en utilisant une combinaison de plusieurs descripteurs bas niveau. Ces descripteurs analysent principalement les textures, les couleurs et l'histogramme de la photographie.

Les photographies touristiques sont les seuls cas où cette technique est efficace, car elle nécessite la présence dans les photographies d'éléments emblématiques du paysage.

Dans un sens inverse, Keita Yaegashi et Keiji Yanai utilisent la position géographique de la prise de la photographie pour aider à détecter le contenu [Yaegashi 09]. Ils considèrent qu'en dehors de l'Afrique, la photographie d'un lion ne peut être prise que dans un zoo. Les prises de vue sur mer ne peuvent provenir que d'un bord de mer ou d'océan... Cette approche utilise une géolocalisation imprécise et très globale.

Le projet européen Tripod [Smart 09, Purves 10] alloue une grande importance à la génération automatique des labels (étiquettes) en se fondant sur la position géographique de la photographie. Le contenu de l'image est décrit par rapport à l'emplacement géographique de la prise de vue. Cette technique est connue sous le nom de géocodage inversé. Une ontologie est mise en place permettant de classer et d'enregistrer les types de scènes. Cette ontologie contient les relations, les qualités et les activités relatives à chaque élément de la scène prise en photo [Purves 08].

Géolocalisation et vidéos

Si l'on connaît la position géographique de prise de vue, les techniques de réalité augmentée rendent possible la superposition d'une photographie aux images réelles d'une vidéo. C'est une des solutions proposées pour recréer le monde urbain ou pour ajouter des détails supplémentaires. Notons que déjà en 1997, les applications en SIG et en urbanisme représentaient plus d'un tiers des applications de réalité augmentée [Haklay 02].

Le projet Snap2Play permet de jouer avec des mobiles équipés de GPS et de caméras en mode réalité augmentée. Le but du jeu est de retrouver en se déplaçant dans une zone l'emplacement des objets mal positionnés virtuellement [Chin 08].

L'ajout d'un simple texte sur une vidéo a fait l'objet des travaux au sein du LIRIS [Bres 09]. Ces travaux ont permis d'annoter les objets sur une vidéo prise par un opérateur mobile dans une zone urbaine. L'utilisateur doit simplement pointer et sélectionner avec la souris un objet (bâtiment, statut...) et le résultat de recherche du nom du bâtiment s'affiche sur la vidéo en permanence, même en déplaçant l'objectif. La texture des bâtiments est l'élément clé utilisé pour retrouver les bâtiments dans une base de données de textures. La technique employée repose sur la robustesse de

l'algorithme de vision par ordinateur SIFT. Une requête dans une base de données en ligne permet de fournir des données textuelles injectées dans la vidéo.

Les travaux du LIRIS automatisent deux processus différents. Le premier est la reconnaissance des bâtiments à partir de la base de données des bâtiments et l'injection de données textuelles à l'emplacement de chaque élément détecté. La seconde est le suivi continu au cours de la vidéo de ces éléments identifiés.

Dans leurs projet Enkin, deux jeunes chercheurs allemand ont proposé un dispositif incluant une webcam, une boussole numérique et un récepteur GPS [Loic 08, Braun 09]. Ces trois types de capteurs inter-connectés à un ordinateur permettent de donner des informations sur la distance entre l'opérateur et certains éléments du paysage urbain.

Ils ont réussi en 2007 à créer une application permettant d'obtenir trois types de cartes :

1. Map : une carte 2D qui ressemble à celle de Google Maps permettant de manipuler les données liés à la carte. Cette interface permet d'ajouter librement des tags ou des itinéraires sur la carte.
2. Landscape : une vue 3D de la carte avec les éléments ajoutés. Cette présentation est semblable à celle offerte par Google Earth. Les éléments ajoutés dans le mode Map sont tagués aussi en 3D.
3. Live : la vidéo capturée intègre les objets ajoutés en 3D. Le nom de chaque objet est affiché avec la distance qui le sépare du point de prise de vue. L'intégration des tags et des itinéraires ajoutés en amont dans la vidéo est assurée en temps réel.

Des chercheurs de Rennes [Colleu 08, Colleu 07, Sourimant 09] ont proposé des méthodes de localisation de l'appareil photo ou de la caméra en analysant la texture des bâtiments. Ces travaux ont cherché à trouver une position initiale de la caméra afin d'assurer une bonne cohérence de la réalité augmentée. L'approche utilisée pour cette initialisation de position se fait en utilisant la position approximative donnée par un GPS puis utilise l'algorithme RANSAC pour affiner ce positionnement. Cette pose initiale est calculée précisément en utilisant les angles entre les façades vues par l'observateur [Bioret 09].

De Boer a travaillé dans sa thèse à l'ajoût de labels sur des photos géoréférencées [Boer 08, Arnoud 07]. Son approche commence par la reconstruction 3D du paysage urbain. Les données de base sont l'Actueel Hoogtebestand Nederland (AHN) qui présente une valeur d'élévation tous les 5 mètres. À chaque bâtiment est attribuée une couleur unique. La seconde phase est la création de rendu dans ce monde virtuel en donnant à la caméra virtuelle les mêmes paramètres que la photo à laquelle on souhaite associer les labels. On obtient alors deux images analogues du point de vue du contenu : une photo du monde réel et une image de synthèse. L'identification des bâtiments sur l'image de rendu est quasi directe puisque chaque couleur représente un bâtiment unique. La dernière phase est l'ajout des labels sur la photo réelle dans

les mêmes emplacements que ceux de l'image de synthèse. La taille de la police des noms des bâtiments dans la photo de vue varie selon la distance par rapport au point de vue.

La reconstruction 3D des bâtiments est un moyen performant pour aider à la localisation. Les bases de données en trois dimensions incluant les textures ont fait l'objet d'un intérêt particulier. En partant d'un ensemble de photographies, [Taillandier 04] parvient à construire un modèle complet de 3D texturé. Un modèle plus fidèle à la réalité est enrichi par l'ajoût de la forme des toitures [Dornaika 08].

Dans sa thèse, [Moslah 11] a produit un modèle urbain capable de détecter les différents éléments des façades comme les fenêtres, les corniches et les balcons. Ce travail a déployé une grammaire descriptive des éléments constituant un paysage urbain.

Dans [Nogueira 09], un appareil mobile prenant en photo un paysage urbain par stéréovision est localisé en analysant ces photos. Un robot mobile est donc capable de se positionner en analysant les points d'intérêts des images stéréoscopiques.

Enfin l'analyse de texture de photographies standards permet aussi de se retrouver en ville en utilisant une base de données de signatures de textures géolocalisées [Suleiman 11].

2.1.3 Calculs de visibilité et photographie

En se positionnant dans la ville, un nombre limité de bâtiments est visible autour de notre position. En effectuant une prise de photo, le nombre diminue encore sauf si on prend une vue panoramique à 360°.

2.1.3.1 Lancer de rayons

Le lancer de rayons ou le ray tracing est reconnu comme la méthode la plus utilisée pour la génération de photographies réalistes par synthèse d'image [Caprani 00].

L'algorithme de lancer de rayons permet à l'origine de déterminer l'illumination (couleur et degré de luminosité) ainsi que la visibilité des différentes surfaces appartenant à la scène [Boyer 11].

Le principe de fonctionnement est simple [Caprani 00]. Pour chaque pixel de l'écran que l'on veut afficher, on trace des rayons imaginaires partant du point d'observation en direction de la scène. Si aucun objet n'est touché par un rayon, on affiche la couleur du fond d'écran à la position de ce pixel. Si on touche un objet, on calcule la couleur et la luminance par rapport à l'objet et aux sources de lumière. Dans le cas où cet objet touché est transparent ou réfléchissant, la procédure d'origine est relancée en reconsidérant le point de vue comme étant ce point d'intersection.

Cet algorithme récursif ne s'arrête que lorsque les rayons ne touchent plus d'objet.

Le ray casting, un sous-algorithme du ray tracing, permet de déterminer uniquement les objets visibles à partir d'un point. Il permet de tracer un polygone autour d'un point de vue. Ce polygone est appelé polygone de visibilité ou Isovist [Rana 06].

2.1.3.2 Notion d'Isovist

La notion d'Isovist est introduite par Benedikt et Davis en 1979 [Davis 79]. Ils définissent l'Isovist d'un point x comme l'ensemble de tous les points de son entourage qui sont visibles à partir de x.

La méthode de calcul de l'espace Isovist vérifie les segments entre le point d'origine x et les points y de l'espace P. Si la totalité du segment $[xy]$ est incluse dans l'espace P, alors son autre extrémité y fait partie de l'Isovist (voir formule 2.1).

$$\forall x, P = \Big\{ y | y \in P \; and \; xy \bigcap P = xy \Big\} \tag{2.1}$$

L'espace de visibilité 2D d'un observateur quand on l'envisage dans un environnement architectural ou urbain est très bien calculé par la technique de l'Isovist [Batty 01, Wimmer 99, Antani 11, Jiang 08, Leduc 11, Woloszyn 11].

Adaptation de l'Isovist pour la photographie

Une adaptation de la technique de détermination de visibilité (Isovist) permet de calculer l'espace visible à partir du point d'où une photographie a été prise.

Les segments testés se limitent à l'angle qui correspond à la focale de la photographie.

L'analyse du monde visible a fait l'objet de plusieurs recherches. Joliveau et Rana [Joliveau 08a, Joliveau 08b] ont proposé d'utiliser la notion d'Isovist pour étiqueter une photographie.

La figure 2.3 montre les Isovists de différents points de vue sur une partie du centre de Lyon. A chaque photographie est associé le nom des rues appartenant à l'Isovist [Joliveau 07].

C'est le résultat obtenu en utilisant l'outil informatique développé par Sanjay Rana sur la plateforme d'ESRI ArcView. Elle ne fonctionne qu'en 2D et dans un espace plat.

Une application de calcul du champ de visibilité 3D par un algorithme d'Isovist est développée par Suleiman [Suleiman 12].

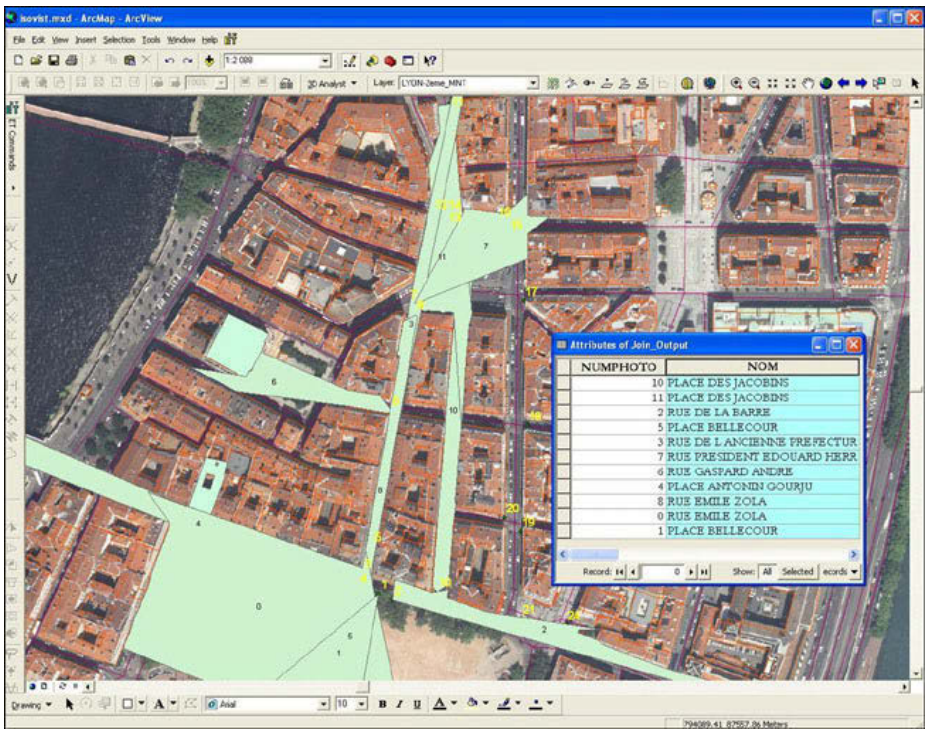


FIGURE 2.3 – Isovists dans le centre de Lyon. (Source : [Joliveau 07])

2.2 Méthodes utilisées dans le projet MapYourPicture

Les approches utilisées par les chercheurs pour coupler des données géographiques avec des données photographiques sont diversifiées. L'enrichissement de la photo en utilisant sa position géographique représente le choix effectué par notre équipe.

Cette section introduit puis détaille les investigations faites dans le projet MapYourPicture pour décrire le contenu d'une photographie en connaissant sa position et en utilisant les SIG. Nous présentons cette démarche en commençant par l'exposition des méthodes de visibilité pour analyser le monde visible à partir d'un point de vue. Ensuite nous présentons le travail qui est l'origine de notre nouvelle stratégie d'extraction de contenu géographique à l'aide des bases de données géographiques.

2.2.1 Approche de départ

Ce travail de thèse s'intègre dans le projet MapYourPicture financé par la région Rhône-Alpes dans un programme Cible. En préparation de ce projet une première phase avait été initiée par Benoît Delaup, un étudiant en master [Delaup 08]. Nous avons ensuite repris, amélioré et prolongé l'application qu'il avait proposée.

Cette sous-section présente cette approche ainsi que son analyse critique.

L'approche consiste à utiliser la connaissance de l'environnement de la photographie, au moyen d'une reconstitution virtuelle extraite d'une base de données géographiques. Cette reconstitution virtuelle est ensuite calée sur les paramètres de prise de vue de la photographie et confrontée à celle ci. Le rendu créé à partir de cet environnement 3D sert d'intermédiaire pour l'extraction du contenu d'une photo.

L'approche emprunte cinq étapes :

1. **la préparation des données de type 2,5D.**

   Les bâtiments qui sont stockés au format vecteur sont posés sur le sol puis subissent une élévation suivant leurs hauteur (extrusion).

2. **l'indexation des données.**

   La base de données, comme déjà souligné, contient les bâtiments. Les bâtiments présents au sol sont schématisés en tant que polyèdres sur un plan.

   On propose de décomposer les murs des bâtiments en segments élémentaires qui correspondent aux limites de l'environnement.

   L'élément de base devient alors le mur. Chaque bâtiment est indexé par les murs qui le limitent, comme les rues et tous les types d'espaces ouverts. Deux raisons poussent à effectuer ce choix :
   - il est impossible d'apercevoir tous les côtés d'un bâtiment en se positionnant au niveau du sol. Habituellement, on est capable d'observer un, deux voire trois côtés (murs) d'un bâtiment.
   - un espace ouvert (parc, place, rue...) est représenté aussi par un polyèdre. Il n'est pas représenté par un objet dans la base de données. Comme il est limité physiquement par des bâtiments, il sera donc défini par les murs des bâtiments qui l'entourent.

   On attribue à chaque mur la hauteur du bâtiment auquel il appartient.

   La méthode d'indexation des routes, exemple d'espace ouvert, et des bâtiments est détaillée dans la figure 2.4.

3. **la création du modèle 3D.**

   Le modèle 3D du paysage urbain est construit pour faire une capture des objets visibles dans le monde virtuel.

   Les paramètres photographiques (x, y, z, azimut et angle d'ouverture) sont chargés dans une caméra virtuelle pour obtenir un rendu identique à la photographie prise avec ces paramètres.

   Le modèle 3D est reconstruit avec un niveau de détail LoD1 (voir section 3.3.1). En effet, il ne contient ni textures ni formes spécifiques de toitures [Kolbe 05].

4. **Association.**

   Les photographies et les objets de la base de données (bâtiments, rues, places...) sont associés. Les couples bâtiment/photographie et rue/photographie sont créés. Ces couples sont enregistrés dans une base de données.

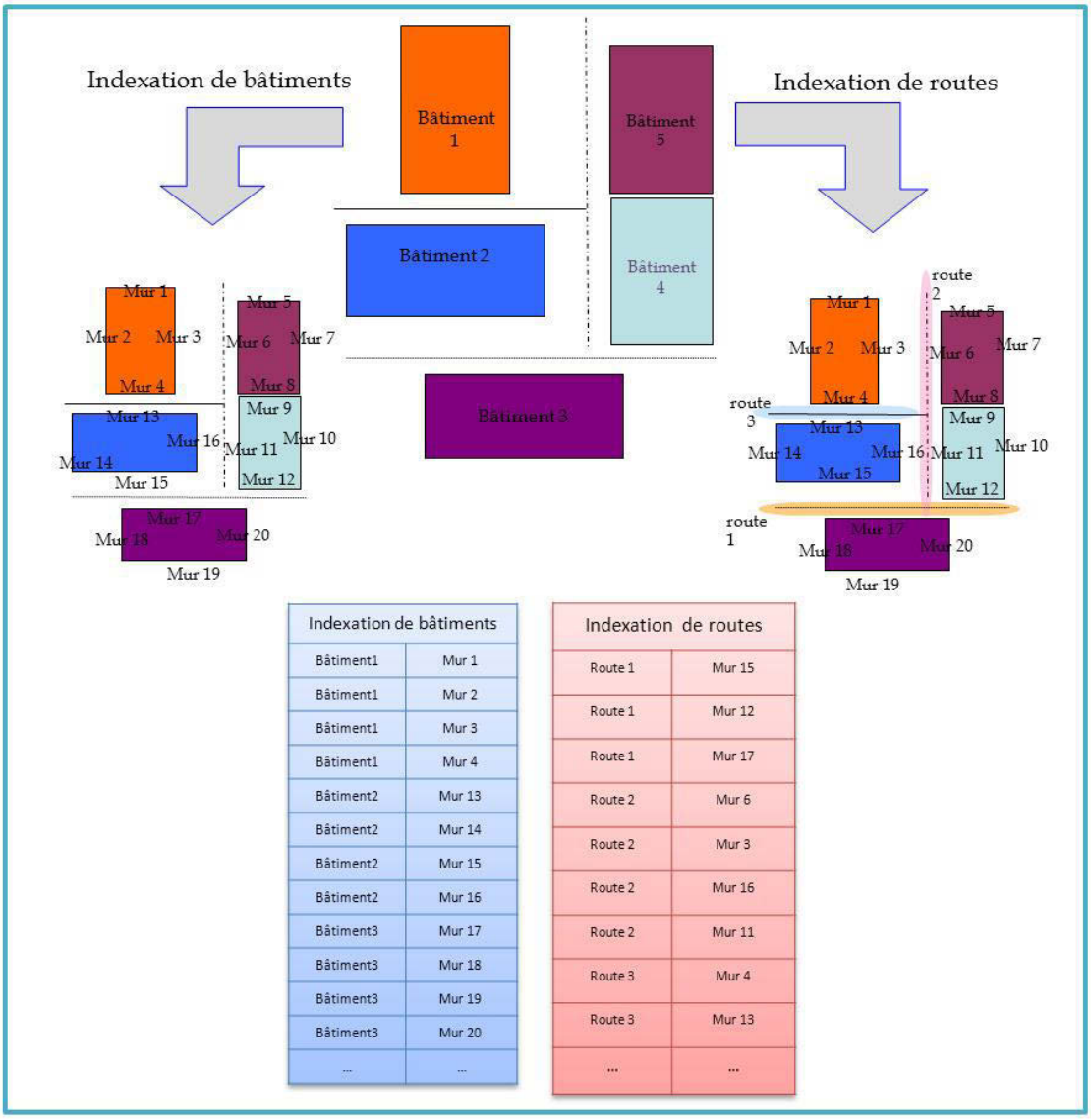


FIGURE 2.4 – Indexation des bâtiments et des rues.

Un coefficient de présence est affecté à chaque bâtiment visible. Ce coefficient est donné par l'équation 2.2.

$$Coefficient\ de\ présence = \frac{Surface\ projetée}{Distance} \tag{2.2}$$

La surface projetée correspond au nombre de pixels que représente l'objet sur la photo.

La distance est celle qui sépare le point de prise de vue photographique d'un bâtiment (le point le plus proche).

5. Interrogation de la base de données pour déterminer les éléments contenus dans une photographie.

   Un filtrage est effectué grâce à un seuillage du coefficient de présence pour sélectionner les informations les plus pertinentes à placer dans les tags et éventuellement dans le fichier EXIF.

2.2.2 Description du démonstrateur

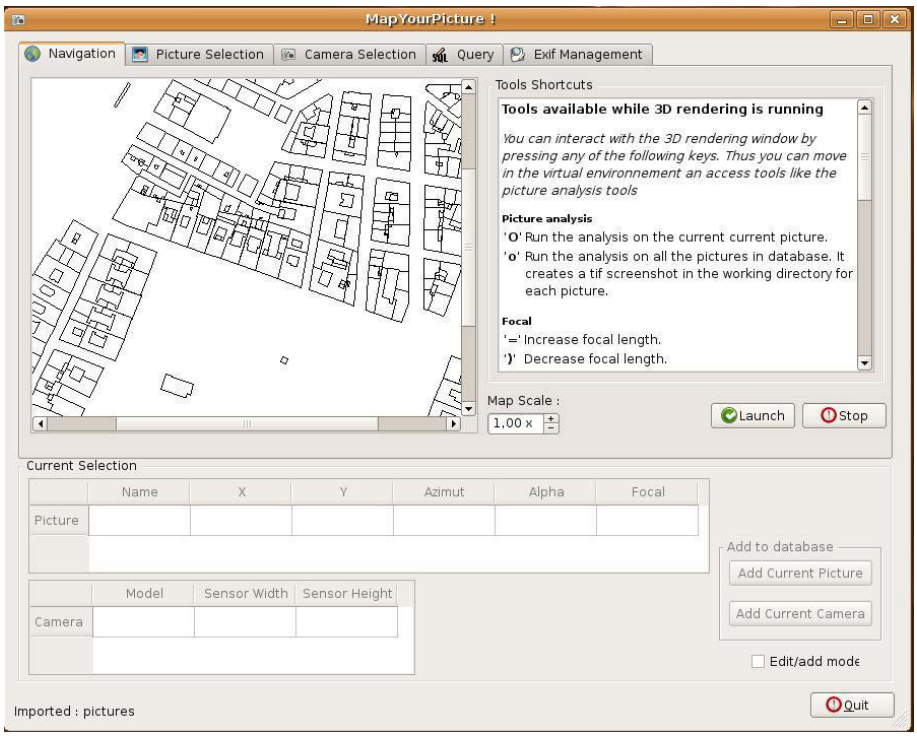


FIGURE 2.5 – Fenêtre principale du démonstrateur.

Le démonstrateur de cette approche est programmé en C++. La bibliothèque OpenGL a permis de faire les traitements 2D et 3D. Le système de gestion de bases de données utilisé est PostgreSQL.

L'application (voir figure 2.5) assure la liste des fonctionnalités suivantes :

- 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 EXIF.
- estimation du coefficient de présence de chaque objet identifié sur la photo.

Description des données utilisées

L'appareil photographique utilisé est un réflex Olympus E-500 couplé avec une boussole classique et placé sur un trépied. Celui-ci indique l'angle de l'axe visé par rapport à la surface sur laquelle il est posé.

Les bâtiments sont fournis par une ancienne version de la couche Bâti de la BD TOPO de l'IGN. C'est une couche vecteur qui contient l'occupation au sol des bâtiments.

Un bâtiment est considéré s'il mesure plus de 50 m^2. Chaque bâtiment possède une hauteur minimale et une hauteur maximale (voir figure 2.6).

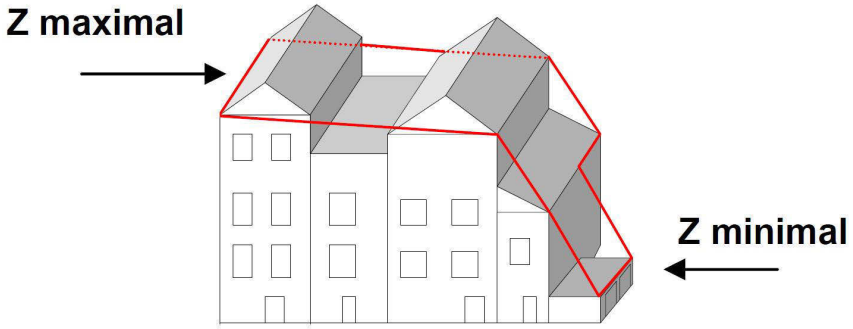


FIGURE 2.6 – Les hauteurs minimales et maximales relatives à un seul bâtiment selon les cartes BD TOPO de l'IGN. (Source : [IGN 08b])

Précisons qu'une hauteur maximale est concrètement la hauteur maximale du pourtour vu du ciel. Il représente alors la hauteur au niveau des gouttières et non pas la hauteur absolue du sommet du bâtiment.

Les bâtiments contigus, superposés qui ont une même fonction, sont considérés comme une seule entité. Le Z min et Z max d'un bâtiment simple sont égaux.

Pour créer les rendus 3D, le choix est fait arbitrairement sur la valeur minimale de hauteur.

2.2.3 Résultats et évaluation

Cette proposition d'extraction de données des photographies avec une reconstitution en 3D de l'environnement a démontré son intérêt.

Le fait d'avoir deux photos comparables, une prise avec un appareil photo et une vue de synthèse générée par rendu, nous permet d'extraire un grand nombre d'informations. On a accès à la liste des bâtiments et des rues visibles, aux distances de chaque objet par rapport au point de prise de vue et à l'importance de chaque objet dans la photographie.

La méthode proposée permet aussi de géolocaliser des photographies qui ne le sont pas précisément, pourvu que l'on sache approximativement où elles ont été prises. En effet, l'application offre la possibilité de naviguer dans l'environnement 3D jusqu'à avoir un rendu analogue à celui d'une photo déterminée. Il est donc possible de localiser celle ci en lui affectant la position et les paramètres de la camera virtuelle puis en les stockant dans son fichier EXIF.

Évaluation

Dans la majorité des cas, la caméra est correctement placée et paramétrée. Cependant plusieurs problèmes apparaissent. La figure 2.7 illustre deux photographies et deux rendus 3D. Le relief n'est pas pris en compte dans le rendu. De ce fait, la pente n'apparait pas identique dans les deux cas de figure (photographie et image de synthèse). C'est le premier point à améliorer et qui sera présenté dans le chapitre suivant..

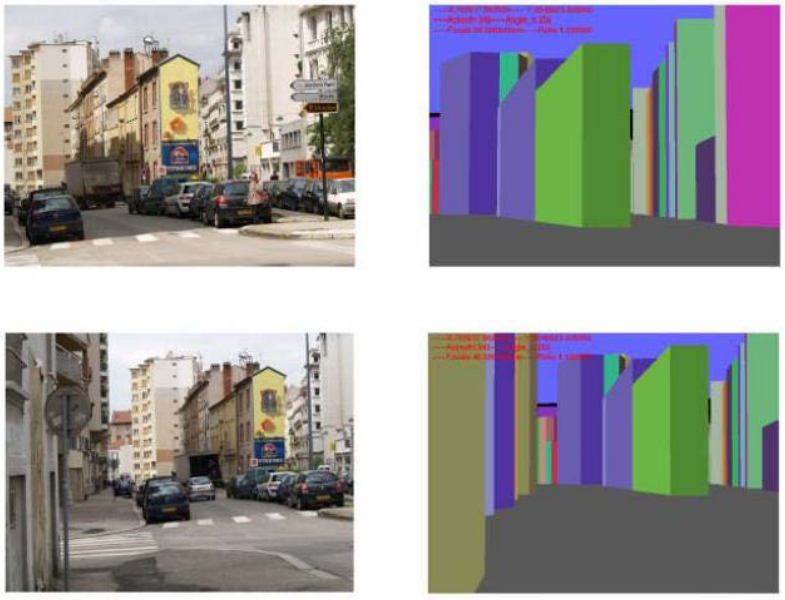


FIGURE 2.7 – Photographie et rendu des photos.

Par ailleurs, les erreurs de mesures font qu'un recalage d'image est parfois nécessaire. Cela est dû aux incertitudes des coordonnées GPS constatées (jusqu'à 8 m d'erreur de localisation) et des mesures des autres appareils utilisés. L'erreur maximale constatée pour la mesure d'orientation par la boussole analogique utilisée est de 14°.

Un exemple complet est présenté dans la figure 2.8. La photo originale prise par l'appareil photo est placée dans le coin haut à gauche. En dessous sont schématisées par deux demi-droites sa position de prise de vue sur la carte, sa direction de visée

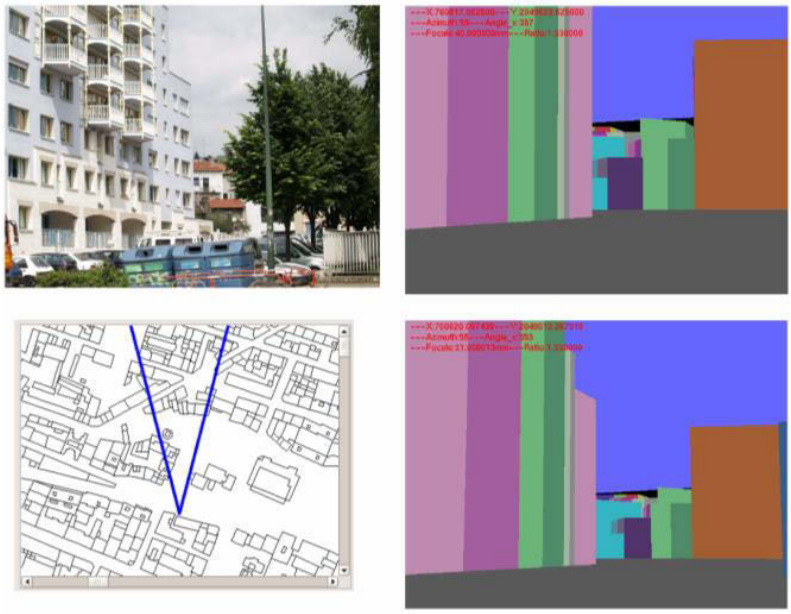


FIGURE 2.8 – Résultat d'extraction du contenu d'une photo.

et sa focale. A droite en haut est affiché le rendu généré, qu'il a fallu réajuster manuellement pour le rendre encore plus proche de la photo initiale. Le rendu recalé qui correspond plus à la réalité est affiché dans le coin bas à droite de la figure.

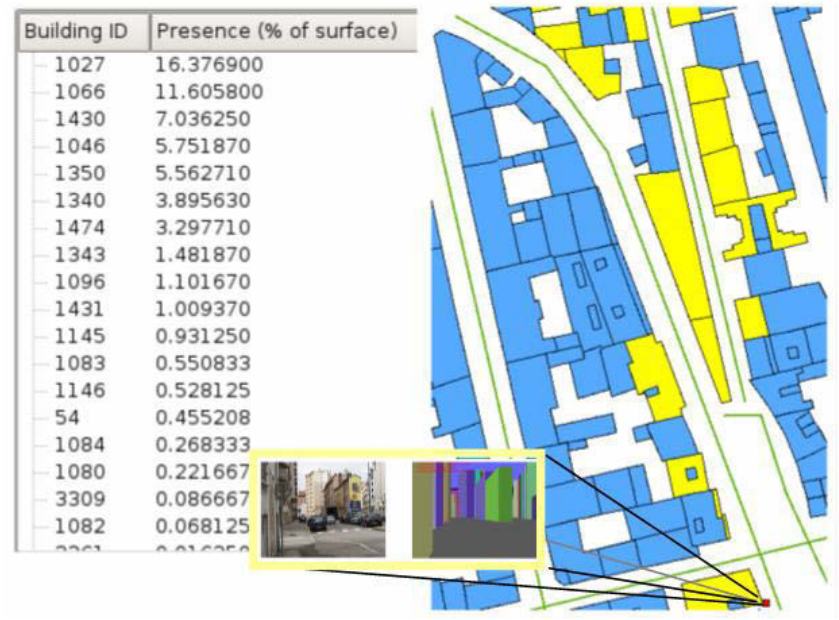


FIGURE 2.9 – Bâtiments visibles à partir du point de vue.

La figure 2.9 présente le pourcentage de surface de chaque bâtiment identifié sur

cette photographie.

Le point rouge en bas de la figure localise la position du point de prise de vue.

Conclusions

Nous pouvons conclure que les incertitudes ne constituent pas un obstacle absolu à l'extraction du contenu d'une photographie. En effet, les bâtiments détectés sur une photo demeurent les mêmes avant et après rectification. Seul leur ordre de classement selon leurs niveaux de présence dans la photo est affecté par les incertitudes.

Le coefficient de présence est une réponse simple à une question complexe. Il est en effet difficile de trouver une clé d'entrée permettant de deviner ce qui intéresse l'observateur. Est-ce que ce sont les objets les plus présents dans l'image ou les objets les plus proches de la photographie ? Il serait possible de faire intervenir dans la formule la taille réelle de l'objet, ce qui permettrait d'avoir une idée de son niveau de détail. Une recherche de méthodes de validation formelle s'impose, qui pourrait s'appuyer sur des techniques sensorielles de suivi de regard afin de valider expérimentalement, ce qui intéresse les utilisateurs dans les photographies qu'ils recherchent. Il s'agirait d'appliquer à des photographies urbaines ce qui a été déjà fait pour des paysages [Ghadirian 08].

2.3 Conclusion

Ce chapitre a présenté les différentes logiques de passage du monde géographique au monde photographique à travers les méthodes utilisées au sein de notre équipe de travail et dans la littérature.

Nous avons détaillé l'approche basée sur la reconstruction virtuelle du paysage urbain pour définir le contenu photographique. Nous avons surtout insisté sur les lacunes et imperfections de la méthode auxquelles il faudra remédier pour aboutir à de meilleurs résultats.

CityGML ou Collada offrent des informations sur les formes des bâtiments ainsi que sur la texture. Ces types de sources de données 3D plus détaillées donnent plus de précision au résultat et rendent le rendu plus proche de la réalité observée. Cependant les informations de ce type ne sont disponibles que pour certaines grandes villes dans le monde.

Il est aussi nécessaire d'accroître la précision des données de localisation.

Une première solution est d'adopter un équipement GPS plus puissant. Les GPS différentiels seraient une solution envisageable mais lourde à mettre en œuvre et s'avèrent mal adaptés à une utilisation courante de la photographie.

Une deuxième solution serait de mettre en œuvre un algorithme de vision, afin de détecter les contours des objets et d'effectuer ensuite la correspondance entre les

formes dans la photographie et dans le rendu 3D. Ce rendu représenterait exactement le même contenu que la photographie.

La solution imaginée et mise au point dans cette thèse est d'apparier les lignes repérables dans la photo réelle et celles présentes dans le rendu 3D.

Chapitre 3 : Localisation de photographies par appariement de lignes

Chapitre 3

Localisation de photographies par appariement de lignes

Nous traitons en détail dans ce chapitre la démarche que nous proposons pour l'aide à la localisation de photographies.

Le chapitre précédent a permis de présenter les techniques développées dans notre laboratoire. Les différentes limites du démonstrateur présenté précédemment ont amorcé une nouvelle réflexion pour mieux coupler des bases de données géographiques avec des bases de données photographiques.

Nous avons décidé d'opter pour une approche basée sur le traitement d'image afin d'optimiser l'extraction de contenu photographique.

La procédure proposée au cours de notre travail de thèse ne représente pas une rupture avec l'existant mais une continuation et un approfondissement de la méthode déjà engagée par l'équipe.

Ce chapitre détaille la démarche en justifiant les choix technologiques adoptés et les techniques utilisées. La présentation de tous les traitements et données manipulées est détaillée dans les sous-sections suivantes.

Nous présentons enfin l'algorithme que nous proposons pour interfacer deux images par appariement de lignes.

3.1 Présentation globale

Une photographie prise par un appareil photo en milieu urbain possède une position géographique. Cette position est donnée par un GPS, par un autre système de positionnement ou bien affectée manuellement.

Une photographie contient des objets du paysage urbain. Notre objectif est de retrouver la localisation de prise de vue en comparant le contenu de la photographie avec celui de l'image de synthèse.

Pour ce faire, on utilise une base de données SIG qui contient les bâtiments du milieu urbain, l'élévation du sol au pied de chaque bâtiment et sa hauteur. Toutes ces données permettent de créer un monde virtuel représentatif du monde réel.

Dans un monde virtuel du milieu urbain généré par la synthèse d'image, la navigation permet de positionner une caméra dans un emplacement quelconque.

Nous proposons d'utiliser les paramètres de positionnement de la photographie prise pour les injecter dans les coordonnées de la caméra virtuelle. La position, comme indiquée dans les chapitres 1 et 2, est loin d'être précise en milieu urbain. Nous sommes confrontés au même problème d'incertitude qu'avec la boussole numérique.

Nous allons chercher à retrouver la position et l'orientation exactes de notre photographie en nous positionnant dans le monde virtuel de telle façon qu'on ait le même contenu d'image. C'est donc le contenu de l'image qui permet de déduire la localisation.

Un contenu identique entre une photographie et une photo de synthèse n'est possible que si les deux images partagent la même position, la même focale et la même direction de visée. C'est l'apport de notre travail : retrouver la position avec une précision accrue dans un milieu urbain en utilisant la vision par ordinateur.

La démarche que nous proposons permet de retrouver l'image de synthèse qui correspond à une photo réelle au moyen d'un appariement de lignes entre deux images.

Deux images appariées partageront donc la même position géographique, la même direction de visée, la même distance focale et par conséquent le même contenu.

Dans la pratique, nous pouvons considérer cinq phases dans notre procédure.

1. Création du monde virtuel en se basant sur une couche de données géographiques contenant les bâtiments. Les données utilisées feront l'objet d'une présentation détaillée dans le chapitre suivant.
2. Sélection d'images de synthèse aux alentours de la position présumée de la photographie réelle, objet de l'appariement. Dans une zone centrée sur la position géographique de la photographie dont on souhaite décrire le contenu, les images de synthèse sont réalisées idéalement avec un intervalle régulier de 2 mètres. Le choix de cet intervalle est le résultat d'une expérimentation par la pratique.
3. Préparation de la photographie en éliminant le sol.
4. Essai d'appariement de la photographie traitée avec une des images de synthèse sélectionnées à la phase 2.
5. Affectation à la photographie des paramètres de positionnement des images de synthèse avec qui elle s'apparie le mieux.

Notre approche d'appariement pour fusionner les données géographiques avec les photographies est présentée en détail dans la figure 3.1. Les opérations indiquées

dans cette figure sont détaillées dans les sections suivantes.

Le principe est de détecter des lignes dans les deux images. Ensuite nous cherchons à retrouver les lignes détectées dans la photographie parmi les lignes détectées dans la photo de synthèse.

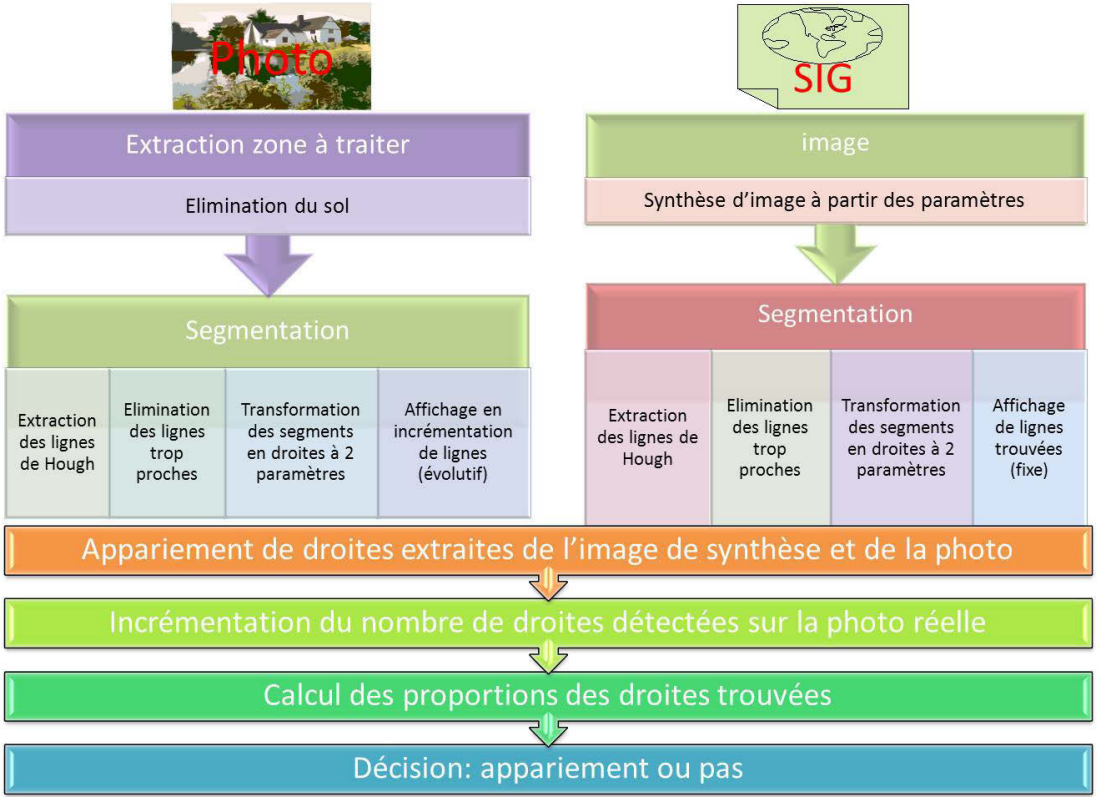


FIGURE 3.1 – Méthode globale d'interfaçage par appariement de lignes.

3.2 La démarche d'interfaçage par appariement de lignes

L'appariement des images consiste à passer d'une comparaison entre deux images à une comparaison entre deux listes de lignes. C'est une correspondance (matching) entre les lignes détectées dans une image de synthèse et les lignes détectées dans une photographie.

3.2.1 Présentation des données

Une ville contient principalement des bâtiments. Il est donc très naturel et logique de chercher ces bâtiments dans les images urbaines. La quasi-totalité des frontières

des bâtiments sont des droites. Ce sont des lignes qui délimitent les bâtiments, les toits, les bordures des routes...

Nous avons choisi d'utiliser les lignes qui définissent parfaitement les bâtiments dans une ville et de les apparier avec celles qui apparaissent sur les images de synthèse.

La transformation de Hough

Paul V. C. Hough a proposé en 1962 une méthode qui permet de détecter dans une image les formes paramétriques comme les droites et les cercles [Hough 62, Maître 85]. Cette technique [Valentin 09] a été revue par d'autres chercheurs [Duda 72, Rosenfeld 69] dix ans plus tard pour passer de la détection d'alignements sur un oscilloscope à une méthode de détection de lignes. La transformation de Hough est une technique "optimale" [Crowley 10] pour détecter les lignes spécialement dans des images très bruités.

Cette technique ne demande pas une continuité des droites pour les détecter. Elle permet par exemple de trouver les bords de routes avec un éclairage minimum [Kneepkens 05]. Cependant, les éléments détectés par cette transformation ne sont pas des segments, mais des droites.

Cette transformation permet de passer d'un point sur une image à une ligne droite dans un espace de paramètres [Duda 72]. L'espace de paramètres est défini par les paramètres de représentation des lignes dans le plan de l'image. La description de cette transformation est l'objet du paragraphe suivant.

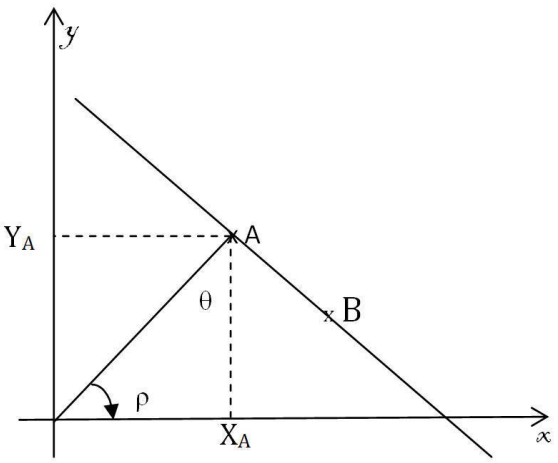


FIGURE 3.2 – Représentation d'un point.

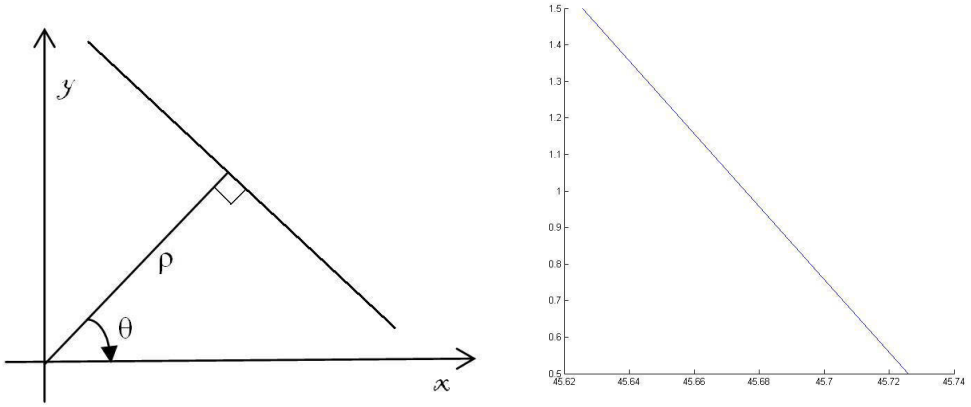
Notons que d'après les expérimentations, une des meilleures méthodes de détection de lignes, pour une image bruitée [Cohen 77] est la transformée de Hough.

Un autre argument d'utilisation des droites de Hough est lié à la représentation

des lignes. En effet chaque ligne détectée est caractérisée parfaitement par seulement deux paramètres numériques :

- Une distance ρ : c'est la distance exprimée en pixels qui sépare la ligne de l'origine du repère. La valeur maximale que peut prendre ρ est celle de la diagonale de l'image.
- Une inclinaison θ : elle est exprimée en radians et varie de $-\Pi$ à Π.

Ces deux paramètres sont schématisés dans la figure 3.3. L'exemple donné sur cette figure est le rendu obtenu par Matlab pour le traçage d'une ligne avec les paramètres $\rho = 45$ et $\theta = 0.1$.



(a) Paramètres ρ et θ

(b) Droite avec les paramètres $\rho = 45$ et $\theta = 0.1$

FIGURE 3.3 – Représentation des lignes.

Principe de la transformée de Hough

On considère A et B deux points d'une droite (voir figure 3.2). Ces points vérifient les équations 3.1.

$$\begin{cases} Y_A = m \times X_A + p \\ Y_B = m \times X_B + p \end{cases} \; avec\, m = \frac{(Y_B - Y_A)}{(X_B - X_A)} \; et\, p = Y_A - m \times X_A \tag{3.1}$$

La droite qui passe par ces deux point est régie alors par l'équation 3.2.

$$Y = mX + p \tag{3.2}$$

On appelle m et p paramètres cartésiens d'une droite. Cependant, la largeur de l'intervalle de valeurs que peut prendre ces deux paramètres varie de $-\infty$ *à* $+\infty$,

large écart qui n'est pas souhaitable pour les mathématiciens. De même les droites très verticales ont des coefficients directeurs très élevés alors que ces droites ne sont pas particulières. On utilise alors un paramétrage dit normal [Duda 72, Nozick 10]. Il s'agit des paramètres polaires qui dériveront les droites selon l'équation 3.3.

$$\rho = X\cos(\theta) + Y\sin(\theta) \; tel \; que \; \begin{cases} \rho = \frac{|X_A Y_B - X_B Y_A|}{\sqrt{(Y_B - Y_A)^2 + (X_B - X_A)^2}} \\ \theta = -\arctan(\frac{X_B - X_A}{Y_B - Y_A}) \end{cases} \tag{3.3}$$

Nous constatons alors que ρ prendrait des valeurs allant de 0 à $+\infty$ et que θ serai comprise entre $-\Pi$ et Π (figure 3.2).

En faisant varier θ entre $-\Pi$ et Π, on obtient une sinusoïde. Chaque point de cette sinusoïde possède les paramètres d'une des droites parmi l'infinité de celles qui passent par ce point.

La démarche se résume par les quatre propriétés suivantes [Duda 72] :

1. D'une manière générale pour chaque point, on peut tracer une sinusoïde dans le plan des paramètres.
2. Un point se situant dans le plan des paramètres représente par bijection une droite dans le plan de l'image.
3. Les points alignés dans l'image correspondent à des courbes qui ont un point en commun dans le plan des paramètres. Cette propriété est facilement démontrée par le fait que la droite qui passe par des points alignés possède les mêmes paramètres.
4. Les points d'intersection des courbes dans le plan des paramètres correspondent aux lignes relatives à ces points dans l'image.

Un exemple d'exécution de la détection de lignes par la transformée de Hough est illustré dans la figure 3.4.

Algorithme

L'algorithme global qui traduit le processus de détection de lignes par la transformation de Hough se présente comme suit :

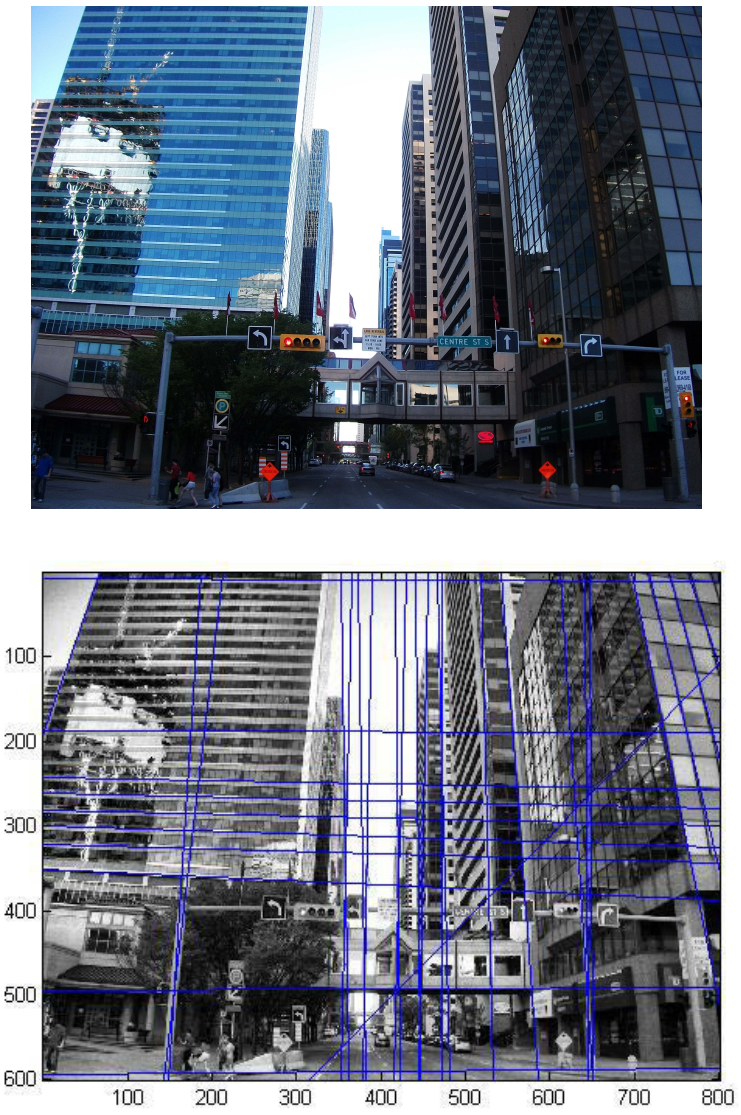


FIGURE 3.4 – Détection de lignes par la la transformation de Hough.

Input (image I, seuil S)
Détecter les contours C dans l'image I
Pour chaque point A du contour C faire :
    Incrémenter l'accumulateur Acc de la droite passant par A
    Si $Acc \geq$ seuil S alors faire :
        Reconstruire la droite D qui correspond à A
        Mette Acc à zéro
    Fin si
Output(tableau de lignes T)

Cette méthode de détection de lignes a fait ses preuves pour de nombreux types d'images. Les traits imaginaires présentés à la figure 1.9 sont détectables par la transformation de Hough. Le résultat de la détection des droites de Hough avec un seuil très bas est schématisé par la figure 3.5.

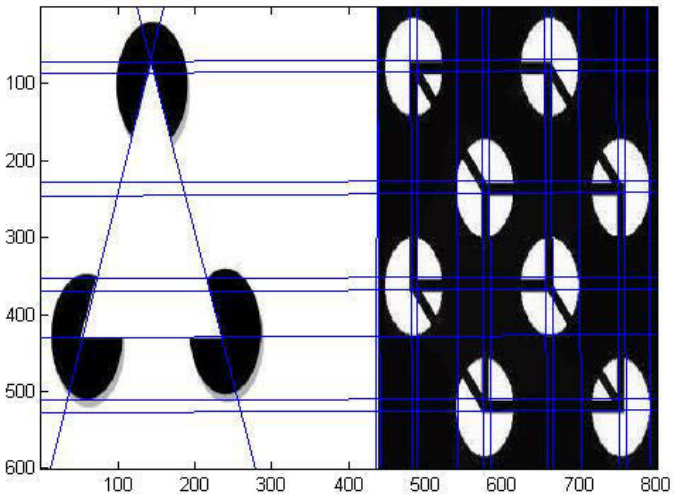


FIGURE 3.5 – Lignes imaginaires détectés par la transformée de Hough.

Généralisation et optimisation

La transformée de Hough ne se limite pas uniquement à la détection des droites. Elle se généralise pour toute forme géométrique ayant des paramètres représentatifs. Les cercles par exemple possèdent la représentation paramétrique de la formule 3.4.

$$(x-a)^2 + (y-a)^2 = r^2 \tag{3.4}$$

Chaque point (x,y) de l'image est l'origine d'un cône dans l'espace des paramètres (a, b,r). Si on fixe r (rayon) on obtient un cercle.

En ajoutant d'autres paramètres, la transformée de Hough permet la détection des droites en 3D [Bhattacharya 00].

L'élimination ou la réduction du bruit par l'un des filtres détecteurs de contraste et donc de contours favorise la perfection de la détection des droites par la transformée de Hough [Horiki 98]. Cette approche de détection de lignes est confrontée tout de même à quelques problèmes dont le principal est l'apparition de doublons de

droites. Ce cas de figure se présente souvent dans le cas où une ligne dans l'image est large ou encore plus fréquemment quand les points ne sont pas parfaitement alignés [Nozick 10].

Des techniques ont été proposées pour trouver le bon candidat parmi les lignes détectées. On retrouve dans la littérature plusieurs améliorations et adaptation selon les objets étudiés par la transformée de Hough [Maître 85, Aggarwal 06].

3.2.2 Déroulement de la démarche

Pour détecter les lignes se trouvant dans l'image de synthèse, nous affectons au seuil de détection de lignes de Hough une valeur permettant de détecter un nombre maximal de lignes tout en évitant de détecter les marches d'escalier.

Les lignes trouvées sur l'image de synthèse délimitent uniquement les bâtiments. Cela est dû au fait que les bâtiments ne possèdent pas de textures et sont les seuls objets présents dans le monde virtuel.

On cherche ensuite le nombre de lignes en commun. Un appariement est constaté entre deux lignes si ces deux lignes partagent une même distance ρ et un même angle θ.

Concrètement, ceci est traduit par une distance ρ qui appartient à l'intervalle centré sur la valeur ρ de la ligne dans l'image de synthèse. La largeur de cet intervalle est appelée l'intervalle de confiance.

Le degré d'inclinaison θ possède aussi son intervalle de confiance. Le choix de ces valeurs est expérimenté dans le chapitre suivant.

Si une ligne A détectée sur une image de synthèse possède une distance ρ et un angle d'inclinaison θ appartenant tous les deux aux intervalles de confiance respectifs d'une droite B détectée sur une photographie, alors la droite A est considérée comme identique à la droite B.

Chaque ligne détectée sur l'image de synthèse est recherchée dans la photographie réelle.

Considérons que les droites détectées dans les images de synthèse sont des droites de type S. Les droites détectées par transformation de Hough sur les photographies seront appelées droites de type P.

Le nombre de droites de type S est fixe et reste inchangé durant tout le processus d'interfaçage.

Le traitement commence par la détection d'un minimum de lignes de type P. Ensuite on cherche les lignes de type S parmi les lignes de type P.

Nous diminuons ensuite le seuil de détection de lignes de type P pour détecter plus de lignes. La recherche des lignes S parmi les lignes P est lancée à nouveau.

Si on retrouve une certaine portion des lignes S parmi les lignes P, et que le nombre de lignes P détectées est inférieur au triple de lignes S, alors on considère

que les deux images partagent le même contenu photographique.

Si on détecte le triple du nombre des lignes de type S dans la photographie sans retrouver un nombre considérable de lignes correspondantes, on estime que l'appariement n'a pas lieu entre la photographie et l'image de synthèse.

Rappelons que le nombre de lignes détectées dans les images de synthèse est largement inférieur à ceux détectés dans les photographies à cause des effets de textures, d'obstacles...

Le choix du triple de valeurs de lignes de type S est donné par expérimentation sur les photographies urbaines. En effet les 15 à 45 lignes détectés sur les photographies permettent de bien décrire les bâtiments et une partie des textures et des obstacles.

3.2.3 Algorithme

La démarche décrite dans le paragraphe précédant peut être présentée sous forme de l'algorithme généralisé suivant :

Input(Image S ,image P)
Transformation de l'image S et P en niveaux de gris
Détecter LS lignes de l'image S
Filtrer les lignes trop rapprochées parmi DS
Tant que la condition C n'est pas vérifiée faire

Détecter LP lignes de l'image P avec le seuil SP
Filtrer les lignes trop rapprochées parmi LP
Pour chaque ligne des LP

similitude = 0
Pour chaque ligne de LS

Si ($\rho P \approx \rho S$ et $\theta P \approx \theta S$) alors

similitude = similitude +1

Fin si

Fin Pour

Fin Pour
Si (similitude $\geq LP \times 4/5$ et $LP \leq 3 \times LS$) alors

C = vrai
Déclarer appariement parfait

Fin si

Si ($LP \geq 3 \times LS$) alors

C = vrai

Fin si
Seuil SP = SP - 0.01

Fin tant que
Output(Lignes en communs *L*, *similitude*)

3.3 Choix technologiques et méthodologies

Comme on l'a expliqué au premier chapitre, ce travail de thèse fait intervenir trois principales disciplines a priori distinctes.

La première discipline est la géomatique car les bases de données géographiques sont la source des données pour identifier le contenu photographique.

Le second domaine est la vision par ordinateur. Nous analysons les images pour parvenir à les localiser avec précision. Nous nous intéressons en particulier aux lignes extraites de ces images.

La fusion entre les deux types de données issues des deux premières disciplines est la troisième branche. Il est nécessaire d'adapter les données d'un SIG pour les apparier avec une photographie.

Les outils SIG sont divers et permettent de gérer les données géographiques avec une panoplie de fonctionnalités toujours plus large. Ce qui est valable pour les SIG l'est aussi bien pour le traitement d'image.

Cette section propose une argumentation sur les choix effectués pour la création du rendu de la scène urbaine. Ensuite nous présentons la plateforme Matlab qui a permis d'effectuer l'appariement de lignes.

3.3.1 Création de rendu

Une image de synthèse est définie comme "l'affichage, sur un support physique, d'un ensemble numérisé de points lumineux (pixels) engendrés par le modèle mathématique des objets de la scène, éclairés à l'aide d'un modèle mathématique de la lumière et observés par l'intermédiaire du modèle mathématique de l'optique d'une caméra." [Porchet 02].

La synthèse d'image se définit simplement par l'utilisation d'un ordinateur pour générer des images [Boyer 06].

Les trois phases de la synthèse d'images sont :

1. La modélisation : les objets sont représentés mathématiquement pour être placés sur la scène. Dans notre paysage urbain, les objets sont les bâtiments. Les données sont enregistrées dans des bases de données géographiques. Elles sont déjà modélisées dans un shapefile (fichier de formes) . C'est un des standards le plus utilisé par les plateformes SIG. Plusieurs librairies propres à des langages de programmation différents sont disponibles [LeSage 04] pour manipuler ce formalisme du logiciel ArcGis de la société ESRI.
2. La gestion du modèle : il s'agit du choix de ce qu'on souhaite visualiser et de la manière de le réaliser. Notons qu'il existe plusieurs niveaux de détails (ou LoD pour Levels-of-Detail) pour l'affichage en 3D. Ces niveaux tel qu'ils sont définis par CityGML (modèle standard de représentation des objets en 3D) sont illustrés dans la figure 3.6.

   Ce LoD varie de la simple élévation du sol au modèle 3D avec textures et modélisation de l'intérieur des bâtiments en haute définition. Nous nous contentons dans ce travail du LoD1 qui permet un affichage des bâtiments sur le sol sans fournir de détails sur la forme des toits ni sur la texture.
3. La production de l'image : le rendu est créé en se fondant sur la description du modèle. Une image est produite avec les paramètres de position, d'orientation et de focale de la caméra virtuelle. L'image obtenue contient les objets visibles avec ces paramètres.

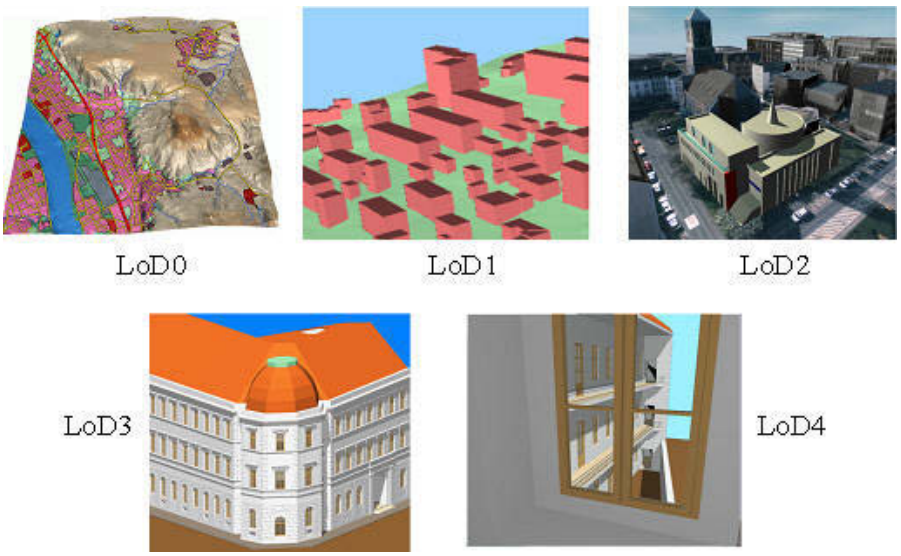


FIGURE 3.6 – Les cinq niveaux de détails définis par CityGML. (Source : [Kolbe 05])

La création d'un rendu qui respecte une position de la caméra, une orientation et une focale permet d'obtenir une image. Cette image est analogue à celle prise par un appareil photographique dans les mêmes conditions que dans le monde réel.

La focale de l'objectif d'un appareil photographique permet de cadrer précisément ce qu'on souhaite photographier. La focale [Jullier 98] représente alors la portion du monde capturable par l'objectif de l'appareil photo. Il est donc nécessaire dans un monde virtuel d'afficher uniquement les objets correspondant à cette focale.

La valeur de l'illumination de chaque pixel est déterminée principalement par deux méthodes. Le premier est le lancer de rayons qui a prouvé sa rapidité et le réalisme de son résultat. Le Z-buffer [Chen 02] est une alternative qui permet d'effectuer un test de profondeur (par rapport au point de vue) pour tous les éléments du monde virtuel, puis de sélectionner les éléments les plus proches pour les afficher.

Outils utilisés

Nos images de synthèse sont créées par deux moyens différents : la plateforme MapYourPicture et ArcScene.

3.3.1.1 MapYourPicture

La plateforme MapYourPicture déjà développée permet de naviguer dans le monde urbain 3D de la ville de Saint-Étienne et dans le centre de la ville de Lyon. Elle fournit un rendu assez fidèle à la réalité. Ce rendu est créé en utilisant l'algorithme du Z-buffer. Nous avons détaillé et illustré cette plateforme dans le deuxième cha-

pitre. Rappelons simplement que cet environnement ne prend pas en considération le relief et considère que le sol est toujours plat.

Le relief

Nous avons constaté dans la section précédente que l'une des principales limites et sources d'erreurs dans le processus adopté par le démonstrateur MapYourPicture est la non prise en compte du relief. La conceptualisation du modèle de cette plateforme ne permet pas de prendre en charge le relief. C'est la raison principale qui nous a conduit à créer une nouvelle modélisation 3D de l'espace urbain.

La figure 3.7 schématise l'influence du relief sur l'étendue du monde visible.

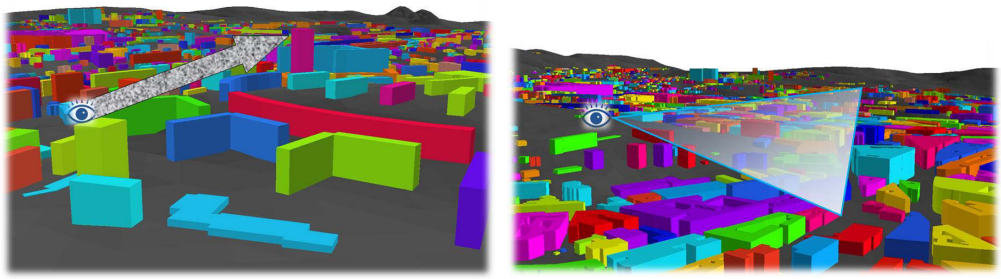


FIGURE 3.7 – Importance du relief pour une visibilité fidèle à la réalité.

Un sol plat ne représente pas suffisamment une ville possédant du relief. En effet l'observateur est capable d'apercevoir les bâtiments qui sont situés en hauteur. Sans prise en compte du relief, il est impossible de voir à partir d'un point de vue en ville d'autres bâtiments à part ceux du voisinage immédiat ou les plus hauts étages des gratte-ciels.

Dans l'autre sens, un observateur se situant en hauteur est capable de voir tous les bâtiments qu'il peut surplomber si sa vue est bien dégagée.

La position d'un appareil photographique dépend du relief. Dans une montée, on est amené à incliner l'objectif vers le haut pour ne pas photographier une grande partie du sol.

En conclusion, il a fallu utiliser l'altitude du sol au pied de chaque bâtiment pour le placer à sa bonne altitude dans l'espace urbain. Ces altitudes sont fournies par le modèle numérique de terrain qui sera décrit dans le chapitre suivant. Ces données nous ont permis de se rapprocher encore plus du monde réel tout en restant au niveau de détail LoD1.

Traitement d'image sur les images de synthèse réalisés par MapYourPicture

Le rendu synthétisé par le démonstrateur de MapYourPicture offre un rendu fourni par l'algorithme de Z-Buffer et exécuté par la bibliothèque OpenGL.

L'accélération du processus de création de rendu est optimisée par une limitation de la résolution. Il en résulte des marches d'escalier visibles au niveau des lignes obliques. Les lignes droites apparaissent sous forme d'un ensemble de petits segments. Cet effet est assez visible dans la figure 3.8.

La détection de lignes de Hough conduirait à déceler plusieurs lignes pour une seule droite sur l'image.

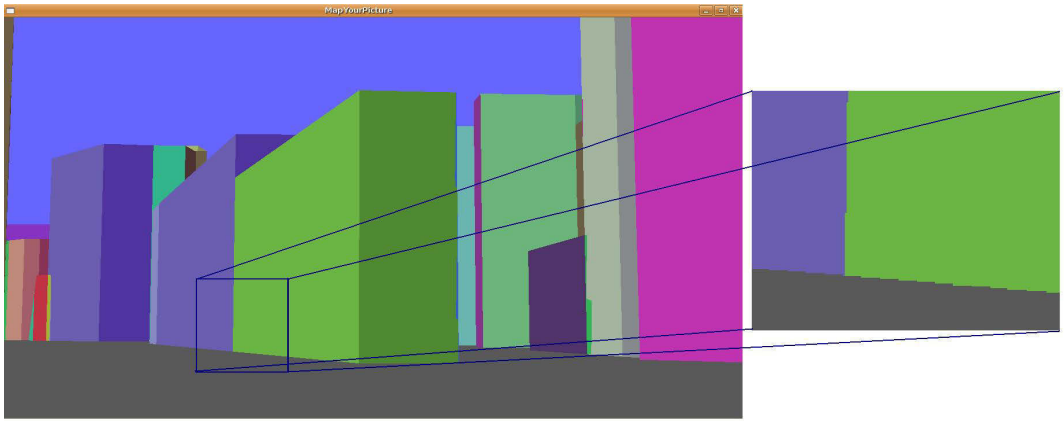


FIGURE 3.8 – Effet de marches d'escalier sur la détection de ligne dans les images de synthèse. (Source : capture d'écran du monde virtuel crée par le démonstrateur de MapYourPicture)

Nous procédons en conséquence à une augmentation du seuil utilisé pour définir le nombre minimal de points alignés. En utilisant cette manipulation, nous obtenons moins de segments de marches d'escalier, considérées comme autant de lignes différentes.

Une autre anomalie apparaît dans un cas particulier de ce type d'images de synthèse. Elle est aussi due à l'accélération souhaitée de l'affichage du monde virtuel. On ne demande d'afficher qu'une zone de la ville à partir d'une base de données géographique restreinte. En dehors de cette zone un fond noir apparait comme illustré par la figure 3.9. Nous observons alors une zone noire comprise entre le ciel et la sol.

Ce défaut d'affichage interfère avec un bon appariement entre une photographie et une image de synthèse puisque les éléments visibles sur la photographie sont absents en totalité ou en partie. Ces rendus sont trop partiels et imprécis.

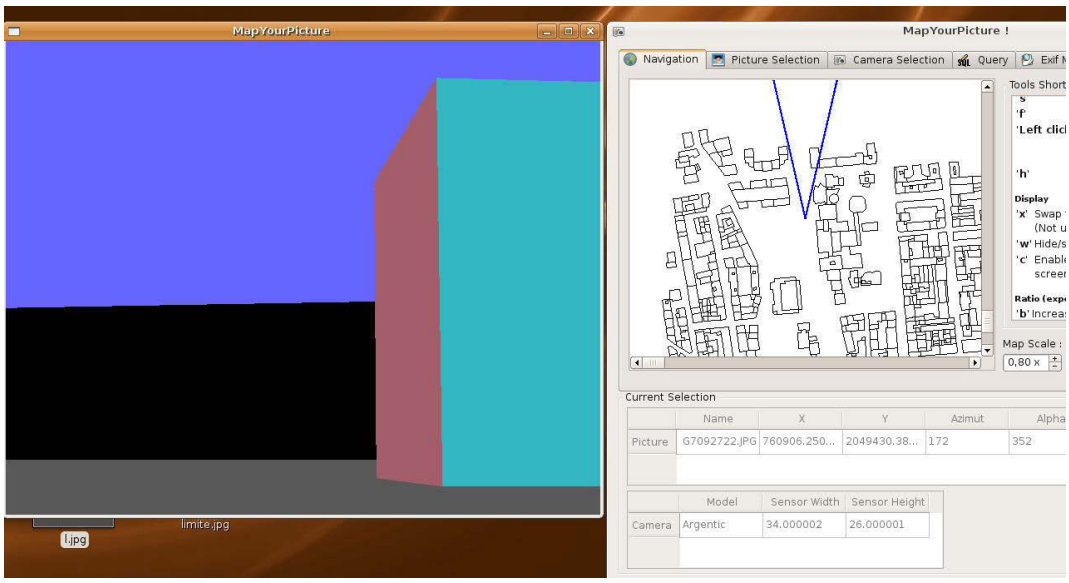


FIGURE 3.9 – Affichage des zones inconnues dans l'image de synthèse. (Source : capture d'écran du démonstrateur de MapYourPicture)

Le remède à cette anomalie est la restriction de l'utilisation de l'appariement aux photographies dont tous les bâtiments susceptibles d'être visibles sont modélisés dans le monde virtuel.

Cette sélection n'est pas automatique, d'autant plus que le simple fait de se positionner au bord de la zone modélisée n'est pas suffisant pour éliminer la photographie prise. Il faut en plus vérifier que la visée est dirigée vers la zone non modélisée et non pas vers la zone parfaitement modélisées de la base de données.

3.3.1.2 ArcScene

L'outil ArcScene d'ESRI nous permet d'effectuer une navigation libre dans un modèle 3D. Il offre le choix des paramètres de la prise de vue souhaitée. Nous illustrons à la figure 3.10 l'interface d'ArcScene 10.0 avec le formulaire qui permet la saisie des données relatives à la position de l'objectif de la caméra virtuelle dans la ville de Saint Étienne.

Notons qu'il est possible de préciser :
- les coordonnées X, Y et Z de l'observateur.
- les coordonnées X, Y et Z de la cible.
- l'angle de vision qui correspond à la focale.
- l'angle d'inclinaison verticale et horizontale.

ArcScene, la visionneuse 3D de la suite ArcGIS permet aussi d'exporter les images de rendu sous différents formats et notamment dans le format de compression JPEG.

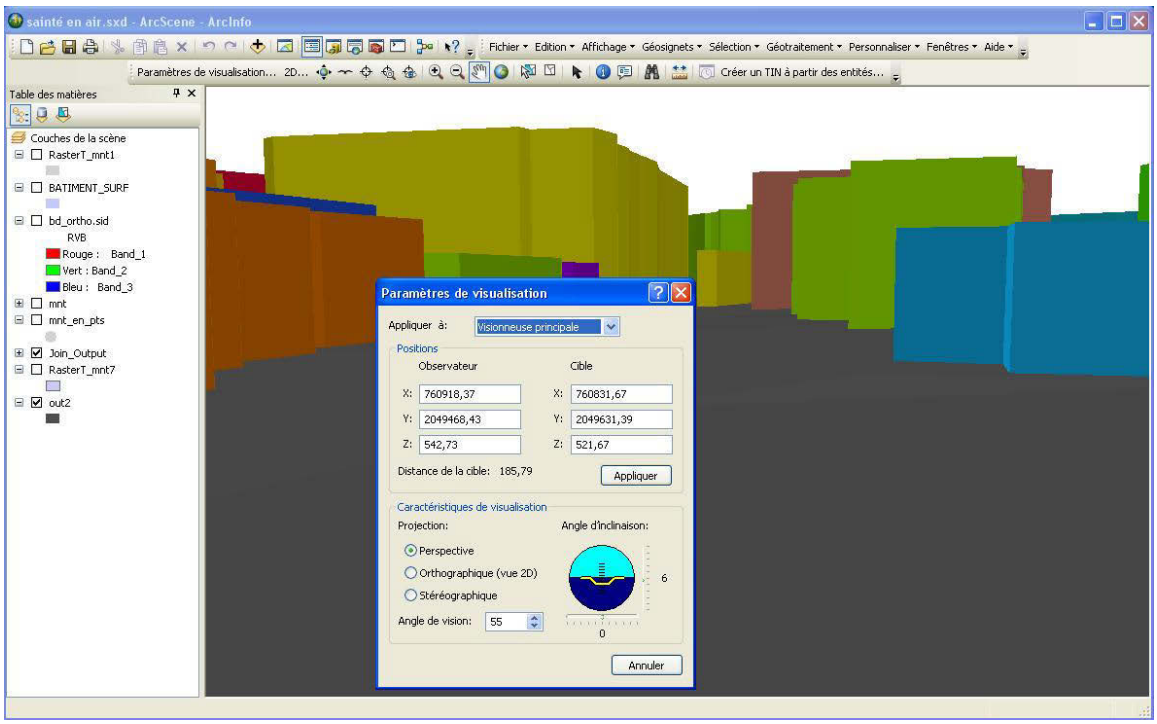


FIGURE 3.10 – Paramètres de visualisation dans ArcScene. (Source : capture d'écran d'ArcScene. ESRI)

Traitement d'image sur les images de synthèse réalisées par ArcScene

ArcScene offre des images en haute définition. L'effet de marches d'escalier y est donc moins visible.

La limite de la zone modélisée dans le monde virtuel n'affiche pas d'anomalies d'affichage. Les zones inconnues prennent la couleur de l'arrière plan. La figure 3.11 montre les deux derniers bâtiments, le vide à l'arrière qui correspond au sol dont les données sont plus étendues que celles des données vectorielles des bâtiments, et le ciel qui prend la couleur blanche. C'est la couleur de l'arrière plan contrairement à celle du démonstrateur de MapYourPicture qui propose 3 couleurs en plus de celles des bâtiments, un ciel bleu, un sol gris et un arrière plan noir.

Le problème des images contenant des entités manquantes qui émanent des bordures des zones modélisées est résolu de la même manière que les images créées en utilisant le démonstrateur de MapYourPicture.

Les images de synthèse ne contiennent que des bâtiments. Les lignes détectées sur ce type d'image délimitent ces bâtiments. Les lignes sont alors bien représentatives du contenu de l'image.

Les images de synthèse subissent quatre traitements particuliers.

1. Transformation des images en couleur en niveaux de gris. Cette étape est obli-

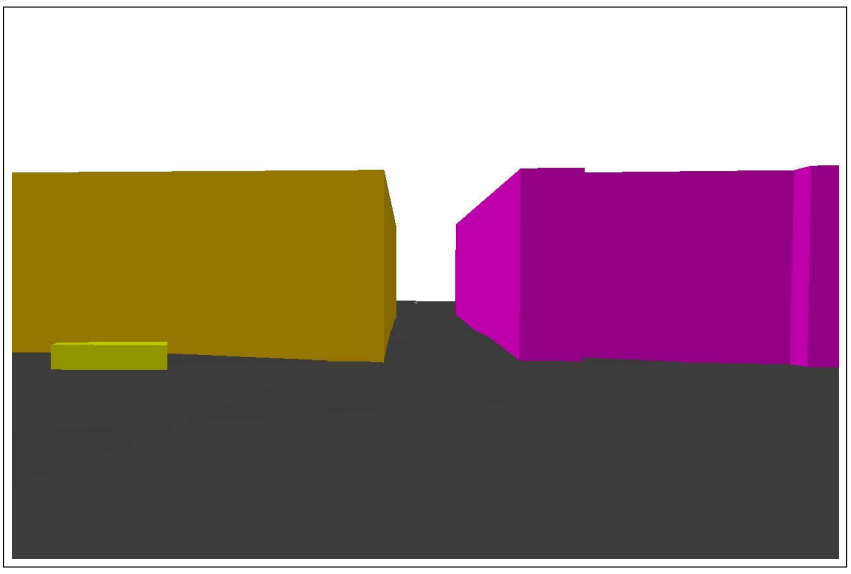


FIGURE 3.11 – Illustration des limites de la base de données utilisée. (Source : capture d'écran d'ArcScene)

gatoire pour permettre à la méthode de détection de lignes par transformation de Hough de prendre comme attribut d'entrée une image en niveaux de gris.

2. Augmentation du seuil pour la transformation de Hough, pour moins détecter les marches d'escalier.

   Nous éliminons les lignes détectés sur les marches d'escalier également avec le troisième traitement.

3. Filtrage des lignes. Nous avons créé un petit module permettant d'éliminer les lignes trop proches les unes des autres. Dans un milieu urbain les lignes sont habituellement bien espacées. Ce module traite les lignes par ordre et cherche les lignes très voisines. Il les éliminent en ne gardant que la première ligne détectée de chaque groupe de lignes voisines.

   Le voisinage est défini par expérimentation. Deux lignes parallèles ayant un même degré d'inclinaison θ et dont la différence entre leurs deux distances ρ ne dépasse pas 10 pixels, sont considérées comme voisines.

4. Augmentation du contraste. Les couleurs des bâtiments peuvent être proches entre elles ou voisines de la couleur du sol surtout en transformant les images du mode couleur au mode niveaux de gris. L'augmentation du contraste permet une détection plus facile des contours et par conséquence des lignes.

Sur les images de synthèse, nous détectons usuellement de 5 à 15 lignes au maximum. C'est un nombre assez faible par rapport aux lignes détectées sur les photographies qui s'explique par la simplification géométrique que le modèle produit par rapport à la réalité.

3.3.2 Matlab

Matlab [Hahn 07] est un système informatique à base de matrices conçu pour aider à résoudre les problèmes scientifiques et d'ingénierie. C'est un langage de programmation [Burel 01] de haut niveau permettant de programmer rapidement différents calculs scientifiques.

Matlab est devenue une plateforme permettant d'effectuer des calculs sur des types de données simples qui s'est adaptée à la majorité des traitements de données existantes. Des boîtes à outils spécifiques sont commercialisées pour enrichir et élargir la couverture des données traitées.

Dans ce travail nous utilisons la boîte à outils permettant de manipuler les photos et d'effectuer les traitements d'analyse d'image les plus connus.

Image Processing Toolbox [MathWorks 10b] propose un ensemble d'algorithmes et d'outils graphiques de référence pour le traitement, l'analyse, la visualisation et le développement d'algorithmes en traitement d'images. Cette boîte à outils permet de restaurer des images bruitées ou dégradées, d'améliorer la netteté d'une image, d'analyser des formes et des textures ainsi que de nombreuses autres fonctionnalités.

La plupart des fonctions de la boîte à outils sont écrites en langage MATLAB. Cela permet d'analyser les algorithmes, de modifier le code source et de créer ses propres fonctions personnalisées.

Matlab offre aussi une autre boîte à outils pour gérer les données géographiques. La Mapping Toolbox [MathWorks 10c] propose les outils et les utilitaires d'analyse spatiale et d'affichage de cartes. Elle permet aussi d'importer des données vectorielles et raster depuis différents format de fichiers. L'affichages des cartes en 2D est correct. Cependant, l'outil manque encore de réalisme en 3D.

La schématisation sous forme de graphes et l'appariement de graphes sous Matlab nécessitent l'utilisation d'une troisième boîte à outils.

Bioinformatique Toolbox [MathWorks 10a] fournit des algorithmes et des techniques de visualisation pour plusieurs types de données comme les graphes et les micro-réseaux.

Matlab demeure actuellement la seule plateforme qui englobe toutes ces fonctions dans un seul outil. Les matrices qui forment les images, les tableaux qui représentent les lignes, les graphes et les cartes sont des données manipulées conjointement dans une seule application.

Matlab offre non seulement la possibilité d'effectuer des traitements de haut niveau comme l'analyse de forme sur une image mais aussi d'effectuer des traitements de très bas niveau. L'accès et la modification des données au niveau du pixel est ainsi envisageable avec cet outil. Ces possibilités et sa polyvalence ont conduit à considérer Matlab comme l'outil le plus adapté à nos besoins de recherche.

3.4 Traitement d'image sur les photographies

Les photographies dans un environnement urbain ne contiennent pas uniquement des bâtiments. On y trouve des arbres, des pancartes publicitaires, des signalisations, des voitures, des piétons, des cyclistes...

La détection de lignes sur ces photographies engendre un nombre important de lignes qui n'appartiennent pas aux bâtiments. Ces lignes sont source de fausses détections. Il est donc important de les éliminer ou au moins d'en éliminer le maximum.

Les éléments indésirables qui apparaissent sur une image prise en milieu urbain, se trouvent dans la majorité des cas au sol. Une grande partie de ces éléments se situe au pied des bâtiments.

L'élimination du sol est une des solutions possibles pour éliminer les objets indésirables du paysage urbain.

Cette section détaille la procédure de l'élimination du sol et ensuite présente la spécificité de la détection des lignes dans une photographie en milieu urbain.

3.4.1 Élimination du sol

Cette section argumente l'utilité de l'élimination du sol. Ensuite elle décrit la solution proposée fondée sur les travaux d'une équipe américaine. Enfin elle illustre des effets de l'élimination du sol sur les photographies du milieu urbain.

3.4.1.1 Importance de l'élimination du sol

Dans une photographie urbaine, le sol contient des éléments indésirables ou parasites. Les piétons, les animaux de compagnie, les voitures, les bacs à fleurs... sont les objets les plus courants qu'on trouve au pied des bâtiments. Ces éléments ne sont pas modélisés dans les bases de données géographiques utilisées.

Certains de ces éléments parasites sont décrits dans des bases de données spécifiques. Ces bases sont capables de modéliser les arbres, les panneaux de signalisation, les bacs à fleurs, mais jamais les objets en mouvement comme les piétons, les vélos...

En étudiant l'espace urbain stéphanois, nous avons constaté aussi des lignes parfaitement détectées au niveau du sol. Ces lignes correspondent aux bordures des trottoirs, aux barrières de sécurité et aux rails des lignes de tramway.

Tous ces éléments nous poussent à éliminer le sol de la photographie afin de minimiser le nombre des fausses détections de lignes qui devrait idéalement correspondre uniquement à des bâtiments.

3.4.1.2 Méthode utilisée

Derek Hoiem, Alexei A. Efros et Martial Hebert [Hoiem 05] ont proposé une démarche permettant d'analyser le contexte d'une image. Ils s'appuient sur une heuristique qui permet de segmenter l'image en trois parties (surface horizontale ou base, zones verticales et ciel) en combinant quatre critères pour déterminer l'appartenance d'une zone à une de ces trois parties [Hoiem 07] :

1. La localisation de la zone traitée : le ciel se situe toujours au sommet de l'image. La partie basse de l'image est à priori représentative du sol. La zone centrale verticalement a de fortes chances qu'elle représente les objets verticaux de l'image.
2. La couleur : le ciel possède une teinte bleu ou blanche dans la majorité des cas. Le sol est de couleur brune ou verte en milieu rural et grise en ville.
3. La texture : une combinaison de filtres permet de retrouver les orientations des textures et des contours.
4. Le point de vue : l'étude des points de fuite permet de reconnaître l'orientation des objets.

En combinant ces quatre descripteurs et en utilisant une heuristique basée sur un apprentissage, la distinction entre sol, objets verticaux et ciel s'avère assez pertinente. De plus, cette équipe détecte l'orientation des objets verticaux par l'examen de l'orientation des textures et l'analyse du point de vue. Des flèches de direction indiquent le sens d'orientation des bâtiments.

La figure 3.12 illustre un exemple de segmentation d'une photographie d'un hôtel. Le ciel est teinté en bleu, la zone verticale est en rouge et le sol est colorié en vert.



FIGURE 3.12 – Résultats de segmentation d'une image urbaine. (Source : image personnelle | rendu généré avec Matlab)

Nous nous intéressons uniquement à la partie sol que nous éliminons de l'image en remplaçant chaque pixel estimé comme appartenant à la zone 'sol' par la valeur nulle, correspondant au noir. Un exemple de ce traitement est illustré dans la figure 3.13.



FIGURE 3.13 – Élimination du sol sur une photographie. (Sources : image personnelle | rendu généré avec Matlab)

Notons que la méthode utilisée pour l'élimination du sol élimine aussi dans certains cas une partie du bâtiment. L'impact de cet effet indésirable est analysé dans le chapitre suivant.

3.4.2 Extraction des droites de Hough

Une fois le sol éliminé de la photographie, nous avons à disposition une image avec un minimum de lignes parasites détectables. Nous utilisons uniquement les images sans sol.

Les photographies passent du mode couleur, très important pour la phase d'élimination du sol, au mode niveaux de gris. C'est la condition d'utilisation de la transformée de Hough. En effet, nous analysons les contours sur ces photographies. La transformation de Hough permet de savoir si ces contours forment des lignes ou pas.

La texture des bâtiments contient des fenêtres, des balcons... . Ces éléments contiennent des segments qui sont détectés comme les lignes délimitant les bâtiments. Nous procédons alors à une augmentation du seuil pour ne considérer que les lignes qui sont représentées par plusieurs dizaines de pixels. Cette manœuvre élimine la majorité des lignes détectées dans les textures ou les fenêtres.

Ce seuil est incrémenté à chaque boucle de notre algorithme jusqu'a retrouver la totalité des lignes précédemment détectées dans la photo de synthèse, candidate à l'appariement.

3.5 Conclusion

En conclusion à ce chapitre rappelons que nous avons détaillé notre approche permettant de retrouver une photographie dans un monde virtuel. La technique adoptée est l'interfaçage par appariement de lignes. Nous utilisons les lignes comme élément de comparaison.

L'importance d'avoir le relief dans un monde en trois dimensions représentatif du paysage urbain est explicite surtout dans une ville entouré de collines comme c'est le cas de la ville de Saint Étienne.

Les photographies subissent un traitement spécifique permettant une meilleur détection des lignes. Ces photographies sont comparées ensuite avec les images de synthèse selon notre approche.

Une image de synthèse qui est comparable à une photographie nous renseigne sur la localisation et l'orientation de la photographie. En ayant ces éléments en main, nous pouvons utiliser l'information du SIG pour décrire le contenu photographique en caractérisant les bâtiments et les rues de la photographie.

Le chapitre suivant expérimente cette approche proposée pour démontrer son efficacité.

Chapitre 4 : Validation du concept

Chapitre 4

Validation du concept

Ce chapitre permet d'expérimenter l'approche proposée au cours des travaux de thèse et d'évaluer l'efficacité de la démarche. L'objectif est de chercher la position la plus exacte possible d'une photographie en utilisant des images de synthèse de la ville issues de la base de données géographique.

La première partie de ce chapitre présente la démarche adoptée pour l'analyse des données. Dans la deuxième section nous décrivons le jeu de données utilisé. Nous présentons en détail les photographies et la base de données géographique qui ont servi à la validation du concept.

Ensuite nous détaillons le module de validation de notre concept d'interfaçage par appariement de lignes. Nous évaluons le degré de pertinence des résultats et nous précisons les principales limites et lacunes de la démarche.

La conclusion donne les statistiques générales sur toutes les expérimentations effectuées.

4.1 Méthode d'analyse des données

Notre méthode d'analyse des données relève de la fouille de données (Data Mining) qui est traduit littéralement par "creusage des données" [Jambu 98].

La démarche méthodologique d'un processus de fouille de données passe par 6, 8 ou 10 étapes selon la complexité des données [Jambu 98, Larose 05, Lefébure 98, Tufféry 02].

Nous avons opté pour la méthodologie standardisée mise au point en 1996 par le CRISP-DM (CRoss-Industry Standard Process for Data Mining) qui fournit un processus du cycle de vie en six phases [Larose 05] :

1. **Compréhension de la problématique d'étude**
   Nous cherchons à trouver la position géographique ainsi que l'orientation (rotation autours des 3 axes de rotation dans l'espace) d'une photographie ne

possédant pas ces informations.

Cet objectif général sera spécifié dans le chapitre suivant en trois scénarios différents.

2. **Compréhension des données**

   Les données dont nous disposons sont les taux d'appariement entre des photographies et des images de synthèse prises à partir d'un modèle 3D. Chacune des images de synthèse possède une position géographique, une focale et une orientation.

   Les données sont collectées dans un tableur et extraites à partir de calculs effectués avec Matlab. La description des données est l'objet de la seconde section de ce chapitre.

3. **Préparation des données**

   Nous avons sélectionné 20 photographies (figure 4.1) situées sur la figure 4.2 pour les analyser. Nous comparons ces photographies avec 73 images de synthèse réparties dans le voisinage de 29 positions (figure 4.3).

   Les taux d'appariement sont triés du plus grand au plus petit. Les moyennes des taux sont calculées.

4. **Modélisation**

   Chaque photographie est reliée à une image de synthèse par un taux d'appariement (voir section 4.3). La moyenne des taux d'appariement avec toutes les images de synthèse est comparée à la valeur des taux d'auto-reconnaissance (définition dans la section 4.3.2).

   Nous modélisons un arbre de décision (section 4.3.3) dans lequel un appariement est considéré comme bon si sa valeur est supérieure à la moyenne de tous les taux relatifs à la photographie.

5. **Evaluation**

   L'évaluation permet de caractériser la qualité et l'efficacité de l'arbre de décision créé. Si l'image de synthèse correspondant véritablement à une photographie n'a pas le meilleur taux d'appariement, son taux est toujours supérieur à la moyenne des taux. Dans la quasi totalité des cas une photographie est appariée concrètement quand son taux s'écarte d'au moins un écart type de la moyenne des taux.

   La partie 4.3 présente en détail cette phase d'évaluation.

6. **Déploiement**

   La réalisation de rapport et l'extraction de résultats à partir de notre démarche d'interfaçage par appariement de lignes fait l'objet des trois scénarios d'utilisation donnés dans le chapitre 5.

Ces 5 premières phases permettent de produire un arbre de décision (phase 4) à partir d'une sélection des 20 photographies de la phase d'apprentissage. L'évaluation

(phase 5) se fait ensuite sur d'autres données (13 photographies) pour vérifier les règles de l'arbre construit.

On trouvera en conclusion de ce chapitre les statistiques relatives à ces deux groupes d'échantillons.

4.2 Données utilisées pour la validation de la démarche

L'expérimentation de la démarche nécessite une base de données de photographies. La méthode de la prise de vue ainsi que les différents appareils utilisés sont détaillés dans le premier paragraphe de cette section. Nous décrivons ensuite la base de données géographique utilisée en détaillant les paramètres servant à construire le monde virtuel.

4.2.1 Prises de photographies

Le jeu de données d'apprentissage est composé de 20 photographies prises dans le milieu urbain stéphanois composé de bâtiments anciens du XIXe siècle ou plus récents du XXe siècle et du XXIe siècle.

Elles sont présentées dans la figure 4.1. Notons que les photographies sont traitées en niveaux de gris comme expliqué dans le chapitre précédent. C'est la raison pour laquelle nous les présentons ainsi, même si la prise de vue originale est en couleur.

Trois de ces photographies ont été prises au même emplacement avec une orientation ou simplement une focale différente.

Une fois l'analyse de ces photographies réalisée, nous avons effectué la phase d'évaluation de notre démarche d'interfaçage par appariement de lignes avec 13 autres photographies. Cela fait donc un total de 33 photographies prises dans la ville de Saint-Étienne.

Ces 20 photographies de la figure 4.2 nous permettent d'effectuer la phase initiale d'apprentissage. Nous avons choisi dans la majorité des cas des situations simples avec un minimum d'obstacles et des orientations simples de l'appareil photographique.

Ces photographies sont prises aux alentours de l'université et dans la Grande Rue aux bâtiments divers par le style architectural.

Les pointeurs rouges sur la carte de la figure 4.2 indiquent les positions des appareils photographiques lors des prises de vue.

Rappelons que la première opération subie par ces photographies est l'élimination du sol (exemple : la photographie 4.1b). La totalité des 20 photographies est présentée dans l'annexe A.



(a) Photo 1 (b) Photo 3 (c) Photo 4 (d) Photo 5



(e) Photo 6 (f) Photo 7 (g) Photo 9 (h) Photo 10



(i) Photo 11 (j) Photo 12 (k) Photo 15 (l) Photo 19



(m) Photo 20 (n) Photo 23 (o) Photo 24 (p) Photo 26



(q) Photo 27 (r) Photo 28 (s) Photo 29 (t) Photo 33

FIGURE 4.1 – Jeu de données photographiques d'apprentissage.



FIGURE 4.2 – Position des 20 prises de vue.

Mesure de positionnement et d'orientation

Pour chacune des photographies, une position GPS et une orientation ont été prises. Les données d'orientation peuvent être données par une boussole ou par une centrale inertielle, qui, comme on l'a vu dans le chapitre 1 section 1.2.3, offrent une précision très accrue même dans des conditions extrêmes. Nous avons opté pour les centrales embarquées sur les Smartphones qui offrent une précision suffisante pour nos besoins à un coût identique.

Appareils utilisés

Les photographies ont été prises en utilisant cinq appareils différents. Chacun de ces appareils est décrit précisément dans l'annexe B. Nous nous contentons de caractériser succinctement les trois familles d'appareils dans cette section :

1. Deux appareils photographiques sont équipés d'une puce GPS intégrée qui donne directement la position géographique de la prise de vue si la réception GPS est valide. Ces données et d'autres (nombre de satellites, altitude, date et heure...) sont stockées directement dans l'extension EXIF des images compressées en même temps que les caractéristiques de la caméra et les paramètres de la prise de vue (focale, position, direction de visée...).
2. Deux Smartphones, décrit comme systèmes hybrides dans le chapitre 1, com-

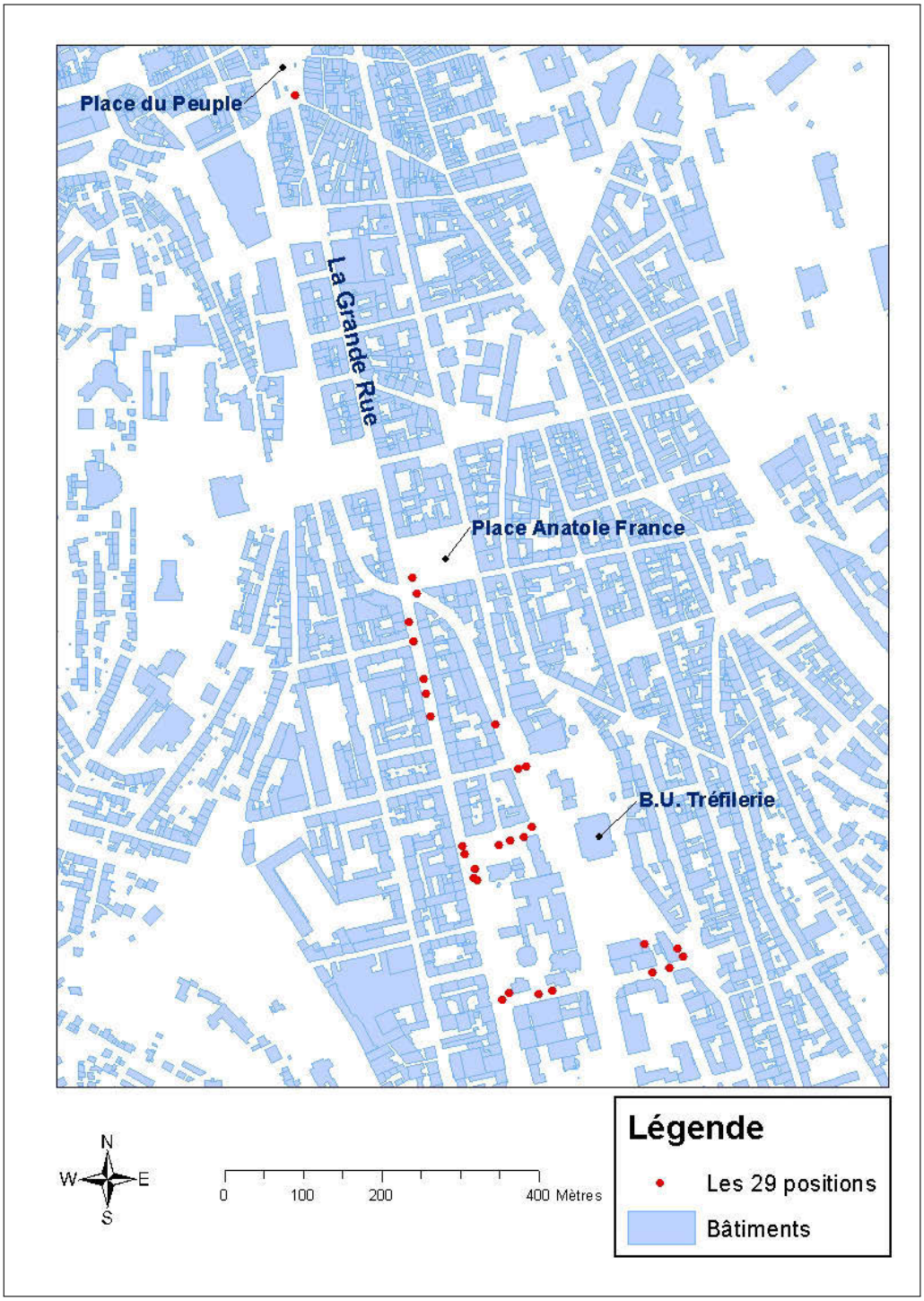


FIGURE 4.3 – Les 29 positions des images de synthèse.

portent une centrale inertielle et une puce GPS ainsi qu'un appareil photo numérique. Les deux utilisent le système d'exploitation embarqué Android de

Google. Nous avons programmé sous la plateforme Eclipse une application sur ces deux Smartphones en utilisant le langage JAVA. Elle permet d'afficher les informations sur la position géographique donnée par le GPS en cas de couverture par les satellites de ce type de système ou une position beaucoup plus approximative par l'identification de la cellule GSM de la téléphonie.

L'application considère aussi les différents degrés d'inclinaison de l'appareil sur les trois axes de rotation dans l'espace. Toutes ces données sont stockées dans les métadonnées de l'image résultante avec les coordonnées de positionnement autour des trois axes de rotation X, Y et Z (voir l'imprimé écran de cette application dans la figure 4.4).



FIGURE 4.4 – Capture d'écran de la prise N°2 avec les valeurs de rotation sur les axes X, Y et Z en degrés. (Source : capture d'écran de l'XPERIA X10)

3. Un réflex classique permettant de prendre des photographies numériques standard. Il a été couplé avec un GPS pour récupérer la position géographique et avec une boussole analogique pour préciser la direction de visée pendant la prise de la photographie.

4.2.2 Base de données géographiques

Nous avons décrit dans le chapitre précédent les deux plateformes qui ont permis la création des images de synthèse que nous voulons apparier avec les photographies. Le démonstrateur de MapYourPicture et ArcScene sont les deux outils utilisés. Les données géographiques utilisées par chacun de ces outils sont détaillés dans les deux paragraphes suivants.

4.2.2.1 Démonstrateur de MapYourPicture

Le démonstrateur de MapYourPicture utilise des données géographiques anciennes fournies par l'IGN. Les bâtiments sont ceux de la version de 2006 de la

Google. Nous avons programmé sous la plateforme Eclipse une application sur ces deux Smartphones en utilisant le langage JAVA. Elle permet d'afficher les informations sur la position géographique donnée par le GPS en cas de couverture par les satellites de ce type de système ou une position beaucoup plus approximative par l'identification de la cellule GSM de la téléphonie.

L'application considère aussi les différents degrés d'inclinaison de l'appareil sur les trois axes de rotation dans l'espace. Toutes ces données sont stockées dans les métadonnées de l'image résultante avec les coordonnées de positionnement autour des trois axes de rotation X, Y et Z (voir l'imprimé écran de cette application dans la figure 4.4).



FIGURE 4.4 – Capture d'écran de la prise N°2 avec les valeurs de rotation sur les axes X, Y et Z en degrés. (Source : capture d'écran de l'XPERIA X10)

3. Un réflex classique permettant de prendre des photographies numériques standard. Il a été couplé avec un GPS pour récupérer la position géographique et avec une boussole analogique pour préciser la direction de visée pendant la prise de la photographie.

4.2.2 Base de données géographiques

Nous avons décrit dans le chapitre précédent les deux plateformes qui ont permis la création des images de synthèse que nous voulons apparier avec les photographies. Le démonstrateur de MapYourPicture et ArcScene sont les deux outils utilisés. Les données géographiques utilisées par chacun de ces outils sont détaillés dans les deux paragraphes suivants.

4.2.2.1 Démonstrateur de MapYourPicture

Le démonstrateur de MapYourPicture utilise des données géographiques anciennes fournies par l'IGN. Les bâtiments sont ceux de la version de 2006 de la

couche Bâti de la BDTOPO. C'est une couche vecteur qui contient la géométrie au sol des bâtiments ainsi que, pour chacun d'entre eux, une hauteur minimale et une hauteur maximale. Nous avons décrit les caractéristiques de cette base de données géographique dans le chapitre 3.

Techniquement, les fichiers fournis par l'IGN sont au format Shape d'ESRI. Ils sont transformés en entités géométriques dans le système de gestion de base de données spatiale PostGIS/PostgreSQL. Cette transformation permet de passer d'un ensemble de polygones en shape file à une collection d'entités spatiales structurées dans une base de données par application de la fonction "shp2pgsql".

Les données géographiques dans la base de données PostGIS sont affichées sur un fond de couleur grise et sous un ciel bleu. Les caractéristiques de cet environnement ont été présentées dans le chapitre 3.

Afin de garder une continuité avec le travail antérieur, nous allons utiliser quelques (5) images de synthèse créées par cette plateforme. Ce sont des images qui sont compatibles avec les autres images de synthèse obtenues par ArcScene.

4.2.2.2 ArcScene

ArcScene, module 3D de la société ESRI, a permis de créer la quasi totalité des images de synthèse que nous utilisons, en particulier celles situées dans les zones avec un relief prononcé que le démonstrateur ne permet pas de prendre en charge.

Deux types de données géographiques sont utilisés pour bâtir le monde virtuel 3D.

1. Les données vecteur des bâtiments décrivent les bâtiments sous forme de polygones. La couche utilisée est la dernière version à ce jour fournie par l'IGN. Il s'agit de la version 2 de la BD TOPO. La projection utilisée est conforme à la norme RGF93 en projection conique Lambert (voir chapitre 1).

   Notons que cette version comporte plusieurs évolutions, notamment dans la manière de classifier les bâtiments. La classe Bâtiment a été divisée en onze classes, de façon à faciliter et accélérer les applications utilisant ces bâtiments, en particulier les bâtiments possédant une fonction particulière [IGN 08a].

   Nous avons pris en considération certaines classes pour représenter tous les bâtiments de la zone qu'on souhaitait modéliser. Les quatre classes sélectionnées sont celles qui possèdent une hauteur dans cette base de données et font plus de 20 m^2. Les classes qui respectent ces deux conditions sont : BATI_INDIFFERENCIE, BATI_REMARQUABLE, BATI_INDUSTRIEL, RESERVOIR. Elle sont décrites dans l'annexe C [IGN 11b]. Ces quatre classes ont sept attributs en commun présentés en détail dans l'annexe C. Nous utilisons l'attribut HAUTEUR pour modéliser les bâtiments par extension.

   Cette hauteur correspond à la différence d'altitude entre le point le plus élevé du pourtour du bâtiment et un point situé au pied du bâtiment [IGN 11b](figure

4.5). Sa valeur est arrondie au mètre.

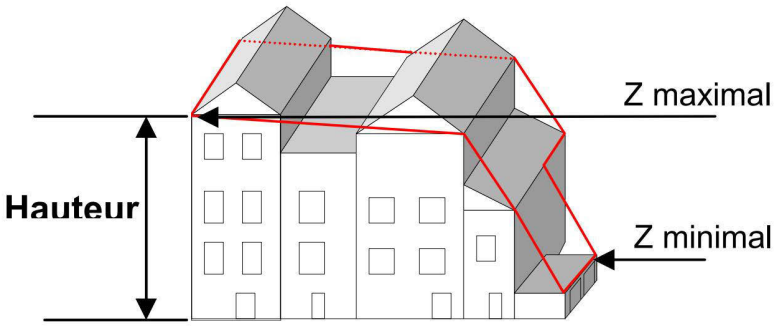


FIGURE 4.5 – Description de la hauteur et des altitudes maximale et minimale dans la BD TOPO. (Source : descriptif de contenu de la version 2 de la BD TOPO)

Nous avons reconstruit une partie de la ville de Saint-Étienne qui englobe largement toutes les zones de prises de photographies. Cette zone représente environ un quart de la ville et comprend plus de 2000 bâtiments.

2. Les données que nous utilisons pour déterminer la hauteur du sol sont fournies par le modèle numérique de terrain (MNT) au format raster de l'IGN. Le MNT est calculé à partir de la base de données altimétriques des courbes de niveau et points cotés de la BD TOPO en complétant les zones sans altitude [IGN 11a]. Le MNT est fourni avec le produit BD ALTI en version 1. La couche utilisée a une résolution de 50 m.

   La précision de ces données varie selon la zone géographique. La région stéphanoise est caractérisée par une précision planimétrique proche de 2 mètres et une précision altimétrique variant de 1 à 2.5 mètres [IGN 11a].

   Nous procédons à la transformation du mode raster au mode vecteur avec les outils fournis par ArcMap. Le passage en mode vecteur est obligatoire pour permettre de fusionner ces données avec la couche des bâtiments et pour avoir plus de cohérence entre le sol et les bâtiments. Nous effectuons ce passage afin de passer du mode raster du MNT à une couche vecteur que nous allons utiliser conjointement avec la couche vecteur des bâtiments. La couche vecteur obtenue est discontinue puisque chaque cellule possède une altitude propre. Ce phénomène de discontinuité est illustré dans la figure 4.6.

   Nous procédons alors à une triangulation pour obtenir une représentation continue du sol. L'outil utilisé pour cette triangulation permet de créer une structure TIN (Triangulated Irregular Network : réseau de triangles irréguliers) à partir de la couche des entités spatiales extraites du MNT.

   Il en résulte alors une surface continue sur laquelle nous plaçons nos bâtiments issues de la couche BD TOPO.

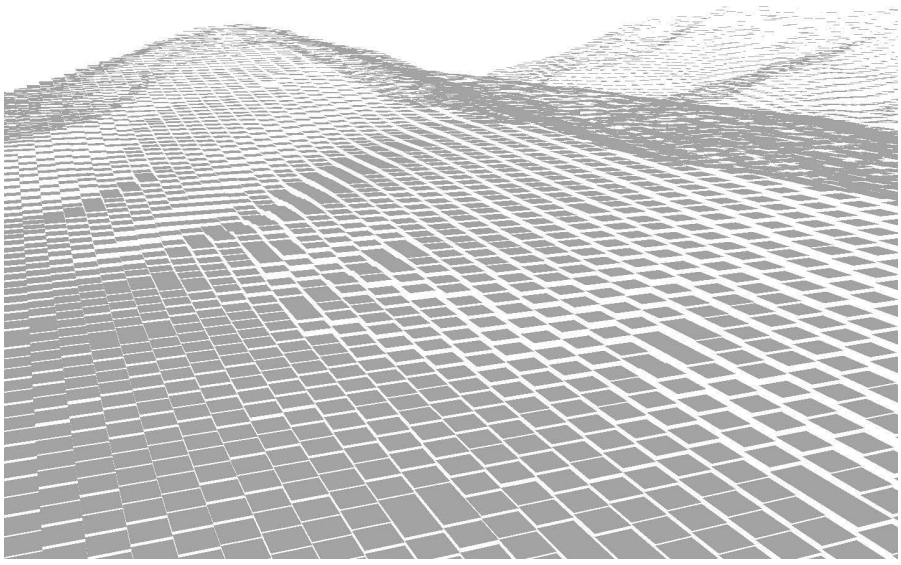


FIGURE 4.6 – Capture d'écran du MNT en couche vecteur.

4.2.3 Création des images de synthèse

Nous présentons dans cette section la création des images de synthèse. Ensuite nous listons les intervalles de confiance considérés pour l'appariement de lignes.

4.2.3.1 Images de synthèse

Nous avons pris pour chacune des positions des photographies du jeu de données d'apprentissage une série d'images de synthèse.

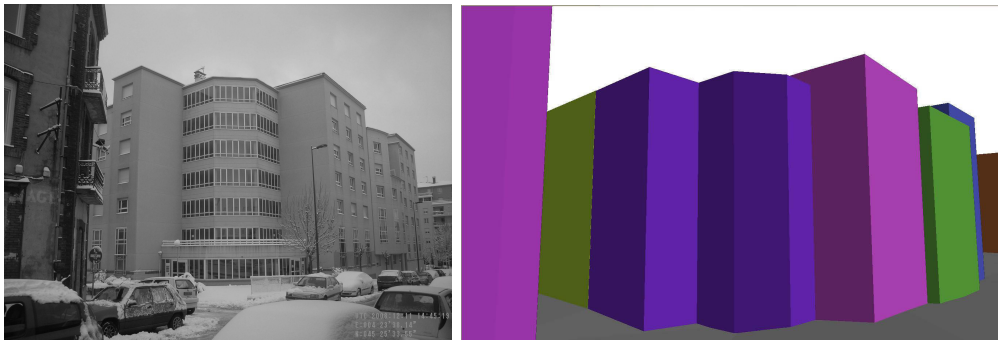
Nous insérons les paramètres de localisation de chaque photographie dans le formulaire approprié dans ArcScene (figure 3.10). Cette opération permet de retrouver la position de la photographie et de reconstituer une image virtuelle ayant le même contenu potentiel que la photographie.

Dans quelques cas, nous optimisons ce rendu en réajustant manuellement le positionnement (localisation et orientation) dans l'environnement virtuel pour reproduire ce qui est visible exactement sur la photographie. Cette opération est nécessaire car nous ne stockons pas la direction de visée pour toutes les photographies et à cause des imprécisions parfois importantes dans la localisation par le GPS en milieu urbain.

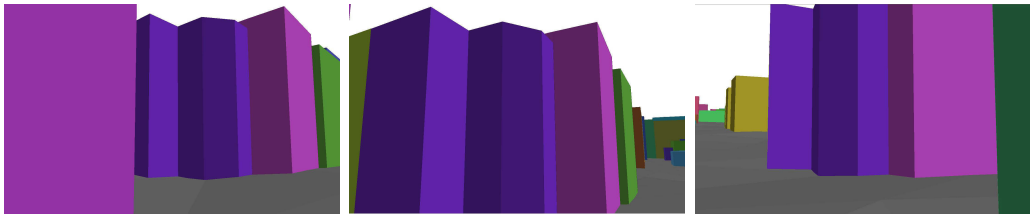
En plus d'une image de synthèse correspondant à chacune des photographies, nous avons pris 1, 2 ou 3 autres images, soit dans la même position avec une orientation ou une focale différente, soit par décalage d'un ou deux mètres de la position constatée de la photographie. Nous avons donc un ensemble de 2, 3 ou 4 images de synthèse que nous affectons à chaque photographie.

Le rendu 3D du monde virtuel est exporté pour fournir une image exploitable.

La base d'apprentissage est constituée de 20 photographies. Nous avons créé 46 images de synthèse prises aux environs de ces photographies. En moyenne nous avons pris 2.3 images de synthèse pour chacune des positions des prises de vue photographiques.



(a) Photographie (b) Image correspondante



(c) Focale modifiée (d) Orientation modifiée (e) Position modifiée

FIGURE 4.7 – Exemple d'images de synthèse prises aux alentours d'une photographie.

Parmi ces 46 images de synthèse, 20 sont très proches de l'une des photographies. Les 26 autres en différent par l'angle de visée, la focale ou l'orientation (figure 4.7).

La base de données d'images de synthèse que nous utilisons est composée de 73 images de synthèse prises dans 29 positions géographiques différentes (figure 4.3) et correspondent aux 33 photographies prises en deux phases : apprentissage et évaluation.

4.2.3.2 Intervalles de confiance

Comme nous l'avons expliqué dans le chapitre précédent, un appariement entre deux lignes est constaté si ces deux lignes partagent une même distance ρ et un même angle θ. Ces deux paramètres sont illustrés dans la figure 3.3.

Concrètement, la distance ρ doit appartenir à l'intervalle centré sur la valeur ρ de la ligne dans l'image de synthèse. La largeur de cet intervalle est appelée l'intervalle de confiance.

Nous avons constaté que les lignes s'apparient trop peu avec un intervalle de confiance inférieur à 20 pixels. À partir des tests effectués, nous avons défini deux degrés d'incertitude dans l'appariement.

Deux lignes sont appariées si leur distance ρ est incluse dans l'intervalle $[\rho - 30, \rho + 30]$ et la variation de l'angle θ est inférieure à 0,2 rad.

Le second degré est plus strict avec ρ= 20 pixels et θ = 0,1 rad.

4.2.3.3 Décalages

Nous avons défini un décalage par translation des lignes détectées sur une image de synthèse afin de les chercher ensuite dans la photographie. Le premier décalage est de 20 pixels vers la droite ou vers le haut pour les lignes horizontales. Nous avons appelé ce décalage 'décalage à droite'.

Le second décalage est de 20 pixels dans l'autre sens. Le décalage négatif permet de faire une translation de 20 pixels de toutes les lignes verticales vers la gauche, et de 20 pixels vers le bas pour les lignes horizontales.

Ces deux décalages permettent d'obtenir une meilleur correspondance entre les images de synthèse et les photographies.

Nous avons collecté et traité un volume de données important mais les tests que nous avons effectués ne font pas apparaître de différence significative dans les résultats entre ces deux décalages.

4.3 Résultats

Nous avons utilisé le jeu d'apprentissage de photographies décrit précédemment. Chaque photographie a été comparée avec les 73 images de synthèse.

Nous avons détecté les lignes par la transformée de Hough. Les deux paramètres caractérisant chaque ligne sont stockés dans des tableaux. Ce sont ces tableaux de données que nous traitons dans l'outil Matlab.

L'appariement entre une photographie et une image de synthèse nous donne un taux d'appariement calculé par la formule 4.1.

$$Taux \ = \frac{Nombre\ des\ lignes\ communes}{X} \times 100 \tag{4.1}$$

X est le nombre minimal de lignes détectées sur la photographie permettant d'avoir le maximum possible de lignes en commun.

4.3.1 Durées de traitement

La procédure d'élimination du sol nécessite un peu moins de 11 minutes (635 secondes exactement). Ce traitement a été effectué sur MATLAB R2010a (version

de février 2010).

La machine utilisée est équipée d'un microprocesseur AMD Athlon 64 X2 Dual Core Processor 4800+. Ce double processeur est cadencé à 2.5 GHz. La machine est dotée de 3 Go de RAM.

La comparaison d'une photographie avec une image de synthèse nécessite entre 12 et 25 secondes de traitement. Ce délai est plus important en affichant une visualisation des étapes de traitement. Nous nous contentons de la visualisation des résultats finaux.

Le temps pour comparer une photographie avec la totalité de la base d'images de synthèse varie entre 15 et 30 minutes. Cette durée dépend de la complexité de la photographie et du nombre de lignes détectées.

4.3.2 Taux d'auto-reconnaissance

Nous appelons taux d'auto-reconnaissance le taux de correspondance entre une image de synthèse et la photographie qui lui correspond véritablement.

Nous présentons les taux d'auto-reconnaissance des 20 photographies de la phase d'apprentissage calculés en utilisant des intervalles de confiance larges : $\rho = 30$ pixels et $\theta = 0.2$ rad. C'est le résultat de la comparaison entre une photographie et l'image de synthèse que nous avons construite pour correspondre à cette dernière. Ces taux sont illustrés dans la figure 4.8.

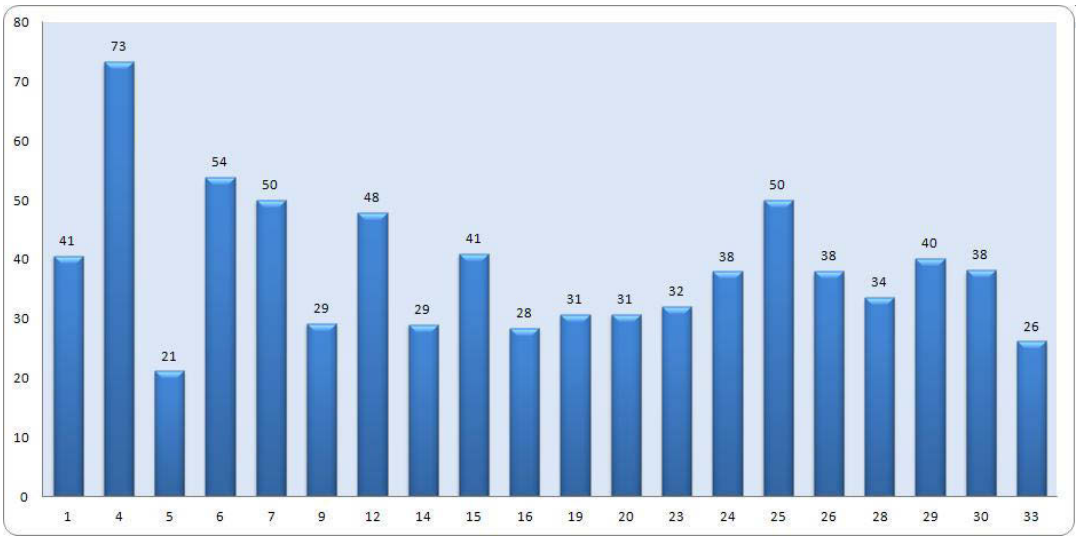


FIGURE 4.8 – Taux d'auto-reconnaissance (% pour chaque photographie).

Le taux minimal constaté dans notre base d'apprentissage est de 21.14 %. Le taux maximal est de 73.3 %, la valeur médiane est de 38 % et l'écart type est de 13.56 %.

Nous constatons que la majorité des taux est comprise entre 25 % et 45 % de reconnaissance. La seule exception est le pic de la deuxième image dont les lignes détectées correspondent pour 73.3 % aux lignes détectées sur l'image de synthèse correspondante.

La valeur moyenne des taux est de 38.59 %. La faiblesse de ces taux s'explique par le grand nombre de lignes parasites que contiennent les photographies.

En utilisant des intervalles de confiance plus stricts avec $\rho = 20$ pixels et $\theta = 0.1$ rad nous constatons des valeurs encore plus faibles de 10 à 30 %. Cela s'explique par la nature des photographies en milieu urbain. Les bâtiments sont habituellement distants de quelques mètres par rapport au point de prise de vue et 20 pixels (1 cm sur la photographie) représentent une distance de confiance relativement faible. Dans notre test, les bâtiments sur une photographie ont une largeur qui va de 7 à 15 cm. Une minuscule rotation avec une focale large représente plus de 2 cm de décalage dans la photographie.

En utilisant les images avant l'élimination du sol, nous constatons que les taux d'auto-reconnaissance sont plus faibles que pour celles dont le sol a été exclu. Il existe quatre photographies où, à l'inverse, l'élimination du sol donne un taux inférieur ou égal au taux d'auto-reconnaissance de la photographie. En analysant ces quatre exceptions, nous constatons que cela est dû à une mauvaise appréciation du sol dans la photographie. Le module détecteur de sol génère dans certains cas des lignes parasites. Nous utilisons dans ces quatre cas les images sans l'éliminer puisque cette opération génère plus de lignes parasites.

4.3.3 Arbre de décision

Une photographie est appariée logiquement à l'image de synthèse ayant le taux de reconnaissance maximal. Nous avons constaté que dans certains cas les deux ou trois valeurs les plus élevées sont assez proches alors qu'une seule d'entre elles représente le bon appariement. Nous expliquons le processus de choix des candidats au bon appariement par quelques exemples et un tableau récapitulatif.

1er exemple

Le premier exemple que nous donnons est celui de la photographie n°7 (figure 4.9). Nous constatons que cette photographie est appariée avec l'image de synthèse 's72' avec un taux de 50 %. C'est le taux maximal constaté pour cette photographie. Dans ce cas on pourrait affecter directement cette image de synthèse à la photographie '7'. Cette situation ne se produit pas toujours. La vérification d'autre paramètres est nécessaire.

La moyenne des taux est de 22.8 % ce qui est inférieur au taux d'auto-reconnaissance. L'écart type est de 12.6. La valeur du taux d'auto-reconnaissance est située à plus

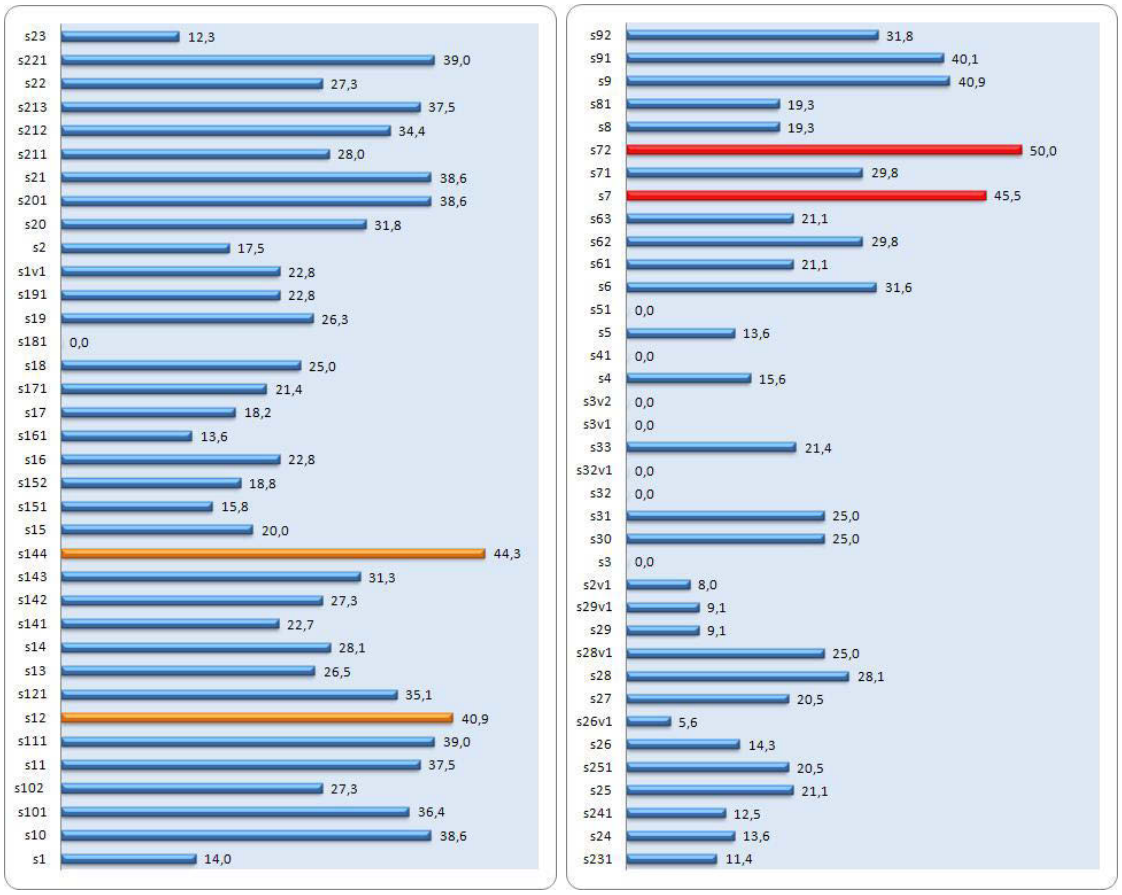


FIGURE 4.9 – Taux d'appariement de la photographie 7 avec les 73 images de synthèse.

d'un écart type de la moyenne. Nous considérons alors que toutes les valeurs supérieures à cette valeur (22.8 + 12.6 = 35.4) sont des taux très élevés.

Nous avons 37 images de synthèse ayant un taux d'appariement avec la photographie '7' supérieur à la moyenne des taux et 14 images de synthèse avec un taux qui s'écarte de plus d'un écart type de la moyenne. Nous nous contentons ici d'illustrer les 3 images ayant les taux les plus élevés en plus de la bonne image.

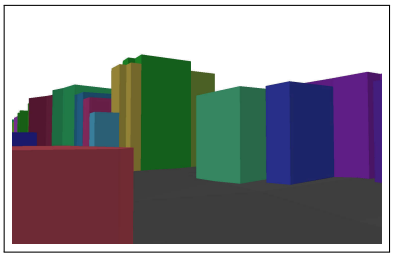
Nous illustrons dans la figure 4.10 le cas de la photographie '7' et de ses quatre images de synthèse.

En interprétant la figure 4.10, nous pouvons dégager les éléments :

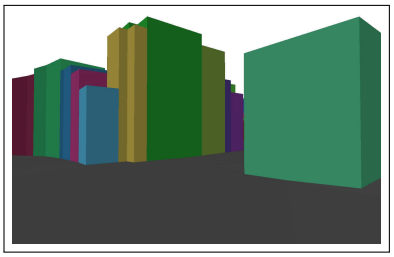
- 's72' a été positionnée manuellement pour bien correspondre à la photographie.
- 's7' a été localisée au même emplacement que l'image de synthèse 's72' mais avec une focale plus serrée.
- 's144' affiche un taux assez élevé de correspondance avec notre photographie même si elle n'avoisine pas l'image '7'. Cette ressemblance est due à la texture des bâtiments de la photographie et aux obstacles présents sur la photographie



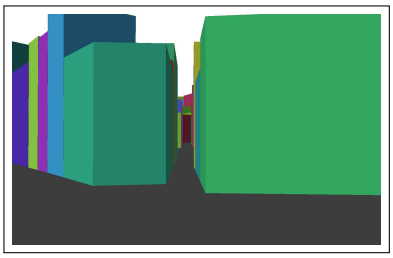
(a) Photographie ’7’



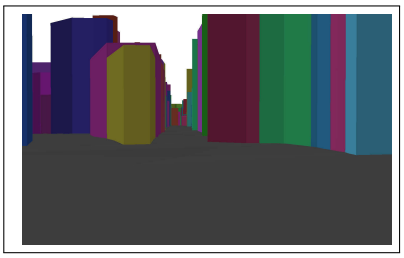
(b) Image ’s72’(taux=50)



(c) Image ’s7’(taux=45.45)



(d) Image ’s144’(taux=44.3)



(e) Image ’s12’(taux=40.91)

FIGURE 4.10 – Les quatre images de synthèse les mieux appariées avec la photographie ’7’.

(arbre, poteau électrique...).

– ’s12’ est positionnée loin de la position de la photographie ’7’. Cependant elle affiche un taux de correspondance assez élevé avec notre photographie (40.91). Cette ressemblance est due elle aussi aux parasites de la photographie qui permettent de détecter un nombre de lignes ne correspondant pas aux limites de bâtiments.

À titre indicatif, les positions de la photographie ’7’ et ’14’ sont présentées dans la figure 4.11. L’image ’s144’ est la quatrième image de synthèse prise dans une position très proche de celle de la photographie ’14’. La distance entre la photographie ’7’ et la photographie ’14’ est de 994 mètres. La distance entre la position ’7’ et celle de l’image ’s12’ est de 74.8 mètres. Le rôle de la distance est analysée dans les scénarios

d'utilisation dans le chapitre suivant.



FIGURE 4.11 – Positions des photographies '7' et '14' sur la carte de Saint-Étienne.

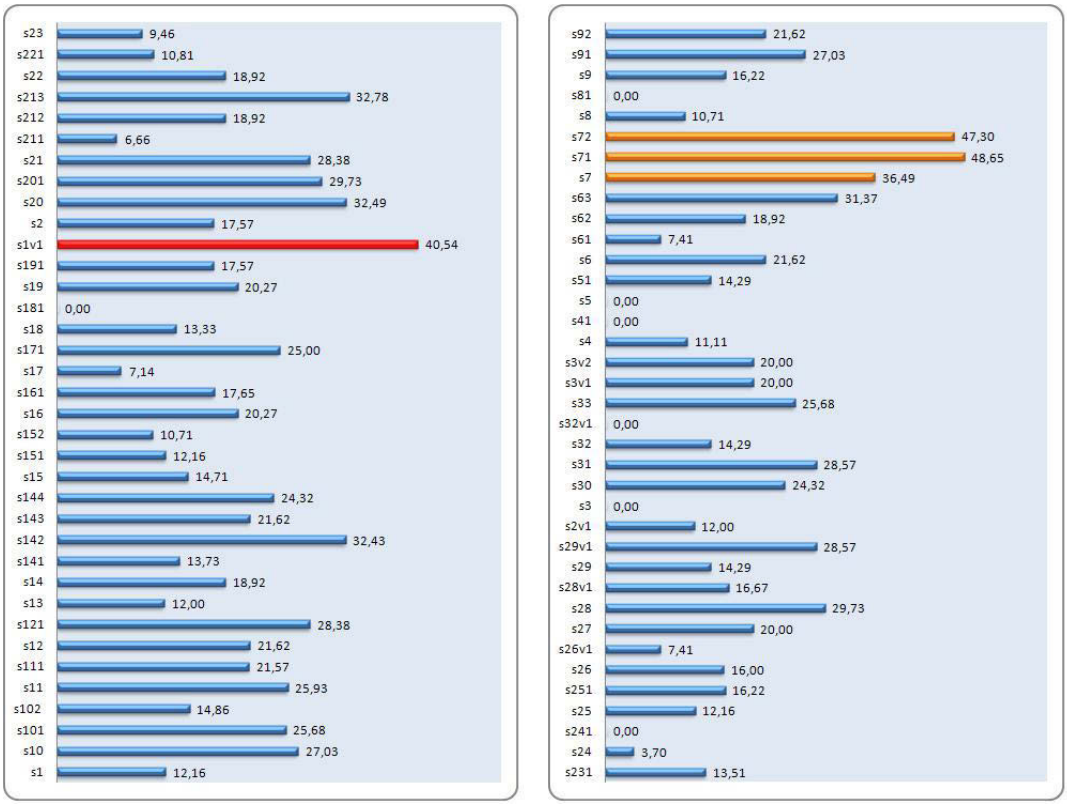
2ème exemple

Un second exemple est celui de l'image '1' qui s'auto-apparie avec un taux de 40.54 %. Ce taux est visible sur le graphique global des taux d'auto-reconnaissance présent dans la figure 4.8 et illustré dans la figure 4.12. Cette dernière figure illustre les différents taux d'appariement de la photographie '1' avec les 73 images de synthèse.

Nous pouvont constater que le taux maximum d'appariement est de 48 % relatif à l'image de synthèse 's71' qui est une des trois images de synthèse prises à la position de la photographie '7'.

Le taux moyen est de 18.4 % et l'écart type est de 11. Nous avons 12 images de synthèse qui s'apparient avec un taux qui s'écarte de plus d'un écart type de la moyenne (29.4).

Le taux d'appariement de notre photographie avec son équivalent est supérieur à ce taux et se présente dans le second ordre d'appariement. Dans notre cas, c'est



(a) (b)

FIGURE 4.12 – Taux d'appariement de la photographie '1'.

la 's1v1' qui détient 40.54 % de lignes appariées avec ceux de la photographie '1'.

Image	Taux d'appariement	Distance (m)
s71	49	213
s72	47	213
s1v1	41	0
s20	36	501
s7	36	213
s213	34	408
s142	32	1200
s63	31	169
s201	30	501
s28	30	197
s29v1	29	178

TABLE 4.1 – Résultat d'appariement de la photographie '1'.

Pour avoir une idée de l'éloignement des candidats à la position de la photogra-

phie '1', le tableau 4.1 présente les distances qui séparent les images de synthèse de la photographie. Ce tableau liste seulement les images appariées avec un taux qui s'écarte de plus d'un écart type de la moyenne (29.4).

Notons que 's1v1' est la seconde image de synthèse prise dans la position de la photographie '1' en changeant l'angle de vue par rapport à l'image de synthèse 's1'.

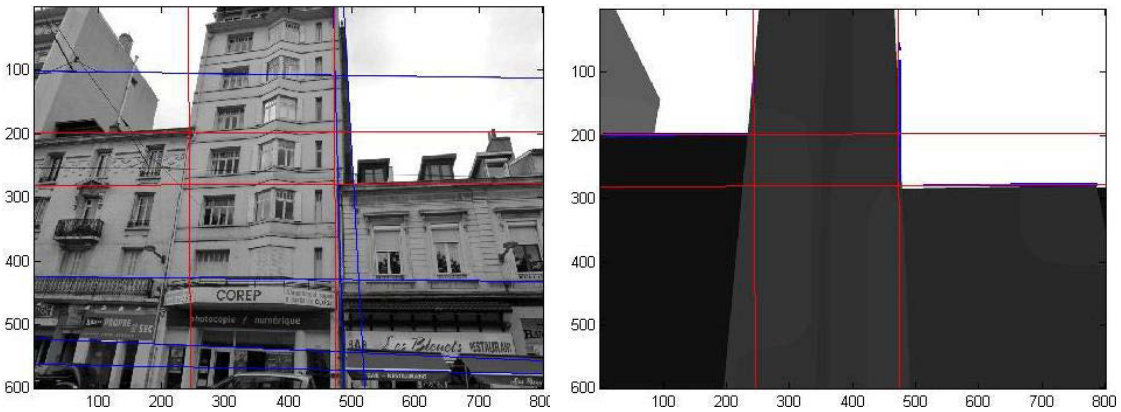
3ème exemple

Nous analysons un troisième exemple d'identification de positionnement de photographie. La figure 4.13 illustre une photographie avec les lignes détectées. Nous colorons en rouge les lignes communes à la photographie et à l'image de synthèse correspondante (4.13b). Les lignes bleues sont celles que nous ne trouvons pas dans l'image de synthèse en question. Nous utilisons le même code couleur et affichons les lignes détectées sur l'image de synthèse.

Le calcul du taux d'appariement peut se faire dans ce cas d'un façon manuelle. Nous observons dans la figure 4.13a les quatre lignes rouges détectées dans l'image de synthèse parmi les 10 lignes détectées dans la photographie '29'.

Le calcul du taux est donné par la simple opération : $\frac{4}{10} \times 100 = 40$ %.

En effectuant l'appariement des lignes de cette photographie avec les 73 images de synthèse nous calculons tout d'abord le taux moyen qui vaut 20.7 % et l'écart type qui est de 11.3. Dans ce cas aussi, l'image de synthèse correspondante est appariée avec un taux qui s'écarte de plus d'un écart type de la moyenne (40 > (20.7+11.3)).



(a) Photographie '29'. (b) Image de synthèse correspondante.

FIGURE 4.13 – Exemple d'image de synthèse prise aux alentours d'une photographie.

Conclusion

L'image de synthèse correspondant véritablement à une photographie est dans toutes les photos du jeu d'apprentissage appariée avec un taux supérieur à la moyenne

des taux. Dans 18 sur 20 cas ce taux s'écarte de plus d'un écart type de la moyenne.

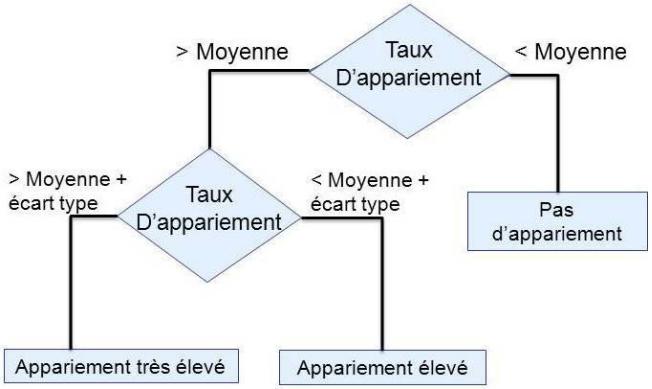


FIGURE 4.14 – Arbre de décision.

Nous avons alors un arbre de décision assez simple permettant de déterminer les candidats en sélectionnant les images ayant un taux supérieur à la moyenne des taux.

La figure 4.14 présente un arbre de décision ne prenant en compte que le paramètre taux d'appariement. Selon la valeur de la moyenne des taux et de l'écart type nous considérons qu'une image de synthèse est candidate à un bon appariement ou pas.

La zone géographique de recherche délimite aussi les candidats en fonction du degré d'imprécision. Nous détaillons cet élément dans le deuxième scénario du chapitre suivant.

4.3.4 Les faux positifs

Nous appelons faux positif toute correspondance donnée par l'appariement effectué entre une image de synthèse et une photographie qui ne concerne pas la même scène.

Une image de synthèse est considérée comme faux positif si elle s'apparie avec une photographie avec un taux d'appariement légèrement inférieur au taux d'auto-reconnaissance relatif à une photographie. Nous avons choisi de définir cette limite par le taux d'auto-appariement moins 10 %. La limite supérieure est naturellement 100 %.

Concrètement si une image s'auto-apparie à 40 % alors toutes les images de synthèse ayant un taux d'appariement compris entre (40-4=36) et 100 % sont considérées comme faux positifs, en dehors de la bonne image de synthèse.

Ce phénomène apparaît principalement dans le cas des photographies qui contiennent beaucoup d'éléments perturbateurs comme les arbres ou des textures prononcées. Ces éléments génèrent des lignes que nous retrouvons dans l'image de synthèse.

Nombre de photos	Nombre de faux positifs	Distance moyenne en m
3	0	—
3	1	173
3	2	494
6	3	285
3	4	521
2	8	179
20	18	Total

TABLE 4.2 – Nombre des faux positifs.

Ce phénomène est constaté dans 17 photographies parmi les 20 qui constituent notre base de d'apprentissage. Cela représente 85 % des cas mais avec différents degrés.

Ces photographies possèdent un seul ou plusieurs faux positifs. Le maximum qu'on a pu dégager est de 8 faux appariements. La moyenne est de 2.75 images de synthèse appariées alors qu'elles ne correspondent pas à la scène véritablement.

Le tableau 4.2 récapitule le nombre de faux positifs détectés. Par exemple, dans notre base d'apprentissage, nous comptons 3 photographies avec un seul faux positif. La distance moyenne séparant la position de chacune des photographies avec ce faux positif est de 173 mètres.

L'annexe D présente le tableau complet des distances séparant chaque photographie de ses éventuels faux positifs détectés.

Nous revenons sur la position géographique dans le deuxième scénario du chapitre suivant.

Statistiques sur la base d'apprentissage

En considérant les taux d'appariement avec connaissance à priori du bon appariement, nous avons eu dans 15 % des cas une seule solution d'appariement (pas de faux positifs). Dans 15 % des cas un seul appariement en plus du bon est détecté. Ces mauvaises détections concernent des images de synthèse prises en moyenne à 173 mètres de la bonne photographie.

Dans 10 % des cas les candidats d'appariement sont au nombre de 8 qui s'éloignent de la photographie correspondante en moyenne de 179 mètres.

Nous donnons ces distances à titre indicatif et nous détaillons cet éloignement dans les deux premiers scénarios d'utilisation. Rappelons que le seul paramètre utilisé est le taux d'appariement qui nous permet de retrouver une image de synthèse correspondante à une photographie.

Nous donnons ici les statistiques relatives à la base d'apprentissage. Les chiffres

globaux sont présentés dans le paragraphe suivant.

Sur les 55 faux positifs trouvés dans la base d'apprentissage, nous avons uniquement deux images qui sont distantes de moins de 20 mètres de la bonne position recherchée. Cela représente 3 % des faux positifs et 0.13 % de tous les appariements effectués dans notre base d'apprentissage.

Si nous cherchons une localisation d'une photographie avec une incertitude de 50 mètres, nous avons 0.47 % de chance d'avoir plusieurs candidats en plus de la bonne localisation.

Nous avons détecté 16 faux positifs autour du bon appariement dans un rayon de 100 mètres. Ces faux positifs représentent environ 1 % des appariements effectués.

Statistiques générales

Les 83 images de synthèse sont réparties dans 39 positions comme l'illustre la figure 4.15 [1]. On remarque que la densité des images de synthèse est suffisante dans quelques zones et que la couverture est discontinue dans d'autres espaces.

Nous étudions trois niveaux d'imprécision de la localisation d'une photographie. Ces nivaux dépendent de la zone géographique où nous estimons que notre photographie est prise. Le premier niveau que nous testons est dans un rayon de 200 mètres, le second à 100 mètres et le dernier à 50 mètres.

Cette imprécision se traduit par une prise en compte uniquement des images de synthèse localisées dans un disque de rayon 200, 100 ou 50 mètres et centré sur la position supposée.

Si nous avons une incertitude de 50 mètres autour d'une position, en appliquant notre arbre de décision, nous avons une distance moyenne de 12 mètres entre la bonne position de la photographie et le barycentre pondéré (voir chapitre 5) calculé.

L'imprécision de 100 mètres engendre en moyenne 29 mètres séparant la bonne position et celle donnée par notre approche.

Dans un cercle de rayon 200 mètres, nous pouvons trouver en moyenne la position d'une photographie à 55 mètres près de sa position réelle.

Si nous n'avons pas de localisation, nous prenons toutes les images de synthèse ayant un taux de reconnaissance supérieur à la moyenne, nous avons en moyenne un décalage de 218 mètres.

Ces statistiques ont été effectuées sur la totalité des photographies dont nous disposions, sauf la photographie '13' qui ne permet pas d'avoir plus de 5 % d'appariement avec son image de synthèse correspondante. Nous l'avons exclue car cette photographie fait apparaitre plusieurs bâtiments distants l'un de l'autre avec une forte présence d'obstacles.

1. La position de l'image '14' n'est pas visible dans la figure pour permettre d'avoir un bon niveau de zoom sur les autres positions.



FIGURE 4.15 – Couverture des images de synthèse de la zone des prises de photographies.

Les distances séparant les positions des 32 photographies avec la position calculée en appliquant notre approche sont données dans le tableau 4.3. La colonne 'Toute la carte' prend en charge toutes les images de synthèse qui sont englobées dans un rectangle de dimension 250 X 1200 mètres.

4.4 Limites de la méthodologie

Cette section présente les difficultés que rencontre notre méthodologie d'interfaçage par appariement de lignes. Nous distinguons tout d'abord trois éléments qui

Photo	Toute la carte	200	100	50
1	332	74	26	26
2	352	114	0	0
3	335	142	6	6
4	311	70	19	19
5	143	79	50	35
6	385	51	5	5
7	181	45	41	14
8	186	64	62	19
9	97	46	30	0
10	94	67	50	23
11	40	50	48	0
12	91	39	29	0
14	855	0	0	0
15	90	63	45	11
16	236	79	34	6
17	304	98	39	0
18	90	59	28	0
19	231	48	13	6
20	175	10	25	9
21	49	32	51	0
22	242	23	15	16
23	88	69	54	18
24	131	26	37	24
25	186	5	5	8
26	105	41	35	5
27	70	14	26	17
28	138	41	15	16
29	129	42	27	14
30	332	85	39	22
31	265	31	23	23
32	336	85	29	12
33	376	72	20	20
Moyenne	218	55	29	12

TABLE 4.3 – Distances en mètre entre les positions des photographies et les positions calculées par notre méthode.

perturbent le bon déroulement de la reconnaissance des photographies parmi les images de synthèse.

Les bases de données géographiques sont la première cause d'erreur dans l'appariement. En effet les bases de données fournies par l'IGN sont fondée sur le cadastre. Les bâtiments qui ont une fonction commune sont considérés comme un seul bâtiment. Dans un groupe de bâtiments servant à l'habitation, une différence de texture ou de couleur est apparente dans la réalité. Dans certains cas, la BD TOPO fusionne ces bâtiments en un seul. Ces groupements de bâtiments sont cependant moins fréquents dans les nouvelles bases de données ce qui atténue la perturbation de notre méthodologie. La figure 4.16 illustre un exemple de ce phénomène.

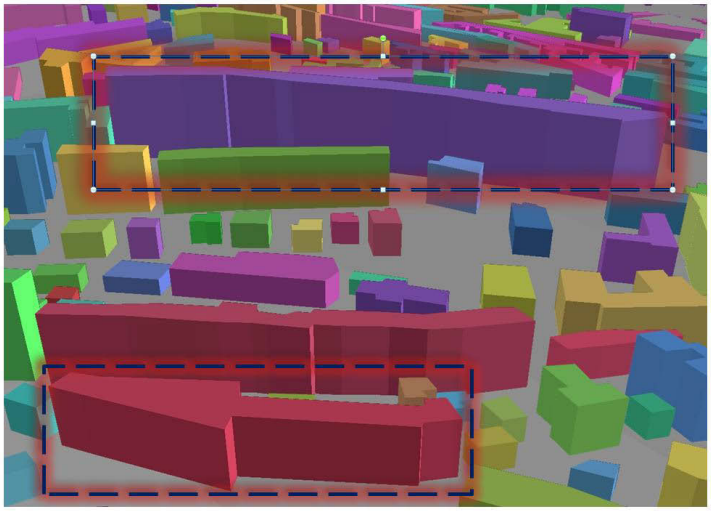


FIGURE 4.16 – Groupement de bâtiments de la BD TOPO. (Source : capture d'écran d'ArcScene. ESRI)

Le deuxième élément perturbateur est le cas des prises de vues dans des situations particulières. Notre méthodologie fonctionne bien en se positionnant au sol et en prenant des photographies à l'horizontal. En se mettant dans des situations spéciales : prise de photographie d'un étage supérieur d'un bâtiment, orientation de l'appareil photographique par rapport à l'axe horizontal... nous obtenons des photographies non disponibles sur la base de données des images de synthèse.

Nous en concluons que notre méthode n'est applicable que si les photographies sont prises approximativement à l'horizontal.

La présence de plusieurs obstacles handicape aussi notre approche. Les décorations festives et la neige sur les arbres ne font qu'accentuer les perturbations et par conséquence le nombre de lignes parasites.

Finalement les lignes qui délimitent les bâtiments sont l'élément de base que nous utilisons pour les comparer et les apparier. Nous avons constaté que des bâtiments différents par la texture et la position géographique mais qui partagent un même

gabarit génèrent un taux d'appariement élevé. La focale de l'appareil photographique et la distance à ces édifices fournissent des bâtiments ayant une même largeur et longueur alors que la réalité ne confirme pas cette similitude. Cet effet fournit des faux appariements et des faux positifs. Dans notre base de photographies, nous avons détecté deux photographies qui ont été appariées avec des images de synthèse présentant ce phénomène.

4.5 Conclusion

La validation de notre méthodologie d'interfaçage par appariement de lignes a été présentée dans ce chapitre en détaillant notre base de données photographique d'apprentissage ainsi que l'échantillon que nous avons pris de la base de données géographique de la ville de Saint-Étienne.

Nous avons démontré par l'exemple que les lignes sont bien représentatives d'un paysage urbain puisqu'elles offrent la possibilité d'apparier une photographie urbaine avec une image de synthèse la représentant.

La recherche d'une photographie parmi les 83 images de synthèse créées pour la validation du concept montre un appariement assez pertinent pour approcher la position de cette photographie. Ce positionnement est la clé pour caractériser le contenu photographique.

Chapitre 5 : Scénarios d'utilisation fondés sur la distance

Chapitre 5

Scénarios d'utilisation fondés sur la distance

Nous exposons trois scénarios d'utilisation de notre approche. Chaque scénario sera exposé à travers un ou plusieurs exemples concrets d'implémentation.

La démarche décrite dans les sections précédentes permet d'évaluer le niveau de correspondance potentielle entre une photo capturée et une image de synthèse. Dans la phase précédente nous avions connaissance de la position et de tous les paramètres des images photographies.

Dans cette phase, nous faisons l'hypothèse que nous tentons de localiser une photographie après avoir généré une collection d'images de synthèse avec un intervalle régulier dépendant de la précision souhaitée. Une capture d'une image de synthèse tous les 2 mètres avec à chaque fois entre 4 à 8 captures prises dans les différentes orientations horizontales devrait convenir. L'objectif serait de localiser une photographie à partir de cette collection.

Si nous ne disposons pas de cette collection régulière, nous avons ajouté 10 images de synthèse pour mieux valider statistiquement nos résultats. Cela nous fait 83 images de synthèse qui seront comparées à chaque photographie de la base d'évaluation.

1er scénario

Le premier scénario d'utilisation de notre approche d'interfaçage des images par appariement de lignes est la vérification de la position géographique d'une photographie et de l'annotation automatique de son contenu.

Nous nous plaçons dans le cas où des photographies argentiques scannées ou numériques possèdent une géolocalisation supposée correcte donnée par GPS ou manuellement par pointage sur une carte. L'objectif de ce scénario est de valider

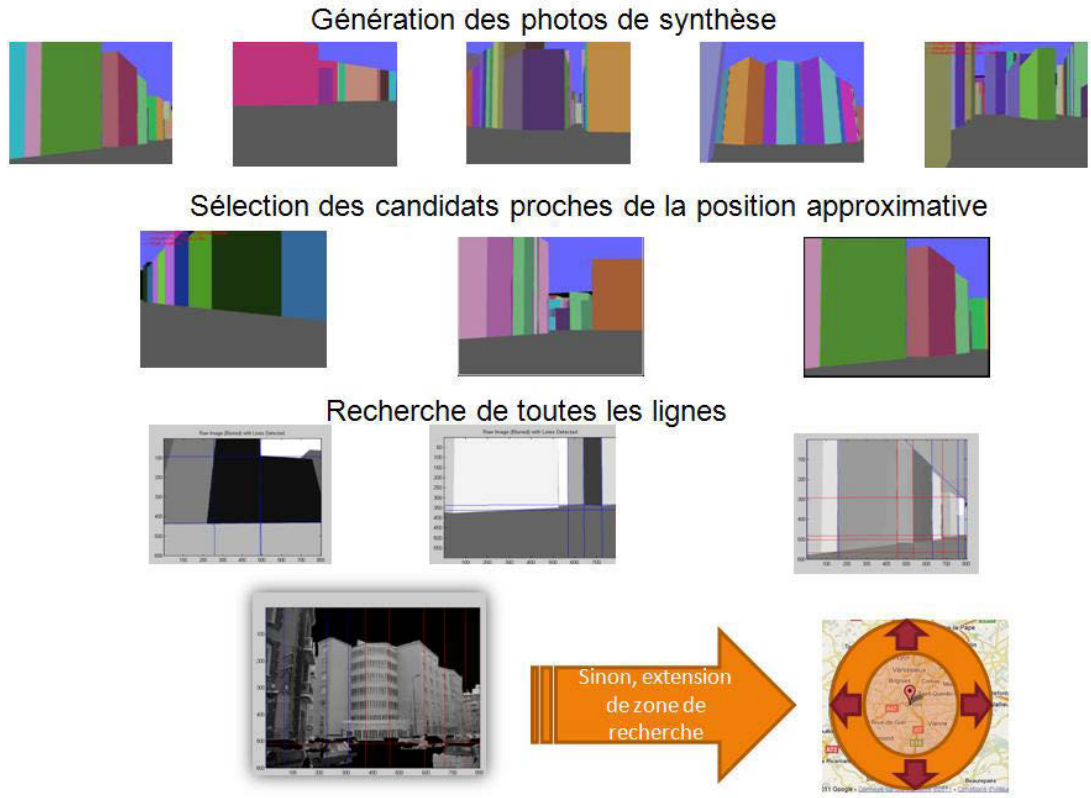


FIGURE 5.1 – Recherche autour d'une position donnée et extension de zone de recherche.

cette localisation. Cette dernière représente le point de départ pour sélectionner les images de synthèse candidates dans le voisinage proche de cette position. Nous utilisons l'incertitude donnée par le GPS. On sélectionne alors toutes les images de synthèse aux alentours de la localisation suggérée par l'utilisateur.

Ensuite on recherche les lignes de la photographie parmi les images de synthèse sélectionnées. A la fin de la procédure de recherche, si on ne trouve pas un bon appariement, on étend la zone de recherche. Cette extension élargit le périmètre de recherche centré sur la position connue de la photographie.

La recherche est relancée pour les images de synthèse se situant entre la zone déjà inspectée et la nouvelle frontière de recherche comme illustré dans la figure 5.1.

La validation du positionnement géographique de la photographie se fait en utilisant deux indicateurs : le taux d'appariement et le la distance séparant les images avec les meilleurs taux et position de la photographie. En combinant les deux, on peut donner un indicateur de la qualité de localisation de la photographie.

Une fois vérifiée sa position, le contenu de la photographie est décrit en utilisant des informations de la base de données géographique : nom de la rue, numéro ou le

nom de l'immeuble, POI associés au bâtiment...

Illustration du scénario

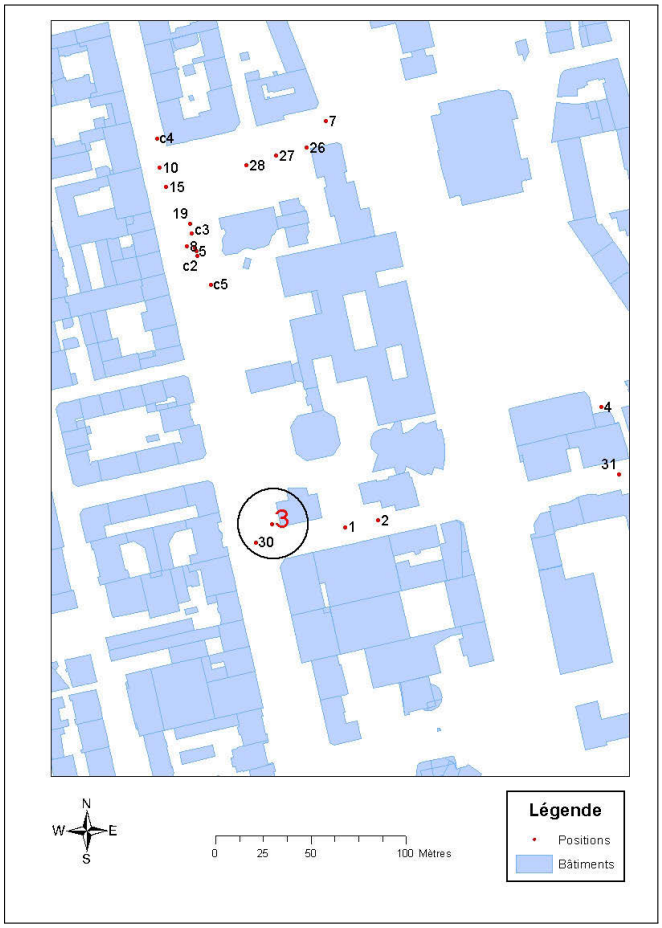


FIGURE 5.2 – Disque autour de la photographie '3' de rayon 18.7 mètres.

Nous avons pris la photographie '3' avec un appareil photographique équipé d'un GPS intégré, nous avons alors une géolocalisation précise de la position de la prise de vue. La position est basé sur le signal de cinq satellites de la constellation GPS et nous avons un DOP de 17/5= 3.4. En prenant la photo en mode 2D, nous avons un HDOP (voir section 1.1.2) donc une imprécision de 18.7 mètres. Nous effectuons la comparaison de cette photographie avec les images de synthèse qui se trouvent dans ses alentours à l'intérieur d'un cercle de rayon 18.7 mètres comme l'illustre la figure 5.2.

Nous remarquons que quatre images de synthèse (réparties sur deux positions) sont incluses dans le cercle centré sur la position enregistrée pour la photographie 3

et de rayon 18.7 mètres. Ces images sont s3, s3v1, s3v2 et s30. Les différents taux d'appariement de ces points sont affichés dans le tableau 5.1.

Image	Taux d'appariement	Distance par rapport à la position 3 (m)
s3v2	36,33	0
s30	31,37	13,02
s3	16,67	0
s3v1	16,67	0

TABLE 5.1 – Résultat d'appariement de la photographie '3' avec les images de synthèse localisées dans un rayon de 18.7 mètres.

Nous remarquons bien que l'image de synthèse 's3v2' est appariée avec un taux supérieur à celle de l'image 30 et aux deux autres images prises à la bonne position (celle de la photo 3) mais avec des orientations différentes.

Le taux moyen général est de 26.4 et l'écart type est de 13. Les taux d'appariement de la photographie '3' avec l'image 's3v2' et l'image 's30' sont supérieurs à la moyenne. L'appariement avec l'image 's3v2' est à moins d'un écart type de la moyenne ($36.33 < 39.4$).

En comparant les taux de ces deux candidats (s3v2 et s30), nous pouvons dire que l'image 's3v2' est plus probablement celle qui a été prise dans l'emplacement 3 et que le contenu de cette dernière est conforme à celui de l'image de synthèse 's3v2'.

2ème scénario

Le deuxième scénario s'intéresse au cas où nous ne possédons pas des coordonnées de la prise de vue mais que nous savons approximativement où elle a été prise.

L'utilisateur peut déclarer un quartier ou une position approximative d'une photographie ancienne scannée ou récente et l'objectif est de lui proposer une localisation probable de la photographie en s'appuyant sur les images de synthèse du voisinage.

Nous avons dans ce cas un seul critère de recherche de positionnement dans notre zone : le taux d'appariement.

Les candidats sont sélectionnés en utilisant des méthodes statistiques permettant de déterminer le groupe de valeurs le plus signifiant en termes de taux de reconnaissance.

En se référant à l'arbre de décision, nous cherchons le groupe des candidats. Nous sélectionnons le lot d'images dont le taux d'appariement est supérieur à la moyenne de tous les taux d'appariement entre la photographie et les images de synthèse de la zone géographique sélectionnée.

Le résultat final de cette démarche est indicatif et dépend de la localisation du groupe d'images de synthèse correspondantes. La situation idéale est d'avoir des

images de synthèse prises en continue avec un intervalle régulier ce qui permettrait d'avoir une approximation plus efficace de la localisation d'une photographie.

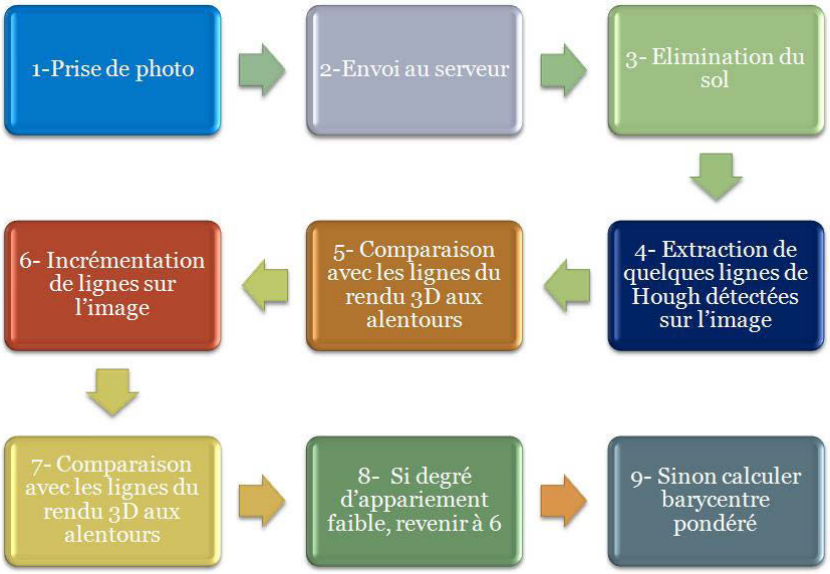


FIGURE 5.3 – Démarche globale du second scénario.

La position suggérée pour la photographie est placée au barycentre pondéré par les taux d'appariement des images de synthèse de la zone. Ce barycentre dépend des taux d'appariement et de l'éloignement des images appariées avec notre photographie.

Le processus de ce scénario est décrit globalement par la figure 5.3.

La position du barycentre pondéré avec le taux d'appariement t_i est calculée selon la formule 5.1.

$$\bar{X}_t = \frac{\sum_{i=1}^{n} t_i x_i}{\sum_{i=1}^{n} t_i}, \bar{Y}_t = \frac{\sum_{i=0}^{n} t_i y_i}{\sum_{i=1}^{n} t_i} \tag{5.1}$$

La distance standard mesure le degré de concentration ou de dispersion des points. Elle peut être pondérée par un poids, dans notre situation c'est le taux d'appariement. C'est une statistique utile puis qu'elle est similaire à la façon dont un écart-type mesure la distribution des valeurs autour de la moyenne statistique. Statistiquement, 68 % des points sont situés dans le cercle de rayon égal à une distance standard, 95 % dans le cercle avec un rayon de 2 distances standard et 99 % des points sont à moins de 3 distances standard.

Une photographie ancienne s'intègre dans ce scénario puisqu'elle ne possède pas de position géographique exacte. Il faut noter que si les photographies sont très

anciennes, on ne dispose que rarement des modèles 3D des bâtiments nécessaires pour les apparier. Toutefois on voit apparaitre des tentatives de modélisation historique des volumes des bâtiments.

Première illustration du scénario 2

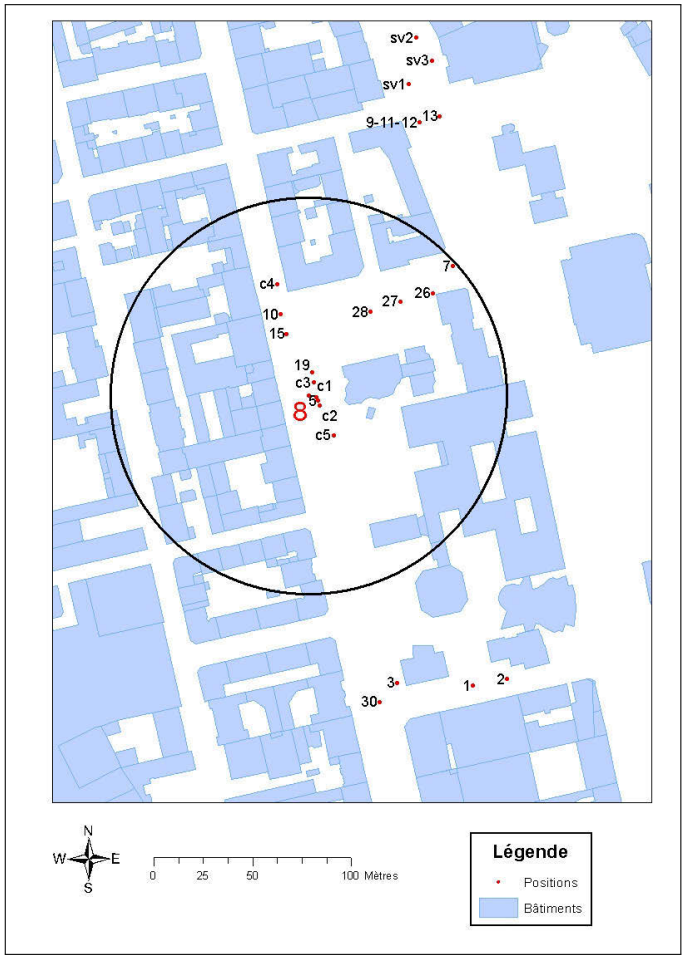


FIGURE 5.4 – Images dans une zone de 100 mètres autour de la photographie '8'.

Nous considérons la photographie '8' localisée à proximité de sa position réelle.Nous testons sa localisation dans un cercle de rayon 100 mètres autour de cette position. Nous prenons donc toutes les images de synthèse situées dans cette zone comme l'illustre la figure 5.4.

Nous trouvons 25 images de synthèse dans ce cercle. Les taux relatifs à ces images de synthèse ainsi que la distance les séparant réellement de la position de l'image '8' sont établis dans le tableau 5.2.

En se fondant sur notre arbre de décision, nous considérons les taux supérieurs

Image	Taux d'appariement	Distance
s26v1	42,86	81,44
s72	42,42	98,19
s28	39,39	52,64
s7	32,00	98,19
s8	30,30	0,00
s71	30,30	98,19
C3	25,00	7,12
s28v1	21,21	52,64
s101	20,00	43,55
c5	17,65	23,71
s26	16,60	81,44
s10	16,00	43,55
C1	15,91	4,06

Image	Taux d'appariement	Distance
c4	15,91	58,66
s81	15,91	0,00
s15	12,12	33,14
s151	10,00	33,14
s27	8,00	66,62
c2	7,41	7,32
s152	7,40	33,14
s102	6,06	43,55
s5	0,00	5,50
s51	0,00	5,50
s29	0,00	11,88
s29v1	0,00	11,88

TABLE 5.2 – Résultat d'appariement de la photographie '8' dans un rayon de 100 mètres.

à la moyenne. Le taux moyen est de 15.63. Nous considérons alors le barycentre pondéré des 15 images ayant les taux d'appariement supérieurs à la moyenne.

Ces 15 images de synthèse sont réparties sur 14 positions et forment les sommets d'un polygone. Notre photographie est prise dans ce polygone et plus précisément proche du barycentre pondéré par le taux d'appariement de chacune des images.

La figure 5.5 illustre la position donnée par le calcul du barycentre ainsi que la distance standard pondérée par les taux d'appariements. Nous constatons que le barycentre pondéré se situe au centre de ce cercle.

La distance standard de la dispersion de nos localisations est présentée dans la figure 5.6 pour la comparer à la distance standard pondérée avec les taux d'appariements de la photographie 8 et de toutes les images de synthèse dans un rayon de 100 mètres. On peut considérer que la photographie a 68 % de chances de se trouver dans le cercle le plus petit, 95 % dans le cercle moyen et 99 % dans le cercle le plus grand.

La distance séparant le barycentre calculé de la bonne position de la photographie est de 40.8 mètres. Cette distance s'explique par l'importance des taux d'appariement avec les images aux positions 7, 26 et 28 qui sont à plus de 50 mètres de la bonne position et qui ont affiché une ressemblance apparente avec les lignes de notre

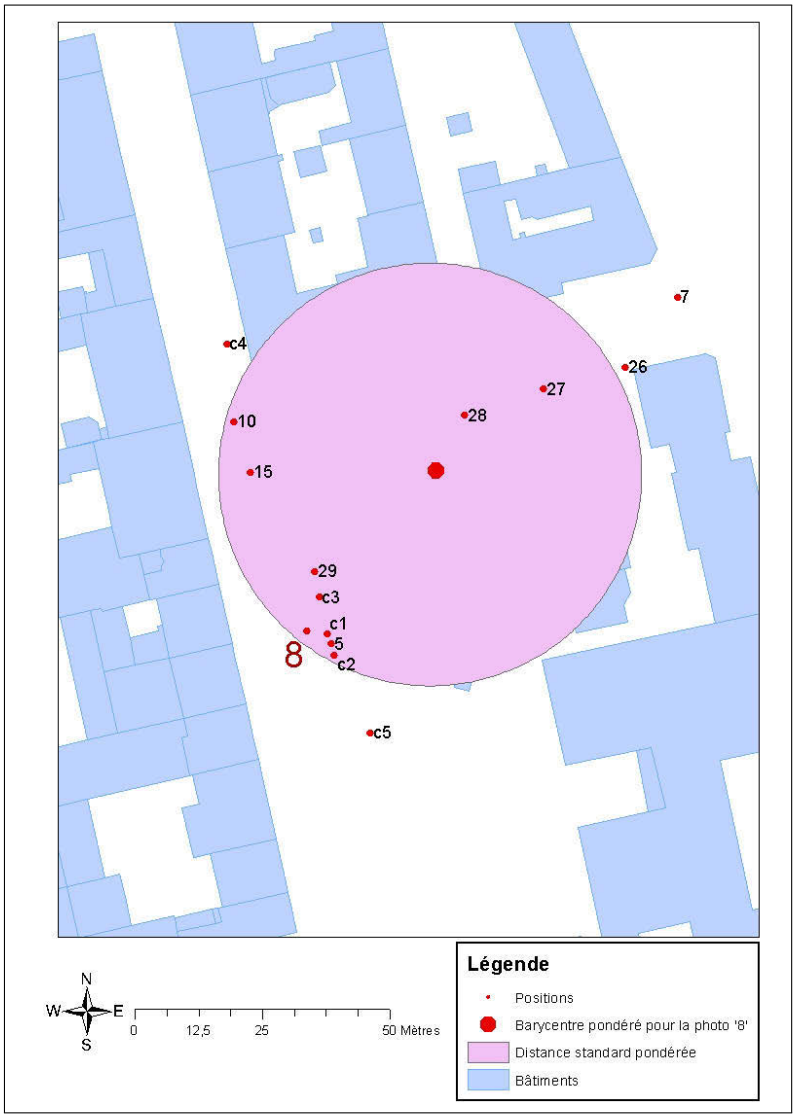


FIGURE 5.5 – Barycentre pondéré calculé pour la position de la photographe '8' et distance standard pondérée.

photographie.

En conclusion, dans le cas présent, une recherche sur un cercle de diamètre de 200 mètres fournit une imprécision constatée de 40.8 mètres.

Cette valeur serait plus faible si nous disposions d'une collection d'images de synthèse régulièrement espacées.

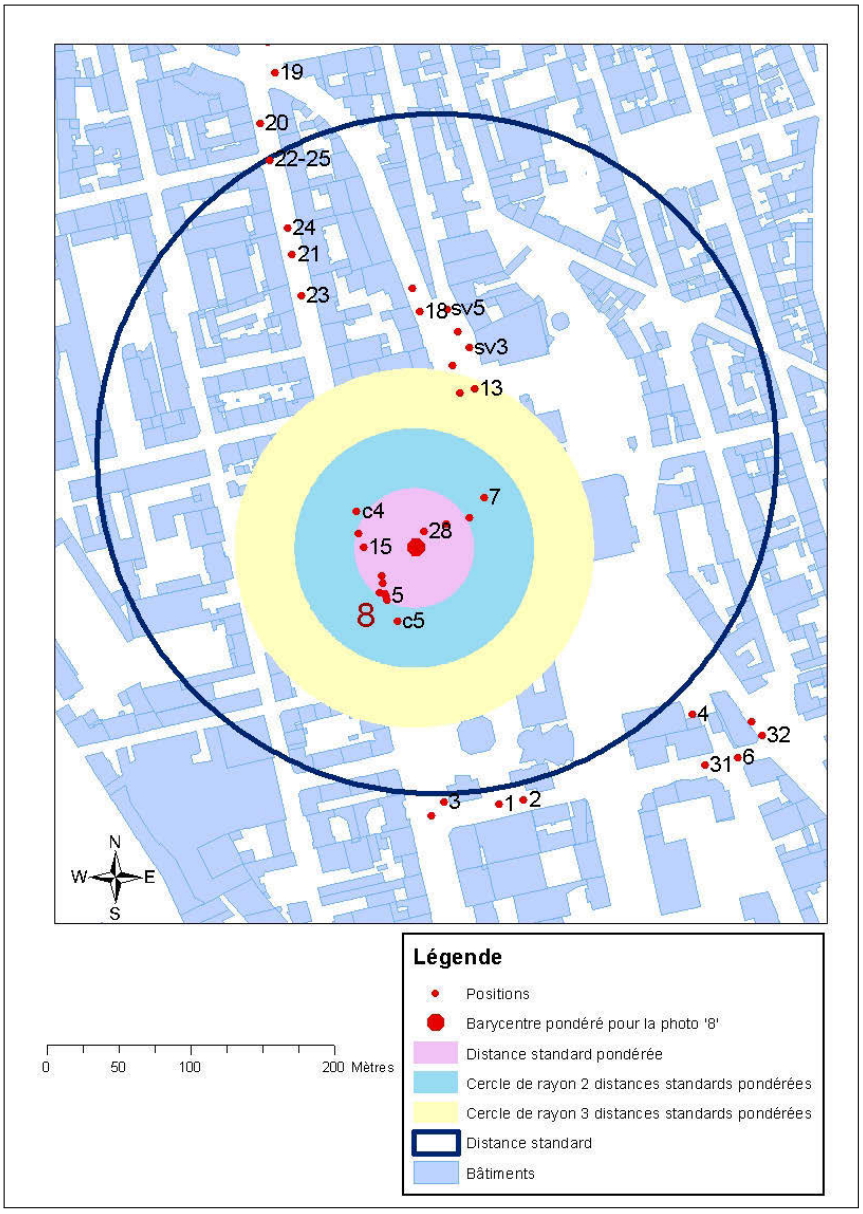


FIGURE 5.6 – Distance standard et distance standard pondérée avec les taux d'appariement de la photographie '8'.

Deuxième illustration du scénario 2

Nous considérons une autre photographie avec une connaissance plus floue de la localisation.

Nous supposons que nous n'avons que l'information indiquant que la photographie est prise dans la ville de Saint Étienne, nous prenons à cet effet une photographie et nous effectuons notre appariement avec la totalité de la collection d'images

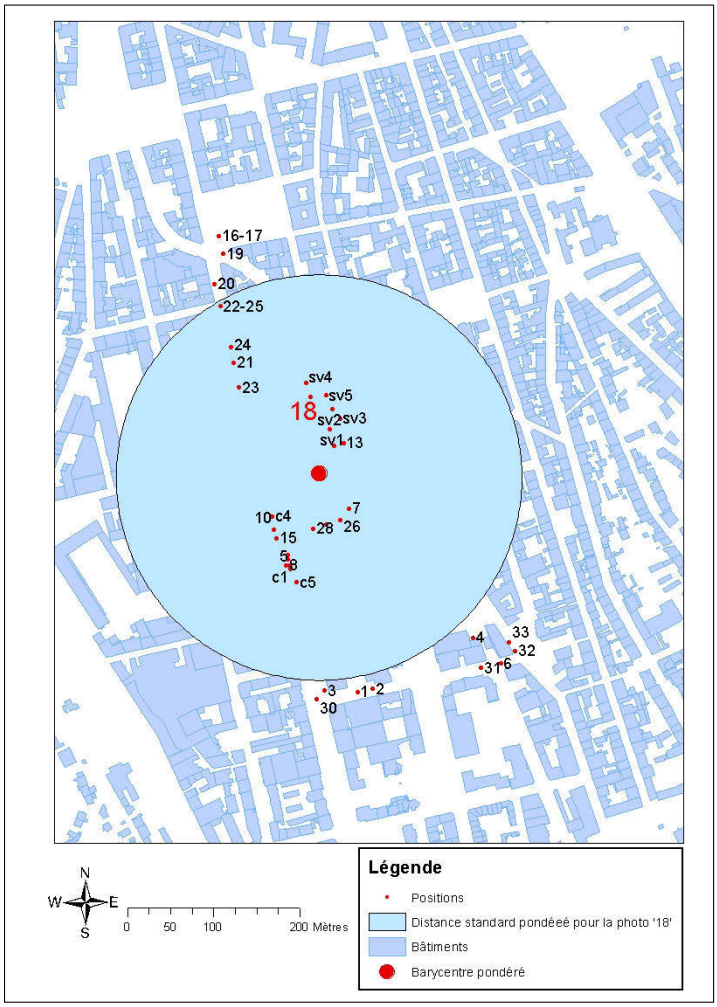


FIGURE 5.7 – Barycentre pondéré calculé pour la position de la photographe '18' et distance standard.

de synthèse qui est en notre possession (83 images réparties sur 39 positions).

Nous ne considérons que les valeurs supérieures à la moyenne des taux d'appariement. Nous avons 41 images de synthèse dans cette catégorie. Nous calculons le barycentre pondéré par les taux d'appariement des images de synthèse. Nous illustrons par la figure 5.7 la position suggérée pour la photographie. La distance standard pondérée présentée dans la même figure a un rayon de 235 mètres.

Nous constatons que 90 mètres séparent la position trouvée et la bonne position de la photographie. Nous avons alors plus de correspondance avec les images proches qu'avec celles qui sont très éloignées. Cette distance s'explique par la présence d'appariements avec des taux importants dans des zones lointaines par rapport à la position de la photographie.

À titre indicatif, on peut comparer cette localisation à 90 m avec le rectangle de

dimension 250 X 1200 mètres qui englobe les images utilisées par le calcul.

3ème scénario

Il peut arriver que des bâtiments aient disparu ou soient apparus sans que la base de données ait été mise à jour. La photographie est plus récente que la base de données. Elle permet de signaler que celle-ci doit être modifiée.

Il est possible de corriger la base de données géographique avec les données à partir d'une photographie localisée précisément à cet effet.

Si la position de la photographie est exacte, le taux de matching avec l'image correspondante devrait être maximal. S'il est faible, la différence dans l'image peut être signe d'un changement dans la réalité et qu'un problème existe au niveau de la base de données géographique.

Nous cherchons alors dans l'image de synthèse la partie où les droites ne s'apparient pas pour signaler les bâtiments qui auraient pu changer.

Concrètement, si on retrouve un appariement faible entre une image de synthèse et une photographie correctement localisée, cela peut indiquer que la base de données ne prend pas en compte l'état urbain réel et qu'il est nécessaire de vérifier si une mise à jour des données géographiques est nécessaire.

Illustration du scénario

Nous considérons une photographie géolocalisée avec précision. En trouvant un appariement faible nous pouvons en constater qu'a priori il existe un bâtiment manquant. Cela est possible en précisant les lignes détectées dans l'image de synthèse qui n'ont pas d'équivalent dans la photographie. Nous affectons alors à chaque couple de lignes appariées le numéro de la ligne correspondante. La liste des lignes restantes dans la photographie est candidate à la création d'une mise à jour de la base de données géographique.

En considérant les lignes verticales nous pouvons suggérer l'existence d'une limite de bâtiment qui manque dans la base de donnée. Un calcul de positionnement prenant en compte la distance à l'édifice, la focale et les différentes inclinaisons est nécessaire pour proposer une modification de structure dans la base de données géographique.

Dans l'exemple que nous affichons à la figure 5.8 la ligne rouge ne trouve pas de ligne correspondante dans l'image de synthèse. Les deux bâtiments de la photographie sont réellement de couleur et texture différentes alors que la bases de données géographique les modélise en un seul bloc. C'est l'indication d'une représentation non fidèle de la base de données.

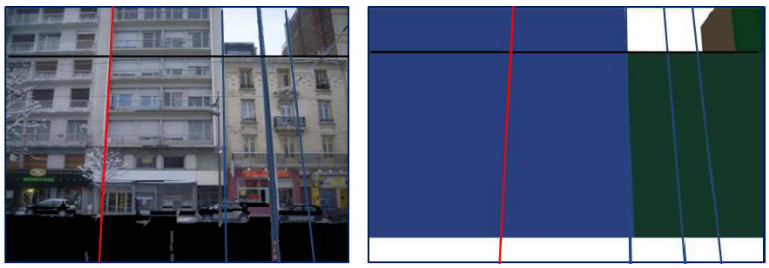


FIGURE 5.8 – Affichage de la ligne non trouvée dans l'image de synthèse.

Conclusion

L'étude de la dispersion des localisations par le biais de la distance standard permet d'avoir un cercle couvrant la position que nous cherchons et des zones de probabilité de présence de la photographie. Le barycentre pondéré que nous calculons représente le centre de la distance standard pondérée avec le taux d'appariement.

Les scénarios présentés permettent de décrire trois façons différentes d'utiliser notre approche d'appariement de lignes.

Les informations que nous avons sur une photographie permettent de choisir la démarche la plus adaptée pour compléter ces données par une localisation, caractérisation de contenu ou utilisation de la photographie pour la maintenance d'une base de données géographique. En fonction des cas, la méthode permet de localiser une photographie avec une indication de probabilité, de décrire son contenu ou de l'utiliser pour la maintenance d'une base de données géographique.

Conclusion

Conclusion

Bilan

Dans l'objectif de localiser et de documenter les photographies prises dans une ville, nous avons émis l'hypothèse qu'il était possible de s'appuyer sur la création d'un monde virtuel représentatif du paysage urbain en question. En comparant la photographie à une collection d'images de synthèse issues de ce monde virtuel il est alors possible d'estimer la localisation de cette photographie et de mettre en relation son contenu avec les informations stockées dans la base de données ayant servi à construire le monde virtuel. Le repérage des bâtiments présents dans la photographie permettrait ainsi d'étiqueter la photographie avec le nom de la rue et la présence d'un monument ou d'un équipement. Il permettrait aussi d'accéder à des informations connexes issues de la base de données géographique ou d'Internet et qui concernent le lieu photographié : information de service, heures d'ouverture d'un magasin, événement culturel, animation commerciale... .

Pour mener à bien ce projet nous avons combiné des techniques et méthodologies issues de plusieurs domaines. Le premier de ces domaines est celui de la géomatique. L'approche proposée a nécessité la mise en place d'une base de données géographiques dans ses dimensions spatiales et attributaires et l'emploi de différentes méthodes de visualisation 2D et 3D, d'interrogation et d'analyse statistique spatiale. La constitution de la base de données photographique a nécessité aussi la mise en œuvre de techniques de positionnement et de localisation à l'aide de satellites ou de réseaux. Même si l'incertitude liée à ces mesures peut être corrigée par un mixage de techniques, une position géographique donnée par un GPS est très imprécise dans un milieu urbain. Dans notre test nous constatons de 2 à 30 mètres d'imprécision horizontale en utilisant des GPS standards, qu'il est possible de réduire en s'appuyant sur des données issues d'une base de données géographiques.

Le deuxième domaine mobilisé est celui de l'imagerie. Un démonstrateur mis au point dans le projet MapYourPicture avait démontré la possibilité de localiser interactivement une photographie en la comparant avec une image de synthèse construite par la technique du Z-buffer à partir des données sur les bâtiments décrits dans la base de données géographiques. Le problème qui se posait à nous était de proposer une méthode pour localiser le plus automatiquement et précisément possible la pho-

tographie à partir d'une ou plusieurs images de synthèse, en retrouvant l'image de synthèse qui se rapprochait le plus fidèlement de la photographie. Notre proposition est d'apparier automatiquement la photographie et l'image en détectant les lignes communes à cette photographie et à cette image au moyen de la transformée de Hough. Les photographies présentent un nombre très élevé de lignes alors que les images de synthèse, compte tenu du niveau de détail faible de la modélisation 3D disponibles, ne contiennent que les limites des bâtiments. Nous avons cherché des indicateurs fiables et discriminants permettant de quantifier le niveau d'appariement et nous avons montré que l'appariement était possible et fonctionnait de manière suffisante pour répondre à nos objectifs de localisation.

Nous avons alors mis en œuvre une approche de fusion de données pour coupler les informations géographiques et photographiques. Nous avons constitué un jeu de données test et validé notre méthodologie par l'intermédiaire d'une fouille de données et d'une analyse statistique. Notre expérimentation valide la pertinence de l'approche. Il est possible par notre méthode d'appariement de lignes de trouver une image de synthèse correspondant à une photographie prise dans la rue, dès lors que la photographie est prise avec un axe à peu près horizontal. La précision de la localisation de la photographie dépend bien évidemment de la densité des images de synthèse dont nous disposons. Plus le maillage utilisé est fin, plus la photographie peut être localisée précisément. Nous avons proposé trois scénarios d'usage différents mettant en œuvre cette méthodologie et illustrant son utilité dans la pratique.

Limites de la méthode et perspectives d'amélioration

Il n'a pas été possible dans le temps de la thèse de constituer un nouveau jeu de données *ad hoc* permettant d'évaluer et de valider les scénarios proposés dans le chapitre 5. Il serait nécessaire pour valider complètement la proposition de constituer cette base d'images de synthèse selon une grille régulière et de tester précisément les scénarios proposés.

La méthode proposée se heurte par ailleurs à des difficultés liées au détail des données de la modélisation 3D. D'abord, la modélisation virtuelle ne prend pas en compte certains éléments importants du paysage urbain. Il faudrait y inclure les arbres, les signalisations, les réseaux ferrés pour obtenir un modèle 3D plus fidèle à la réalité.

Les limites des bâtiments sont modélisées en se fondant sur l'empreinte vue du ciel. Cette représentation diffère de quelques centimètres des dimensions au sol dans la majorité des cas, à cause du débord des toits. L'intégration de la modélisation des toits dans le modèle 3D urbains devrait améliorer les résultats de la méthode.

Les bases de données des textures des bâtiments existent pour certaines zones dans le monde, en particulier les grandes villes, mais elles ne sont pas encore généralisées. La tendance actuelle est de représenter les villes et les zones rurales en 3D

en plaquant les textures capturées par des capteurs photographiques. CityGML ou Google Earth par exemple proposent aujourd'hui leurs propres modèles 3D complets. La présence de textures dans les modèles 3D permettrait de retrouver beaucoup plus de lignes communes à la photographie et à l'image de synthèse.

Une deuxième piste d'amélioration concerne l'élimination des obstacles entre l'appareil photographique et les bâtiments afin de favoriser la ressemblance entre le monde réel et la modélisation 3D. Cette ressemblance permettrait un meilleur appariement entre les images de synthèse et les photographies correspondantes. Les techniques d'élimination des objets situés au-devant de la scène se multiplient dans le monde du traitement de l'image et deviennent de plus en plus performantes. Cela devrait donc permettre d'améliorer encore le taux de reconnaissance.

Un deuxième aspect optimisant notre approche est la possibilité de migrer les traitements pour les exécuter directement sur un Smartphone ou sur une tablette. La mobilité permettrait de faire le traitement juste après la prise de vue dès lors que les processeurs et les mémoires internes dont disposent ces téléphones intelligents convergent vers une capacité de traitement proche d'un ordinateur de bureau.

En effet, il n'est pas nécessaire de sauvegarder les images du monde virtuel sur l'appareil mobile, mais uniquement les deux paramètres des lignes de chacune des images ainsi que les paramètres de localisation. La connexion au nuage (Cloud) permet actuellement un minimum de transfert de données avec un accès illimité à ces derniers.

Cette configuration permettrait d'effectuer les traitements sur un serveur connecté et d'utiliser l'application à partir d'un Smartphone, d'une tablette, d'un ordinateur de bureau, d'un portable ou de tout autre type de terminal ayant un accès au réseau Internet. Dans ce cas un encodage de l'algorithme dans un des langages adaptés au Web tel que C++ ou Java est nécessaire.

Un troisième axe d'amélioration porte sur la correction des bases de données par la constatation de la réalité du terrain. Une suggestion de squelette du ou des bâtiments manquants permettrait de renseigner la base de données géographique d'une façon efficace sur l'absence ou la modification d'un bâtiment.

Annexes

Annexe A

Base de données photographiques

Nous présentons dans cette annexe les photographies de la base de données de test prises par les différent appareils photographiques détaillés dans l'annexe B.

Les photographies sont présentées en couleurs puis sont suivies par les photographies après élimination du sol dans les cas où le sol est détecté. Les zones qui représentent le sol ou le ciel sont colorées en noir dans ces dernières photographies.



(a) Photo 1



(b) Photo 25



(c) Photo 4



(d) Photo 4 sans sol



(e) Photo 5



(f) Photo 5 sans sol



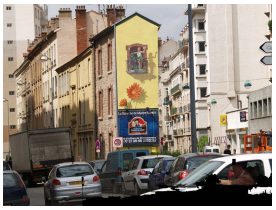
(g) Photo 6



(h) Photo 7



(i) Photo 9



(j) Photo 9 sans sol

FIGURE A.1 – Base de données photographiques (partie 1).



(a) Photo 12 (b) Photo 12 sans sol



(c) Photo 14 (d) Photo 14 sans sol



(e) Photo 15 (f) Photo 15 sans sol



(g) Photo 16 (h) Photo 19



(i) Photo 20 (j) Photo 23 (k) Photo 23 sans sol



(l) Photo 24 (m) Photo 26 (n) Photo 26 sans sol



(o) Photo 28 (p) Photo 28sans sol (q) Photo 29 (r) Photo 30



(s) Photo 33 (t) Photo 33 sans sol

FIGURE A.2 – Base de données photographiques (partie 2).

Annexe B

Appareils utilisés pour la prise des photographies

Les photographies que nous utilisons pour la validation du concept ont été prises en utilisant cinq appareils différents.

1. **Le RICOH Caplio 500SE**

   La spécificité de ce type d'appareils photographiques par rapport aux appareils numériques classiques est qu'ils fournissent des données collectées par des capteurs intégrés. Ces capteurs sont une boussole électronique affichant la direction et une puce GPS affichant la position.

   La position géographique est fournie dans le cas d'une réception GPS valide. Ces données et d'autres sont écrites directement dans l'extension EXIF des images compressées.

   Notons que l'EXIF de ces photographies stocke les caractéristiques du capteur de la caméra et les paramètres de l'image prise. Il offre aussi la possibilité d'associer des notes et donne des détails sur la réception GPS au moment de la prise de vue.

   La figure B.1 illustre ces données relatives à la photographie illustrée par la figure B.2. Ces données concernent la longitude, la latitude, l'altitude par rapport au niveau de la mer, les identifiants des satellites GPS connectés au moment de la capture, la valeur du DOP (voir chapitre 2), la vitesse et la direction de visée.

   Ces photographies prises à 8 Mégapixels incluent dans un coin de l'image capturée l'heure et la date de la prise de vue ainsi que la longitude et la latitude. La longitude est précédée par la lettre 'N' et la longitude est précédée par la mention 'E' qui désignent la direction du référencement. 'E 4°' désigne une position située à 4 degrés vers l'Est.

   Ces données sont affichées en utilisant la couleur orangée sur le coin bas droit

GPS	
Latitude Ref	North
Latitude	45° 25' 49.51"
Longitude Ref	East
Longitude	4° 23' 23.48"
Haltitude Ref	Other
Haltitude	560.0000
TimeStamp	17
Satellites	29,31,21,25,,,,,,,,
Status	Measurement In Progress
Measure Mode	Measurement In Progress
DOP	19/2
Speed Ref	N
Speed	0
Track Ref	True North
Track	129
Img Direction Ref	M
Img Direction	240
Map Datum	WGS-84
Dest Bearing Ref	True North
Dest Bearing	Other
Dest Distance Ref	K
Dest Distance	-1.#IND

FIGURE B.1 – Données de la rubrique GPS des métadonnées d'une photo prise par le CAPLIO 500SE.

de l'image comme l'illustre la figure B.2.

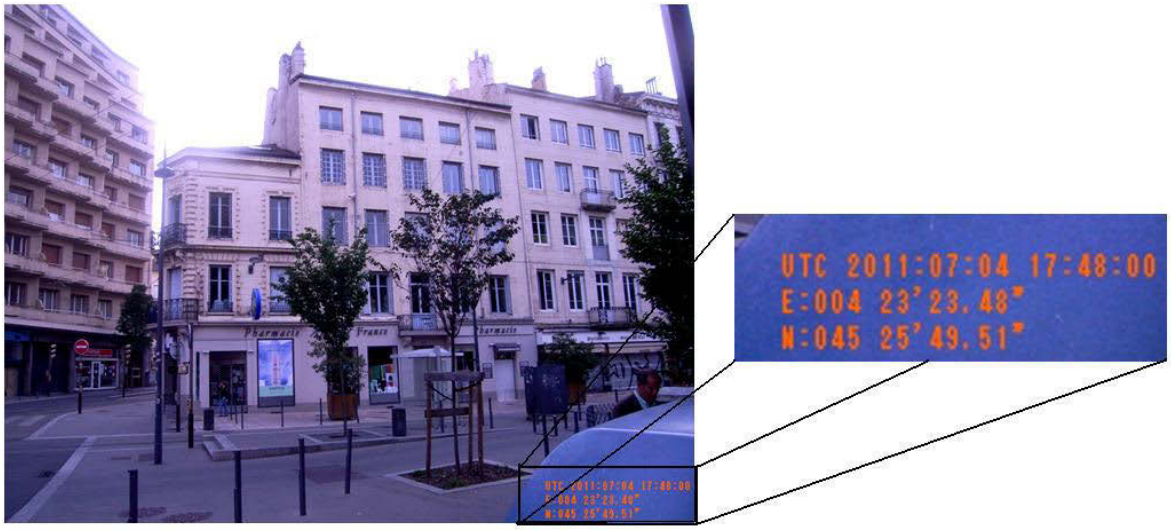


FIGURE B.2 – Affichage des coordonnées sur la photographie prise par le CAPLIO 500SE.

2. **Le Panasonic Lumix DMC-TZ10**

   Il est équipé d'un capteur de 12.1 Mégapixels. Il est équipé d'une puce GPS intégrée. Les données GPS sont placées dans l'extension EXIF de l'image comme illustré par la figure B.3. Nous distinguons dans cette figure la latitude, la longitude, le nombre de satellites disponibles et le DOP.

GPS	
Version ID	2.2.0.0
Latitude Ref	North
Latitude	45° 25' 34.84"
Longitude Ref	East
Longitude	4° 23' 34.40"
TimeStamp	14
Satellites	3
Status	Measurement In Progress
Measure Mode	Measurement In Progress
DOP	3
Map Datum	WGS-84
Processing Method	GPS
Date Stamp	2011:11:28

FIGURE B.3 – Données GPS dans l'EXIF d'une image prise par le Lumix DMC-TZ10. (Source : imprimé d'écran des métadonnées affichées par le logiciel XnView)

3. **Le Smartphone HTC HERO**

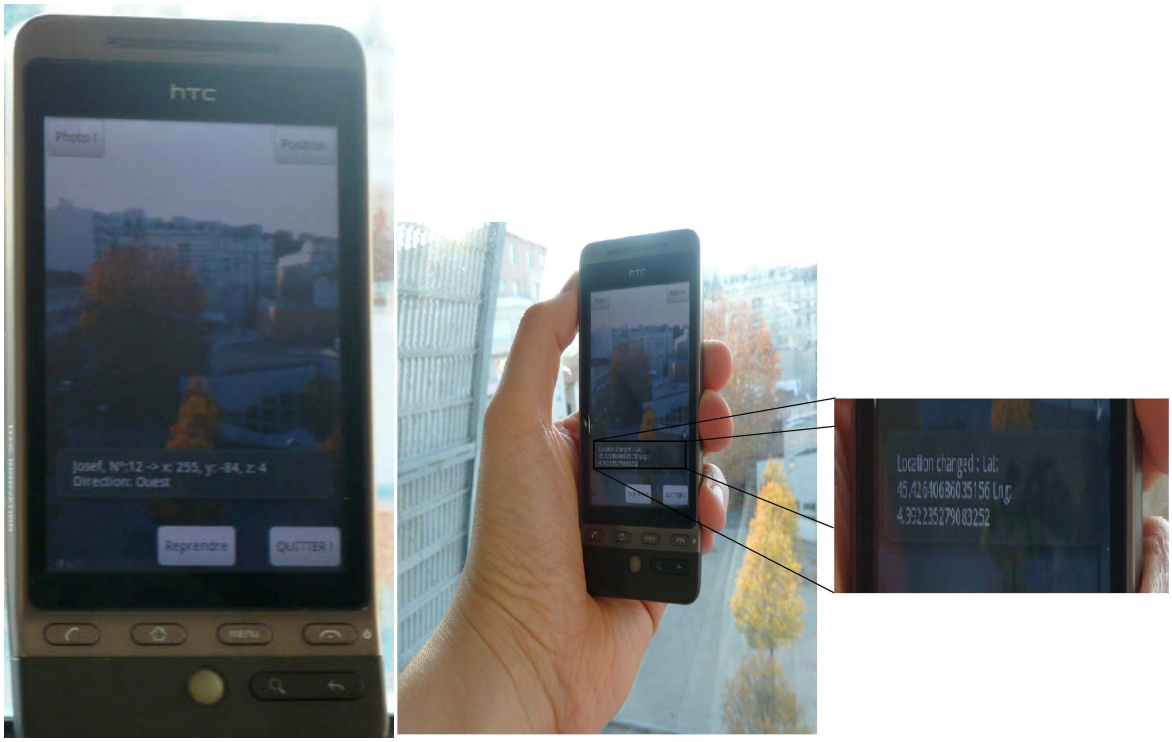
   Il comporte une centrale inertielle et une puce GPS ainsi qu'un appareil photo numérique avec autofocus et 5 Mégapixels (figure B.4). Nous avons ajouté les coordonnées de positionnement autour des trois axes de rotation dans l'espace. Ce sont les axes X, Y et Z dans la figure B.4a.

4. **Le Smartphone Sony Ericsson Xperia X10 mini pro**

   Il partage les mêmes caractéristiques techniques que le HTC Hero (capteur à 5 million de pixels et centrale inertielle), cependant il est plus récent. Le HTC Hero date de la première vague des Smartphones. Il a fallu attendre plusieurs mois pour avoir une mise à jour logicielle a fin de profiter des données proposées par le GPS et la boussole dans l'EXIF de l'image. L'Xperia X10 permet de faire fonctionner la même application programmée pour le HTC Hero. Nous illustrons un imprimé écran de cette application par la figure B.5.

5. **Le reflex Olympus E-500**

   C'est un réflex numérique classique permettant de prendre des photographies à 8 million de pixels. Il a été couplé avec un GPS pour récupérer la position



(a) Les degrés de rotation (b) Latitude et longitude

FIGURE B.4 – Application d'ajout des données de positionnement sur les photographies prises par le HTC Hero.



FIGURE B.5 – Imprimé écran de la prise N°2 avec les degrés de rotation sur les axes X, Y et Z. (Source : capture d'écran de l'XPERIA X10)

géographique et avec une boussole analogique pour préciser la direction de visée pendant la prise de la photographie.

Annexe C

Description des données utilisées pour la génération des images de synthèse avec AcrScene

Les quatre classes sélectionnées parmi les 11 classes de la BD TOPO [IGN 11b] sont celles qui possèdent une hauteur dans cette base de données et font plus de 20 m^2 au sol :

1. BATI_INDIFFERENCIE : ce sont les bâtiments d'habitation, les hôpitaux, les musées, les prisons...
   Les bâtiments indifférenciés représentent 90 % de la totalité des bâtiments.
2. BATI_REMARQUABLE : ce sont les aérogares, les châteaux, les bâtiments sportifs, les bâtiments religieux, les bâtiments de l'état (mairie, préfecture...), les gares... Les bâtiments dits remarquables forment environ 2 % de l'ensemble des bâtiments présents dans la classe BATI de la version 1.2 de la BD TOPO.
3. BATI_INDUSTRIEL : ce sont les bâtiments à caractère industriel, agricole ou commercial et qui représentent environ 10 % de la totalité des bâtiments.
4. RESERVOIR : ce sont les réservoir d'eau ou de liquides industriels de plus de 10 m de diamètre.

Ces quatre classes ont sept attributs en commun :

1. ID : l'identifiant unique pour chaque bâtiment.
2. PREC_PLANI : c'est la précision planimétrique. Elle est donnée en mètre et varie par rapport à chaque source de données (levé GPS, carte...).
3. PREC_ALTI : la précision géométrique altimétrique.
4. ORIGIN_BAT : il définit avec précision d'où est issu le bâtiment. Les cinq origines possibles sont : BD TOPO, cadastre, terrain, autre ou NR (Non Renseigné).

5. HAUTEUR : c'est la hauteur du bâtiment qui correspond à la différence d'altitude entre le point le plus élevé du pourtour du bâtiment et un point situé au pied du bâtiment. Sa valeur est arrondie au mètre.

   C'est la valeur que nous utilisons pour donner une hauteur à nos bâtiments dans la modélisation 3D de la ville.

6. Z_MIN : c'est l'altitude minimale du pourtour d'un bâtiment. Cette valeur n'as pas subi de mise à jour par rapport aux versions précédentes.

   C'est la valeur minimale de l'altitude du pourtour du bâtiment qui coïncide généralement avec l'altitude des gouttières.

7. Z_MAX : altitude maximale du pourtour d'un bâtiment.

Nous avons illustré les trois derniers attributs dans la figure C.1.

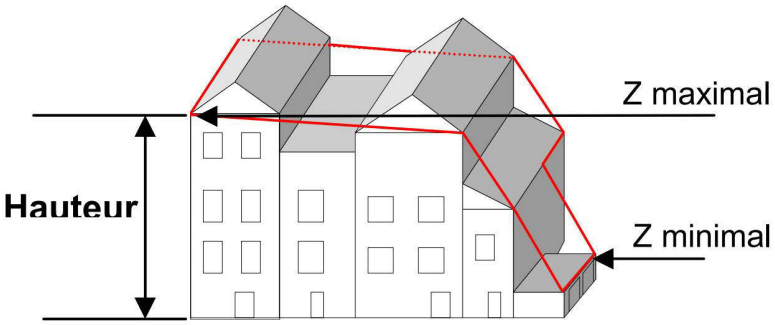


FIGURE C.1 – Description de la hauteur et des altitudes maximale et minimale dans la BD TOPO. (Source : descriptif de contenu de la version 2 de la BD TOPO)

Un attribut "NATURE" qui précise le type du bâtiment s'ajoute aux classes : BATI_REMARQUABLE, BATI_INDUSTRIEL et RESERVOIR.

Annexe D

Tableau des faux positifs

Nous présentons dans un tableau notre base de données de test constituée des 20 photographies prises dans 18 positions différentes.

Ces photographies ont été comparées à 73 images de synthèse. Ces dernières sont réparties sur 29 positions différentes.

Les images de synthèse de la famille des images 9, 11 et 12 partagent la même position géographique. Lors de la prise de vue, c'est la focale ou l'orientation qui change dans les photographies et par conséquence dans les images de synthèse.

C'est le même cas pour le couple d'images 16 et 17 et le couple d'images 22 et 25.

Le tableau D.1 présente la distance calculée point à point entre les coordonnées géographiques de la photographie (colonne de gauche) et l'image de synthèse qui s'y apparente. Nous affichons dans ce tableau les appariements entre chaque photographie et ces faux positifs en rouge.

À fin de faciliter la lecture du tableau, nous donnons deux exemples. Le premier est celui de la photographie 25 qui s'apparie bien avec elle même mais en plus avec l'image de synthèse '10' qui est éloigné de 267 mètres.

La photographie '25' s'apparie avec l'image de synthèse '12' et '19'. La distance qui sépare la position de la prise de photographie '25' et '12' est de 209 mètres.

La valeur moyenne de toutes les distances figurant dans le tableau D.1 est de 322 mètres avec un écart type de 312 mètres. La valeur médiane est de 249 mètres. La valeur minimale est 0 et la maximale est de 1226 mètres séparant les positions des photographies '14' et '6'.

	1	2	3	4	5	6	7	8	9,11 et 12	10	14	15	16 et 17	18	19	20	21	22 et 25	23	24	26	27	28	29	30	31	32	33
S1	0	18	38	149	164	169	213	169	287	212	1200	202	553	347	532	501	408	475	379	426	200	198	197	178	48	146	189	185
S2	18	0	56	131	170	152	211	175	286	218	1202	207	555	347	534	504	411	478	383	430	199	199	199	184	65	128	171	167
S3	38	56	0	183	148	207	213	152	284	196	1189	185	542	341	520	489	395	462	365	413	198	193	189	163	13	184	226	221
S4	149	131	183	0	228	44	208	233	275	264	1187	256	552	338	532	508	423	484	398	439	206	216	225	236	195	36	51	41
S5	164	170	148	228	0	269	96	5	150	48	1040	37	394	199	372	340	246	314	217	265	79	65	52	15	156	251	278	268
S6	169	152	207	44	269	0	252	274	318	306	1227	298	595	381	575	552	467	527	442	483	250	259	269	278	217	23	23	27
S7	213	211	213	208	96	252	0	98	75	91	994	91	351	137	330	303	215	278	190	232	17	32	48	89	224	241	254	242
S8	169	175	152	233	5	274	98	0	149	44	1036	33	390	197	369	336	242	310	213	261	81	67	53	12	159	256	284	274
S9	287	286	284	275	150	318	75	149	0	120	920	127	278	63	257	233	151	209	129	166	87	91	99	138	295	309	317	305
S11 et S12	287	286	284	275	150	318	75	149	0	120	920	127	278	63	257	233	151	209	129	166	87	91	99	138	295	309	317	305
S10	212	218	196	264	48	306	91	44	120	0	993	10	347	160	326	293	199	267	170	218	78	61	45	33	203	289	313	302
S13	289	287	288	272	157	315	76	156	11	129	921	135	281	66	260	237	157	214	137	171	90	96	105	145	298	306	313	300
S14	1200	1202	1189	1187	1040	1227	994	1036	920	993	0	1003	647	857	668	700	794	726	823	775	1004	1003	1004	1025	1196	1222	1220	1208
S15	202	207	185	256	37	298	91	33	127	10	1003	0	357	169	336	303	209	277	180	228	77	60	44	23	192	281	306	295
S16 et S17	553	555	542	552	394	595	351	390	278	347	647	357	0	215	21	56	148	82	178	130	359	357	357	379	549	587	592	579
S18	347	347	341	338	199	381	137	197	63	160	857	169	215	0	194	171	97	148	83	108	148	149	153	185	351	372	379	366
S19	532	534	520	532	372	575	330	369	257	326	668	336	21	194	0	37	127	61	156	109	338	336	335	358	528	566	572	559
S20	501	504	489	508	340	552	303	336	233	293	700	303	56	171	37	0	94	26	123	75	310	307	306	325	496	542	550	538
S21	408	411	395	423	246	467	215	242	151	199	794	209	148	97	127	94	0	68	29	19	221	216	213	232	402	456	467	455
S22 et S25	475	478	462	484	314	527	278	310	209	267	726	277	82	148	61	26	68	0	97	49	285	281	280	299	469	518	526	544
S23	379	383	365	398	217	442	190	213	129	170	823	180	178	83	156	123	29	97	0	48	194	188	185	202	373	430	443	430
S24	426	430	413	439	265	483	232	261	166	218	775	228	130	108	109	75	19	49	48	0	238	233	231	250	420	472	483	470
S26	200	199	198	206	79	250	17	81	87	78	1004	77	359	148	338	310	221	285	194	238	0	17	33	73	209	237	253	241
S27	198	199	193	216	65	259	32	67	91	61	1003	60	357	149	336	307	216	281	188	233	17	0	16	57	203	246	264	252
S28	197	199	189	225	52	269	48	53	99	45	1004	44	357	153	335	306	213	280	185	231	33	16	0	43	198	254	274	263
S29	178	184	163	236	15	278	89	12	138	33	1025	23	379	185	358	325	232	299	202	250	73	57	43	0	171	260	287	276
S30	48	65	13	195	156	217	224	159	295	203	1196	192	549	351	528	496	402	469	373	420	209	203	198	171	0	194	237	232
S31	146	128	184	36	251	23	241	256	309	289	1222	281	587	372	566	542	456	518	430	472	237	246	254	260	194	0	44	44
S32	189	171	226	51	278	23	254	284	317	313	1220	306	592	379	572	550	467	526	443	483	253	264	274	287	237	44	0	12
S33	185	167	221	41	268	27	242	274	305	302	1208	295	579	366	559	538	455	514	430	470	241	252	263	276	232	44	12	0

TABLE D.1 – Tableau des distances en mètres entre les photographies et les images de synthèse.

Glossaire et acronymes

AGPS : Assisted GPS
AHN : Actueel Hoogtebestand Nederland
CC50 : Projection conique conforme du nord de la France
CERTU : Centre d'Études sur les Réseaux, les Transports, l'Urbanisme et les constructions publiques
CNES : Centre National d'Etudes Spatiales
CRENAM : Centre de Recherche sur l'ENvironnement et l'Aménagement
CRISP-DM :CRoss-Industry Standard Process for Data Mining
DGI : Direction générale des Impôts
DGPS : Differential Global Positioning System
DICOM : Digital Imaging and Communications in Medicine
DIPI : Diagnostic et Imagerie des Procédés Industriels
EDGE : Enhanced Data Rates for Global Evolution
Egnos : Européan Geostationary Overlay Service
ESA : European Space Agency
Esri : Environmental Systems Research Institute
EXIF : Exchangeable image file format
FICCDC : Federal Interagency Coordinating Committee on Digital Cartography
GAGAN : GPS and Geo Augmented Navigation
GDOP : Geometry Dilution Of Precision
GPRS : General Packet Radio Service
GPS : Global Positioning System
GRASS : Geographic Resources Analysis Support System
GSM : Global System for Mobile Communications
HDOP : Horizontal Dilution Of Precision
IERP : Institut des Études Régionales et des Patrimoines
IGN : Institut national de l'information géographique et forestière
IRNSS : Indian Regional Navigarional Satellite System

ISIG : Imagerie et Systèmes d'Information Géographique

ISTHME : Image Société Territoire Homme Mémoire Environnement

JDL : U.S. Joint Directors of Laboratories Data Fusion Group

JPEG est l'acronyme de Joint Photographic Experts Group

LoD : Levels-of-Detail

MNT : Modèle Numérique de Terrain

MSAS : Multi Satellite-based Augmentation System

NAVSTAR : Navigation Satellite Timing And Ranging

Pixel : Picture Element

PDOP : Position dilution Of Precision

POI : Point Of Interest

QZSS : Quasi-Zenith Satellite System

RANSAC : RANdom SAmple Consensus

RFID : Radio Frequency Identification

RGF93 : Réseau Géodésique Français calculé pour l'époque 1993

RGP : Réseau GPS Permanent

SBAS : Satellite-Based Augmentation System

SHOM : Service hydrographique et océanographique de la marine

SIFT : Scale-invariant feature transform

Spot : Système probatoire d'observation de la Terre ou Satellite pour l'observation de la Terre

TDOP : Time Dilution Of Precision

TIN : Triangulated Irregular Network

UMTS : Universal mobile telecommunications system

VDOP : Vertical Dilution Of Precision

WAAS : Wide Area Augmentation System

WiFi : Wireless Fidelity

WiMAX : Worldwide Interoperability for Microwave Access

WPS : Wifi Positionning System

Table des matières

Remerciements 3
RÉSUMÉ 5
ABSTRACT 6

Introduction **11**
Cadre de la thèse 11
Logique de l'exposé 12

1 Géomatique, infographie et fusion de données **17**
1.1 Notions de géomatique 18
1.1.1 Bases de données géographiques 19
1.1.1.1 Géoréférencement 21
1.1.1.2 Outils SIG 21
1.1.2 Les technologies de positionnement 23
1.1.2.1 Positionnement Outdoor (en extérieur) 23
1.1.2.2 Positionnement Indoor (en intérieur) 27
1.1.3 Techniques de mesure d'orientation 29
1.1.3.1 Boussole 29
1.1.3.2 Centrale à inertie 29
1.1.3.3 Systèmes hybrides 30
1.2 Notions en traitement d'image 31
1.2.1 L'image numérique 32
1.2.2 Opérations sur les images 32
1.2.3 Les filtres 33
1.2.4 Segmentation 36
1.2.5 Détection de contours 37
1.3 Processus de la fusion de données 39
1.3.1 Appariement de données géographiques 40

1.3.2 Appariement de graphes 41
1.3.3 Appariement de lignes 42
1.4 Conclusion 43

2 Couplage de BD photo et géographiques 47
2.1 Méthodes de couplage dans la littérature 49
2.1.1 Méthodes orientées utilisateur 50
2.1.2 Méthodes automatiques 50
2.1.3 Calculs de visibilité et photographie 53
2.1.3.1 Lancer de rayons 53
2.1.3.2 Notion d'Isovist 54
2.2 Méthodes utilisées dans le projet MapYourPicture 55
2.2.1 Approche de départ 55
2.2.2 Description du démonstrateur 58
Description des données utilisées 59
2.2.3 Résultats et évaluation 59
2.3 Conclusion 62

3 Localisation de photographies par appariement de lignes 67
3.1 Présentation globale 67
3.2 La démarche d'interfaçage par appariement de lignes 69
3.2.1 Présentation des données 69
La transformation de Hough 70
3.2.2 Déroulement de la démarche 75
3.2.3 Algorithme 76
3.3 Choix technologiques et méthodologies 77
3.3.1 Création de rendu 78
3.3.1.1 MapYourPicture 79
3.3.1.2 ArcScene 82
3.3.2 Matlab 85
3.4 Traitement d'image sur les photographies 86
3.4.1 Élimination du sol 86
3.4.1.1 Importance de l'élimination du sol 86
3.4.1.2 Méthode utilisée 87
3.4.2 Extraction des droites de Hough 88
3.5 Conclusion 89

4 Validation du concept **93**
4.1 Méthode d'analyse des données . . . 93
4.2 Données utilisées . . . 95
4.2.1 Prises de photographies . . . 95
4.2.2 Base de données géographiques . . . 99
4.2.2.1 Démonstrateur de MapYourPicture . . . 99
4.2.2.2 ArcScene . . . 100
4.2.3 Création des images de synthèse . . . 102
4.2.3.1 Images de synthèse . . . 102
4.2.3.2 Intervalles de confiance . . . 103
4.2.3.3 Décalages . . . 104
4.3 Résultats . . . 104
4.3.1 Durées de traitement . . . 104
4.3.2 Taux d'auto-reconnaissance . . . 105
4.3.3 Arbre de décision . . . 106
4.3.4 Les faux positifs . . . 112
4.4 Limites de la méthodologie . . . 115
4.5 Conclusion . . . 118

5 Scénarios d'utilisation fondés sur la distance **121**
1er scénario . . . 121
2ème scénario . . . 124
3ème scénario . . . 131
Conclusion . . . 132

Conclusion **135**
Bilan . . . 135
Limites de la méthode et perspectives d'amélioration . . . 136

Annexes **141**

A Base de données photographiques **141**

B Appareils utilisés pour la prise des photographies **143**

C Description des données utilisées sous AcrScene **147**

D Tableau des faux positifs **149**

Glossaire et acronymes **151**

Table des matières **153**

Table des figures **157**

Liste des tableaux **161**

Bibliographie **165**

Table des figures

1.1 Projection conique Lambert.(Source : Sigea(Systèmes d'Information Géographique pour l'Enseignement Agricole)) 22
1.2 Système GPS. (Sources : centre national d'études spatiales (CNES) et [EDUCNET 11]) 24
1.3 Perturbation du signal par multi-trajets. (Source : production personnelle inspirée de [Ray 09]) 25
1.4 L'appareil Ricoh Caplio S500SE équipé de puce GPS et d'une boussole numérique. (Source : documentation de l'appareil) 31
1.5 Mise en place d'un filtre sur une image. 34
1.6 Filtre de Laplace. 34
1.7 Utilisation du filtre de Laplace pour détecter les contours. (Source : photo personnelle | rendu par Matlab) 35
1.8 Segmentation d'une image par la couleur. 37
1.9 Vision des contours par l'œil. (Source :[Kaiser 09, Petitesexperiences 11]) 38
1.10 Détection de lignes avec le filtre de Robinson et la transformation Laplace-Gauss. 39
1.11 Fonctions de la fusion de données. (Source : Y. Attia à partir de [Grandin 06] et [Liggins 08]) 40

2.1 La tour Montparnasse vue de la Tour Eiffel. (Sources : Google Earth et 123rf.com) 49
2.2 Photos géolocalisées manuellement sur le site loc.alize.us et locr. . . . 50
2.3 Isovists dans le centre de Lyon. (Source : [Joliveau 07]) 55
2.4 Indexation des bâtiments et des rues. 57
2.5 Fenêtre principale du démonstrateur. 58
2.6 Les hauteurs minimales et maximales relatives à un seul bâtiment selon les cartes BD TOPO de l'IGN. (Source : [IGN 08b]) 59
2.7 Photographie et rendu des photos. 60
2.8 Résultat d'extraction du contenu d'une photo. 61

2.9 Bâtiments visibles à partir du point de vue. 61

3.1 Méthode globale d'interfaçage par appariement de lignes. 69
3.2 Représentation d'un point. 70
3.3 Représentation des lignes. 71
3.4 Détection de lignes par la la transformation de Hough. 73
3.5 Lignes imaginaires détectés par la transformée de Hough. 74
3.6 Les cinq niveaux de détails définis par CityGML. (Source : [Kolbe 05]) 79
3.7 Importance du relief pour une visibilité fidèle à la réalité. 80
3.8 Effet de marches d'escalier sur la détection de ligne dans les images de synthèse. (Source : capture d'écran du monde virtuel crée par le démonstrateur de MapYourPicture) 81
3.9 Affichage des zones inconnues dans l'image de synthèse. (Source : capture d'écran du démonstrateur de MapYourPicture) 82
3.10 Paramètres de visualisation dans ArcScene. (Source : capture d'écran d'ArcScene. ESRI) 83
3.11 Illustration des limites de la base de données utilisée. (Source : capture d'écran d'ArcScene) 84
3.12 Résultats de segmentation d'une image urbaine. (Source : image personnelle | rendu généré avec Matlab) 87
3.13 Élimination du sol sur une photographie. (Sources : image personnelle | rendu généré avec Matlab) 88

4.1 Jeu de données photographiques d'apprentissage. 96
4.2 Position des 20 prises de vue. 97
4.3 Les 29 positions des images de synthèse. 98
4.4 Capture d'écran de la prise N°2 avec les valeurs de rotation sur les axes X, Y et Z en degrés. (Source : capture d'écran de l'XPERIA X10) 99
4.5 Description de la hauteur et des altitudes maximale et minimale dans la BD TOPO. (Source : descriptif de contenu de la version 2 de la BD TOPO) 101
4.6 Capture d'écran du MNT en couche vecteur. 102
4.7 Exemple d'images de synthèse prises aux alentours d'une photographie. 103
4.8 Taux d'auto-reconnaissance (% pour chaque photographie). 105
4.9 Taux d'appariement de la photographie 7 avec les 73 images de synthèse. 107
4.10 Les quatre images de synthèse les mieux appariées avec la photographie '7'. 108
4.11 Positions des photographies '7' et '14' sur la carte de Saint-Étienne. . 109

4.12 Taux d'appariement de la photographie '1'. 110
4.13 Exemple d'image de synthèse prise aux alentours d'une photographie. 111
4.14 Arbre de décision. 112
4.15 Couverture des images de synthèse de la zone des prises de photographies. 115
4.16 Groupement de bâtiments de la BD TOPO. (Source : capture d'écran d'ArcScene. ESRI) 117

5.1 Recherche autour d'une position donnée et extension de zone de recherche. 122
5.2 Disque autour de la photographie '3' de rayon 18.7 mètres. 123
5.3 Démarche globale du second scénario. 125
5.4 Images dans une zone de 100 mètres autour de la photographie '8'. . 126
5.5 Barycentre pondéré calculé pour la position de la photographe '8' et distance standard pondérée. 128
5.6 Distance standard et distance standard pondérée avec les taux d'appariement de la photographie '8'. 129
5.7 Barycentre pondéré calculé pour la position de la photographe '18' et distance standard. 130
5.8 Affichage de la ligne non trouvée dans l'image de synthèse. 132

A.1 Base de données photographiques (partie 1). 141
A.2 Base de données photographiques (partie 2). 142

B.1 Données de la rubrique GPS des métadonnées d'une photo prise par le CAPLIO 500SE. 144
B.2 Affichage des coordonnées sur la photographie prise par le CAPLIO 500SE. 144
B.3 Données GPS dans l'EXIF d'une image prise par le Lumix DMC-TZ10. (Source : imprimé d'écran des métadonnées affichées par le logiciel XnView) 145
B.4 Application d'ajout des données de positionnement sur les photographies prises par le HTC Hero. 146
B.5 Imprimé écran de la prise N°2 avec les degrés de rotation sur les axes X, Y et Z. (Source : capture d'écran de l'XPERIA X10) 146

C.1 Description de la hauteur et des altitudes maximale et minimale dans la BD TOPO. (Source : descriptif de contenu de la version 2 de la BD TOPO) 148

Liste des tableaux

4.1 Résultat d'appariement de la photographie '1'. 110
4.2 Nombre des faux positifs. 113
4.3 Distances en mètre entre les positions des photographies et les positions calculées par notre méthode. 116

5.1 Résultat d'appariement de la photographie '3' avec les images de synthèse localisées dans un rayon de 18.7 mètres. 124
5.2 Résultat d'appariement de la photographie '8' dans un rayon de 100 mètres. 127

D.1 Tableau des distances en mètres entre les photographies et les images de synthèse. 150

Bibliographie

Bibliographie

[Aggarwal 06] Nitin Aggarwal & William Clem Karl. *Line Detection in Images Through Regularized Hough Transform.* IEEE Transactions on Image Processing, vol. 15, no. 3, pages 582–591, March 2006.

[Ambauen 03] R Ambauen, S Fischer & Horst Bunke. *Graph edit distance with node splitting and merging, and its application to diatom identification.* Graph Based Representations in Pattern Recognition, vol. 2726, pages 95–106, 2003.

[Amiot 12] Maxime Amiot. *La géolocalisation veut entrer dans les centres commerciaux.* http ://bit.ly/wlhzj1, 02 2012.

[Ampère 32] Jean-Jacques Ampère. La chine et les travaux d'Abel rémusat. 1832.

[Antani 11] Lakulish Antani, Anish Chandak, Micah Taylor & Dinesh Manocha. *Efficient finite-edge diffraction using conservative from-region visibility.* Applied Acoustics, 2011.

[Arnoud 07] De Boer Arnoud. *Label placement in 3D georeferenced and oriented digital photographs using GIS technology.* PhD thesis, TU Delft, OTB, Geo-Database Management Center, December 2007.

[Aron 06] Michael Aron, Gilles Simon & Marie-Odile Berger. *Utilisation d'un capteur inertiel comme aide au suivi basé vision.* In 15ème congrès francophone Reconnaissance des Formes et Intelligence Artificielle - RFIA 2006, Tours, France, January 2006.

[Attia 09a] Yousssef Attia, Thierry Joliveau, Eric Favier & Benoit Delaup. *Extraction de contenu photographique en utilisant les SIG.* MajecSTIC 2009, November 2009.

[Attia 09b] Yousssef Attia, Thierry Joliveau, Eric Favier, Benoit Delaup & Sanjay Rana. *SIG et extraction de contenu photographique. Une proposition méthodologique.* SAGEO 2009, November 2009.

[Aufaure 00] Marie-Aude Aufaure, Karine Zeitouni & Laurent Yeh. *Fouille de données spatiales.* In CEPADUES, editeur, Le temps, l'espace et l'évolutif en Sciences du Traitement de l'Information, page 319 à 328. 2000.

[Baillie 03] Jean-Christophe Baillie. *Segmentation.* Rapport technique ES322, ENSTA, Paris, 2003.

[Batty 01] M. Batty. *Exploring isovist fields : space and shape in architectural and urban morphology.* Environment and Planning B, vol. 28, no. 1, pages 123–150, 2001.

[Becker 08] Thomas Becker, Claus Nagel & Thomas H. Kolbe. *A Multilayered Space-Event Model for Navigation in Indoor Spaces.* In Jiyeong Lee & Sisi Zlatanova, editeurs, 3D Geo-Information Sciences, pages 61–77. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[Bergounioux 09] Maitine Bergounioux. *Méthodes mathématiques pour le Traitement d'Image.* Cours de traitement d'image, Université d'Orléans, May 2009.

[Bhattacharya 00] Prabir Bhattacharya, Haiying Liu, Azriel Rosenfeld & Scott Thompson. *Hough-transform detection of lines in 3-D space.* Pattern recogn. lett., vol. 21, no. 9, page 843 à 849, 2000.

[Bioret 09] N. Bioret, M. Servieres & G. Moreau. *Géolocalisation à partir d'images en environnement urbain en s'appuyant sur un SIG 2D.* Ingénierie des Systèmes d'Information, vol. 14, no. 5, 2009. Numéro spécial "Systèmes d'information et géolocalisation".

[Boer 08] Arnoud De Boer, Eduardo Dias & Edward Verbree. *Processing 3D Geo-Information for Augmenting Georeferenced and Oriented Photographs with Text Labels.* In Headway in Spatial Data Handling, pages 351–365. 2008.

[Boeres 04] Maria C. Boeres, Celso C. Ribeiro & Isabelle Bloch. *A Randomized Heuristic for Scene Recognition by Graph Matching.* In Celso C. Ribeiro & Simone L. Martins, editeurs, WEA, volume 3059 of *Lecture Notes in Computer Science*, pages 100–113. Springer, 2004.

[Bouziani 04] Mourad Bouziani & Jacynthe Poulio. *Inventaire et analyse des opérateurs d'appariement géométrique d'objets géographiques linéaire.* Colloque national Géomatique 2004 de l'association canadienne des sciences géomatiques, October 2004.

[Boyer 06] Edmond Boyer. *Synthèse d'images.* Cours master Réseaux Informatique et Communication Multimédia à l'UFRIMA, September 2006.

[Boyer 11] Edmond Boyer. *Lancer de rayons.* Projet master de synthèse et analyse d'images, 2011.

[Braun 09] Max Braun & Rafael Spring. *The whole story, so far*, December 2009.

[Bres 09] Stéphane Bres & Bruno Tellez. *Localisation and augmented reality for mobile applications in cultural heritage.* In 3rd International Workshop 3D ARCH 2009, February 2009.

[Bruslé 07] Lætitia Perrier Bruslé. *SIG : Théorie, définition, applications.* Rapport technique, Université Nancy 2, 2007.

[Burel 01] Gilles Burel. Introduction au traitement d'images : simulation sous matlab. Hermès science publ., Paris, 2001.

[Caprani 00] Ole Caprani, Lars Hvidegaard, Mikkel Mortensen & Thomas Schneider. *Robust and Efficient Ray Intersection of Implicit Surfaces.* Reliable Computing, vol. 6, pages 9–21, 2000.

[Certu 08a] Certu. *Le passage au RGF93 se prépare*, March 2008.

[Certu 08b] Certu. *Les projections coniques conformes 9 zones*, December 2008.

[Certu 08c] Certu. *RGF93 et Lambert 93*, October 2008.

[CETE 06] de l'Ouest CETE. *Le point sur... Le WiMAX.* http ://www.ant.developpement-durable.gouv.fr/le-point-sur-le-wimax-a9.html, December 2006. Site de l'Aménagement Numérique des Territoires.

[Chen 02] Cheng-Hsien Chen & Chen-Yi Lee. *Two-level hierarchical Z-Buffer for 3D graphics hardware.* In IEEE International Symposium on Circuits and Systems, 2002. ISCAS 2002, volume 2, pages II–253– II–256 vol.2. IEEE, 2002.

[Chen 09] Shaopei Chen, Christophe Claramunt, Cyril Ray & Jianjun Tan. *A Multiscale and Multimodal Transportation GIS for the City of Guangzhou.* In Information Fusion and Geographic Information Systems, pages 95–111. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[Cheng 01] H.D. Cheng, X.H. Jiang, Y. Sun & Jingli Wang. *Color image segmentation : advances and prospects.* Pattern Recognition, vol. 34, no. 12, pages 2259 – 2281, 2001.

[Chin 08] Tat-Jun Chin, Yilun You, Celine Coutrix, Joo-Hwee Lim, Jean-Pierre Chevallet & Laurence Nigay. *Mobile phone-based*

mixed reality : the Snap2Play game. The Visual Computer, vol. 25, pages 25–37, August 2008.

[Christine 10] C. Christine. *SIG : définition.* http ://www.scribd.com/doc/42211937/SIG-Definition, December 2010.

[Cisco 06] Cisco. *La géolocalisation par wifi.* Ciscomag, December 2006.

[Clinton 00] Bill Clinton. Statement by the president regarding the united states' decision to stop degrading global positioning system accuracy. May 2000.

[Cocquerez 85] Jean-Pierre Cocquerez & Jean Devars. *Détection de contours dans les images aériennes : nouveaux opérateurs.* 1985.

[Cohen 77] Melvin Cohen & Godfried Toussaint. *On the detection of structures in noisy pictures.* Pattern Recognition, vol. 9, no. 2, pages 95–98, 1977.

[Colleu 07] Thomas Colleu, Gaël Sourimant & Luce Morin. *Une méthode d'initialisation automatique pour le recalage de données SIG et vidéo.* In CORESA'07 - COmpression et REpresentation des Signaux Visuels, Montpellier, France, November 2007.

[Colleu 08] Thomas Colleu, Gaël Sourimant & Luce Morin. *Automatic Initialization for the Registration of GIS and Video Data.* In Proceedings of the 3DTV-Conference 2008 - The True Vision, Capture, Transmission and Display of 3D Video, Istambul, Turquie, May 2008.

[Correia 06] Paul Correia. Guide pratique du GPS. Eyroless, 5 edition, 2006.

[Crowley 10] James L. Crowley. *Détection et Description de contraste*, December 2010.

[Davis 79] Larry S. Davis & Michael L. Benedikt. *Computational Models of Space : Isovists and Isovist Fields.* vol. 11, pages 49–72, September 1979.

[De La Briere 08] Julien Pierre De La Briere. *Annotation sémantique des collections de photo personnelles.* rapport master professionnel réseaux, applications documentaires et images, Université de Caen Basse Normandie, 2008.

[Delaup 08] Benoît Delaup. *Extraction automatique et géolocalisation de contenu photographiques.* Mémoire de master, August 2008.

[Delot 06] Thierry Delot. *Techniques de positionnement en environnements mobiles.* Rapport technique, Laboratoire LAMIH, 2006.

[Denègre 96] Jean Denègre & François Salgé. Les systèmes d'information géographique. Presses universitaires de France, Paris, 1996.

[Deuff 06] Olivier Le Deuff. Folksonomies. 2006.

[Deveryware 11] Deveryware. *Comparez les techniques de localisation.* http ://www.deveryware.com/Comparez-les-techniques-de, 2011.

[Devogele 97] Thomas Devogele. *Processus d'intégration et d'appariement de Bases de Données Géographiques ; Application à une base de donnéesroutières multi-échelles.* PhD thesis, Université de Versailles-Saint Quentin en Yvelines, 1997.

[Donatis 06] Mauro De Donatis & Lorenzo Bruciatelli. *MAP IT : The GIS software for field mapping with tablet pc.* Computers & Geosciences, vol. 32, no. 5, pages 673–680, 2006.

[Dornaika 08] F. Dornaika & M. Brédif. *An efficient approach to building superstructure reconstruction using digital elevation maps.* In In : IAPRS, Volume 37 (Part 3A), July 2008.

[Douglas 08] Hackney Douglas. *Digital Photography Meta Data Overview.* Rapport technique, 2008.

[Duda 72] Richard O Duda & Peter E. Hart. *Use of the Hough transformation to detect lines and curves in pictures.* Communications of the ACM, vol. 15, no. 1, pages 11–15, January 1972.

[Dunkars 03] Mats Dunkars. *Matching of Datasets.* Proceedings of the 9thScandinavian Research Conference on Geographical Information Science, 2003.

[Ecobichon 94] Claude Ecobichon & Claude Ecobichon. L'information géographique : nouvelles techniques, nouvelles pratiques. Hermès, Paris, 1994.

[EDUCNET 11] EDUCNET. *Navstar / GPS.* http ://educol.education.fr/orbito/system/navstar/gps00.htm, 2011.

[Geoconcept 11] Geoconcept. *Geoconcept - Système d'Information Géographique (SIG) : Cartographie, geoportail, geoplanning, geomarketing, geolocalisation, geocodage.* http ://www.geoconcept.com/fr, 2011.

[Ghadirian 08] Payam Ghadirian & Ian D. Bishop. *Integration of augmented reality and GIS : A new approach to realistic landscape visualisation.* Landscape and Urban Planning, vol. 86, no. 3-4, pages 226–232, 2008.

[Grandin 06] Jean-François Grandin. *Introduction à la fusion de données*, March 2006.

[Haala 05] Norbert Haala. *Towards Virtual Reality GIS.* Photogrammetric Week 05, pages 285–294, 2005.

[Habert 00] Elisabeth Habert. *Qu'est ce qu'un système d'information géographique.* Rapport technique, Institit de recherche pour le développement, 2000.

[Hahn 07] Brian Hahn, Daniel T Valentine & Brian Hahn. Essential matlab for engineers and scientists 3rd edition. Elsevier Science Technology Books, Jordan Hill, England, 2007.

[Haklay 02] MMordechai E. Haklay. *Virtual reality and GIS. Applications, trends and directions.* Virtual Reality in Geography, New York : Taylor & Francis, pages 47–47, 2002.

[Haugeard 10] Jean Emanuel Haugeard & Sylvie Philipp-Foliguet. *Recherche d'objets par appariement de graphes combinant contours et régions.* Reconnaissance des Formes et Intelligence Artificielle, January 2010.

[Hays 08] James Hays & Alexei 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.

[Hirsch 95] J Hirsch, R L DeLaPaz, N R Relkin, J Victor, K Kim, T Li, P Borden, N Rubin & R Shapley. *Illusory contours activate specific regions in human visual cortex : evidence from functional magnetic resonance imaging.* Proceedings of the National Academy of Sciences, vol. 92, no. 14, pages 6469–6473, July 1995.

[Hoiem 05] Derek Hoiem, Alexei A Efros & Martial Hebert. Geometric context from a single image, volume 1, pages 654–661. IEEE Computer Society, 2005.

[Hoiem 07] Derek Hoiem, Alexei A Efros & Martial Hebert. *Recovering Surface Layout from an Image.* International Journal of Computer Vision, vol. 75, no. 1, pages 151–172, 2007.

[Horiki 98] Syuji Horiki, Takashi Watanabe, Akio Kimura & Norio Tayama. *Improvement of combinatorial Hough transform to compensate sampling errors of digital lines.* Systems and Computers in Japan, vol. 29, no. 2, pages 57–67, February 1998.

[Hough 62] Paul V. C. Hough. *Method and Means for Recognizing Complex Patterns.* Rapport technique 3,069,654, December 1962.

[IAAT 03] IAAT. *Cahier méthodologique sur la mise en oeuvre d'un SIG.* Rapport technique, Institut

Atlantique d'aménagement des territoires, 2003. http ://www.iaat.org/telechargement/sig.pdf.

[IGN 08a] IGN. *BD TOPO, Comparatif entre les versions 1.2 et 2.* Rapport technique, Institut Géographique National, March 2008.

[IGN 08b] IGN. *BD TOPO Version 2, Descriptif de contenu*, February 2008.

[IGN 11a] IGN. *BD ALTI Version 1, Descriptif de contenu.* Rapport technique, Institut Géographique National, October 2011.

[IGN 11b] IGN. *BD TOPO Version 2.1, Descriptif de contenu.* Rapport technique, Institut Géographique National, October 2011.

[InvenSens 11] InvenSens. *MPU-3000 Triple Axis Gyroscope with Embedded Digital Motion Processor.* http ://invensense.com/mems/gyro/mpu3000.html, 2011.

[Jambu 98] Michel Jambu. Introduction au data mining analyse intelligente des données. Collection technique et scientifique des télécommunications. Eyrolles, Paris, 1998.

[Jang 10] Chuljin Jang, Taijin Yoon & Hwan-Gue Cho. *Digital photo classification methodology for groups of photographers.* Multimedia Tools and Applications, vol. 50, pages 441–463, February 2010.

[Jaschke 08] Robert Jaschke, Andreas Hotho, Christoph Schmitz, Bernhard Ganter & Gerd Stumme. *Discovering shared conceptualizations in folksonomies.* Web Semantics : Science, Services and Agents on the World Wide Web, vol. 6, pages 38–53, February 2008.

[Jiang 08] Bin Jiang & Xintao Liu. *Automatic Generation of the Axial Lines of Urban Environments to Capture What We Perceive.* International Journal of Geographical Information Science, vol. 24, no. 4, page 13, 2008.

[Joel 91] Charre Joel & Miellet Philippe. Pratique des systèmes d'information géographique raster. GIP Reclus, Montpellier, 1991.

[Jolion 05] Jean-Michel Jolion. *Similarité entre images : primitives, graphes et vision humaine*, February 2005.

[Joliveau 07] Thierry Joliveau. *Une nouvelle solution pour coupler bases de données image et bases de données géographiques.* SIG 2007, October 2007.

[Joliveau 08a] Thierry Joliveau & Sanjay Rana. *Coupler photographies et information géographique pour décrire l'espace visible*. Sagéo 2008, Montpellier, page 15, June 2008.

[Joliveau 08b] Thierry Joliveau & Sanjay Rana. *Using Urban Viewsheds for Embedding Geographical Context in Photograph Databases of Urban Areas*. Gisruk, Manchester Metropolitan University. UK, pages 223–225, April 2008.

[Jouili 08] Salim Jouili & Salvatore Tabbone. Applications des graphes en traitement d'images, pages 434–442. 2008.

[Jullier 98] Laurent Jullier. Les images de synthése : de la technologie à l'esthétique. Nathan, Paris, 1998. Ouvrage publié sous la direction de Francis Vanoye.

[Kaiser 09] Peter K. Kaiser. The joy of visual perception. Web book, 2009. http ://www.yorku.ca/eye/thejoy.htm.

[Klipfel 07] J-P. Klipfel, S.Blanc & M.Coutard. *Les différentes technologies SIG Web Open Sources existantes.* http ://www.archive.org/details/LesDiffrentesTechnologiesSig WebOpenSourcesExistantes, December 2007.

[Kneepkens 05] R.E.J Kneepkens. *Hough based road detection.* Master's Thesis, April 2005.

[Kolbe 05] Thomas Kolbe, Gerhard Groger & Lutz Plumer. *CityGML : Interoperable Access to 3D City Models*. In Geo-information for Disaster Management, pages 883–899, 2005.

[Larose 05] Daniel T. Larose. Des données à la connaissance : une introduction au data-mining. Bases de données. Vuibert informatique, Paris, 2005. Traducton et adaptation de Vallaud Thierry.

[Laurini 10] Robert Laurini, Roula Karam, Franck Favetta & Rima Kilany. *Integration of Similar Location Based Services Proposed by Several Providers*. In Networked Digital Technologies, volume 88 of *Communications in Computer and Information Science*, page 136 à 144. Springer Berlin Heidelberg, 2010.

[LAVAL 07] Université LAVAL. *Qu'est-ce que la géomatique ?* http ://www.scg.ulaval.ca/page.php ?nom=geomatique, 2007.

[Lechner 00] Wolfgang Lechner & Stefan Baumann. *Global navigation satellite systems*. Computers and Electronics in Agriculture, vol. 25, no. 1-2, pages 67–85, January 2000.

[Leduc 11] Thomas Leduc, Francoise Chaillou & Thomas Ouard. *Towards a "typification" of the Pedestrian Surrounding Space :*

Analysis of the Isovist Using Digital processing Method. In Advancing Geoinformation Science for a Changing World, volume 1, pages 275–292, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[Lefébure 98] René Lefébure & Gilles Venturi. Le data mining. Collection Informatiques magazine. Eyrolles, Paris, 1998.

[Legifrance 06] Legifrance. *Décret num 2006-272 du 3 mars 2006.* http ://www.legifrance.gouv.fr/affichTexte.do ?cidTexte= JORFTEXT000000813996&dateTexte=, 2006.

[LEMONDE 11] LEMONDE. *Les satellites de Galileo ont été placés sur orbite par la fusée Soyouz.* http ://www.lemonde.fr/planete/article/2011/10/21/la-fusee-russe-soyouz-doit-decoller-de-guyane-apres-un-report-de-24-heures_1591686_3244.html, October 2011.

[LeSage 04] James P. LeSage & R. Kelley Pace. *Arc-Mat, a Toolbox for Using ArcView Shape Files for Spatial Econometricsand Statistics.* In Geographic Information Science, volume 3234, pages 179–190. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[Li 06] S Li & J Gong. *Mashup : a New Way of Providing Web Mapping and GIS Services.* Archives, pages 639–649, 2006.

[Liggins 08] Martin E. Liggins, David L. Hall & James Llinas. Handbook of multisensor data fusion : Theory and practice, second edition. CRC Press, 2 edition, September 2008.

[Lilly 93] Spirkovska Lilly. *A summary of image segmentation techniques.* Rapport technique NASA Technical Memorandum 104022, Ames Resaearch Center, Moffet Feld, California, 1993.

[Lim 08] Hyeeun Lim. *Raster Data.* In Encyclopedia of GIS, pages 949–955. Springer US, Boston, MA, 2008.

[Lin 01] C.-Y. Lin, M. Wu, J.A. Bloom, I.J. Cox, M.L. Miller & Y.M. Lui. *Rotation, scale, and translation resilient watermarking for images.* IEEE Transactions on Image Processing, vol. 10, pages 767–782, May 2001.

[Liu 09] Dong Liu, Xian-Sheng Hua, Linjun Yang, Meng Wang & Hong-Jiang Zhang. *Tag ranking.* 18th International World Wide Web Conference, avril 2009.

[Loic 08] Loic. *Enkin, la navigation réinventée.* http ://www.pointgphone.com/enkin-656, 2008.

[Luqman 10] Muhammad Muzzamil Luqman, Thierry Brouard & Jean-Yves Ramel. *Graphic Symbol Recognition using Graph Based Signature and Bayesian Network Classifier.* Tenth International Conference on Document Analysis and Recognition ICDAR, no. Ea 2101, page 5, 2010.

[Magdelaine 11] Christophe Magdelaine. *Représentation, acquisition des données spatiales.* http ://www.notre-planete.info/terre/outils/sig_1.php, 2011.

[Manzanera 09] Antoine Manzanera. *Traitement et reconnaissance d'images.* cours master 2 IAD, Université Pierre et Marie Curie (UPMC), Paris 6, 2009.

[MathWorks 10a] MathWorks. *Bioinformatics Toolbox.* http ://www.mathworks.fr/products/bioinfo/index.html, 2010.

[MathWorks 10b] MathWorks. *Image Processing Toolbox.* http ://www.mathworks.fr/products/image/index.html, 2010.

[MathWorks 10c] MathWorks. *Mapping Toolbox.* http ://www.mathworks.fr/products/mapping/index.html, 2010.

[Maître 85] Henri Maître. *Un panorama de la transformation de Hough.* Traitement de signal, vol. 2, no. 4, pages 305–317, 1985.

[Mayer 11] Erwin Mayer & Maxime Chipault. *Segment spatial.* http ://mayerwin.free.fr/gnss_spatial.html, 2011.

[Michel Roux 92] Jaime Lopez-krahe et Henri Maître Michel Roux. *Recalage image spot/carte routière.* pages 384–391, Washington, 1992. 17ème Congré ISPR.

[Moslah 11] Oussama Moslah. *Toward large scale urban environment Moedeling from images.* Informatique, Cergy-Pontoise, July 2011.

[Mougel 03] Ludovic Mougel. *Elaboration d'un réseau de points GPS par densification locale du RBF. Utilisation du nouveau système géodésique.* Rapport technique, Ecole Supérieure des Géomètres et Topographes, September 2003.

[Muller 06] Julien Muller. *Développement d'un Système d'Informations Géographiques sur les d'gradations du littoral : l'île de Noirmoutier.* Rapport technique, ENSG, Paris, June 2006.

[Nagel 10] Claus Nagel, Thomas Becker, Robert kaden, Ki-Joune Li, Jiyeong Lee & Thomas H. Kolbe. *Requirements and Space-Event Modeling for Indoor Navigation.* Open Geospatial Consortium, December 2010.

[Nasr 87] Jean-Marc Nasr. L'image numérique. Ed. Fréquences, Paris, 1987.

[Navigation 98] Inertial Navigation. *Inertial Navigation-Forty Years of Evolution.* Evolution, vol. 13, no. 3, pages 140–149, 1998.

[Nixon 02] Mark S. Nixon & Alberto S Aguado. Feature extraction and image processing for computer vision. Elsevier Science & Technology Books, January 2002.

[Nogueira 09] Sergio Nogueira. *Localisation de mobiles par construction de modèles en 3D en utilisant la stéréovision.* Informatique, Université de technologie Belfort-Montbéliard, December 2009.

[Nozick 10] Vincent Nozick. *Détection de lignes.* Rapport technique, Ecole d'ingénieur IMAC, Marne-la-Vallée, 2010.

[Olteanu1 08] Ana-Maria Olteanu1. *Fusion de connaissances imparfaites pour l'appariement de données géographiques : proposition d'une approche s'appuyant sur la théorie des fonctions de croyance.* PhD thesis, Paris-Est, October 2008.

[Petitesexperiences 11] Petitesexperiences. *Illusion d'optique, cube imaginaire.* http ://www.petitesexperiences.com/illusions-doptique/illusion-doptique-cube-imaginaire/, 2011. Dernière visite, le 01/10/2011.

[Porchet 02] Michel Porchet & Michel Porchet. La production industrielle de l'image : critique de l'image de synthèse. L'Harmattan, Paris - Budapest - Torino, 2002.

[Pumain 03] Denise Pumain & Michele Béguin. La représentation des données géographiques : Statistique et cartographie. Armand Colin, 2ème édition edition, August 2003.

[Purves 08] Ross S Purves, Alistair Edwardes & M Sanderson. *Describing the where-improving image annotation and search through geography.* In Proceedings of the workshop on Metadata Mining for Image Understanding, pages 105–113. Insticc Press, January 2008. 3rd International Conference on Computer Vision Theory and Applications, Funchal, Madeira (Portugal).

[Purves 10] Ross S Purves, Alistair Edwardes, Xin Fan, Mark Hall & Martin Tomko. *Automatically generating keywords for georeferenced images.* GISRUK, 2010.

[QGIS 11] QGIS. *A propos de QGIS.* http ://www.qgis.org/fr/a-propos-de-qgis.html, 2011.

[Qureshi 06] R J Qureshi, Ramel J Y & Cardot H. De l'appariement de graphes symboliques l'appariement de graphes numriques : Application la reconnaissance de symboles, pages 31–36. 2006.

[Rabbani 02] Majid Rabbani & Rajan Joshi. *An overview of the JPEG 2000 still image compression standard.* Signal Processing : Image Communication, vol. 17, no. 1, pages 3–48, January 2002.

[Ramel 10] Jean-Yves Ramel. *Graphes en analyse d'images et reconnaissance des formes.* Présentation au cours de la Journée Informatique de la Région Centre à Blois, January 2010.

[Rana 06] Sanjay Rana. *Ray Tracing in GeoComputation.* AAG 2006 Chicago, 2006.

[Ray 09] Cyril Ray. *Mobilité, contexte et Localisation*, September 2009.

[Riesmeier 01] J. Riesmeier, M. Eichelberg, K. Kleber, H. Oosterwijk, S. von Gehlen, D.H.W. Gronemeyer & P. Jensch. *DICOM Structured Reporting : prototype implementation.* International Congress Series, vol. 1230, pages 795–800, June 2001.

[Rollet 05] Claire Rollet. *Géoréférencement en zone intertidale par GPS.* Rapport technique, IFREMER, 2005.

[Rosenfeld 69] Azriel Rosenfeld. Picture processing by computer. Academic press, New York - London, 1969.

[Rosenfeld 82] Azriel Rosenfeld & Avinash C Kak. Digital picture processing. Academic Press, San Diego - New York - Berkeley, 1982.

[Rossant 09] Florence Rossant, Itebeddine Ghorbel, Isabelle Bloch, Michel Paques & sarah Tick. *Automated Segmentation of Retinal Layers in OCT Imaging and Derived Ophtalmic Measures.* In ISBI 2009, Boston, USA, 2009.

[Ruas 02] Anne Ruas. Généralisation et représentation multiple. Hermes science, Paris, 2002.

[Rudiger 11] Romain Rudiger. *Utilisation du GPS dans le domaine de l'industrie et la gestion de production - Wiki de Romain RUDIGER.* http ://romain.novalan.fr/wiki/Utilisation_du_GPS_dans_le_domaine_de_l'industrie_et_la_gestion_de_production#Le_DGPS_.28Differential_Global_Positioning_System.29_et_la_marine., 2011.

[Sagem 11] Sagem. *Equipements de navigation inertielle - Sagem.* http ://www.sagem-ds.com/spip.php ?rubrique184&lang=fr, 2011.

[Sig-france 11] Sig-france. *Sig-france.org : Logiciels SIG.* http ://sig-france.free.fr/dossiers/logiciels.html, 2011.

[Sigea 11] Sigea. *Les projections cartographiques.* Ressources pédagogiques, 2011. (Systèmes d'Information Géographique pour l'Enseignement Agricole).

[Sigurbjornsson 08] Borkur Sigurbjornsson & Roelof Van Zwol. *Flickr tag recommendation based on collective knowledge.* Proceeding of the 17th international conference on World Wide Web WWW 08, page 327, 2008.

[Smart 09] Philip Smart, Florian Twaroch & Christopher Jones. *Final Toponym Ontology Prototype.* Rapport technique 6.5, November 2009.

[Smith 11] Alibaster Smith. *How to Calibrate a Digital Compass | eHow.com.* http ://www.ehow.com/how_4968786_calibrate-digital-compass.html, 2011.

[Sorel 10] Johann Sorel. *Les outils de développement cartographique.* Rapport technique, June 2010.

[Sorlin 05] Sébastien Sorlin & Christine Solnon. *Similarité de graphes : une mesure générique et un algorithme tabou réactif.* In Emmanuel Guéré, editeur, 7èmes rencontres nationales des jeunes chercheurs en intelligence artificielle, RJCIA'2005, pages 253–266. Presses Universitaires de Grenoble, June 2005.

[Sorlin 06] Sébastien Sorlin, Christine Solnon & Jean-Michel Jolion. *A Generic Graph Distance Measure Based on Multivalent Matchings.* Rapport technique RR-LIRIS-2006-013, LIRIS UMR 5205 CNRS/INSA de Lyon/Université Claude Bernard Lyon 1/Université Lumière Lyon 2/École Centrale de Lyon, June 2006.

[Soulier 73] Jacques Soulier & Pierre Berenguier. *Système Sylosat de localisation précise par satellites transit.* May 1973.

[Sourimant 07] Gaël Sourimant. *Reconstruction de scènes urbaines à l'aide de fusion de donnés de type GPS, SIG et Video.* PhD thesis, Thèse de doctorat en Informatique, Université de Rennes 1, France, dec 2007.

[Sourimant 09] Gaël Sourimant, Thomas Colleu, Vincent Jantet & Luce Morin. *Recalage GPS / SIG / Video, et synthèse de textures de bâtiments.* In Actes de la Conférence COmpression et REpresentation des Signaux Audiovisuels, CORESA'2009, Toulouse, France, pages 1–6, Toulouse, France, March 2009. sourimant2009a contrat CRE France Telecom-INRIA / projet collaboratif FUTURIM@GE du pôle Image et Réseaux.

[Suh 07] Bongwon Suh & Benjamin B. Bederson. *Semi-automatic photo annotation strategies using event based clustering and clothing based person recognition.* Interacting with Computers, vol. 19, no. 4, pages 524–544, July 2007.

[Suleiman 11] Wassim Suleiman, Eric Favier & Thierry Joliveau. *Buildings Recognition and Camera Localization Using Image Texture Description.* Paris, July 2011. SAGEO.

[Suleiman 12] Wassim Suleiman, Eric Favier & Thierry Joliveau. *A new algorithm for 3D Isovist.* Bonn, Germany, August 2012. 15th International Symposium on Spatial Data Handling Geospatial dynamics, geosimulation and exploratory visualization.

[Taillandier 04] Franck Taillandier. *Reconstruction du bâti en milieu urbain : une approche multi-vues.* Thèse en mathématiques, Ecole polytechnique, October 2004.

[Theoharatos 06] Ch Theoharatos, V K Pothos, N A Laskaris, G Economou & S Fotopoulos. *Multivariate image similarity in the compressed domain using statistical graph matching.* Pattern Recognition, vol. 39, no. 10, pages 1892–1904, 2006.

[Thomas 05] Berlage Thomas. *Analyzing and mining image databases.* Drug Discovery Today, vol. 10, no. 11, pages 795–802, June 2005.

[Torniai 07] Carlo Torniai, Steve Battle & Steve Cayzer. *Sharing, Discovering and Browsing Geotagged Pictures on the Web.* Rapport technique, Digital Media Systems Laboratory, HP Laboratories Bristol, 2007.

[Tufféry 02] Stéphane Tufféry. Data mining et scoring bases de données et gestion de la relation client. InfoPro Série traitement des données. Dunod, Paris, 2002.

[Tupin 06] Florence Tupin. *Graphes en analyse d'images, vision, reconnaissance*, May 2006.

[Usery 05] E. Lynn Usery, Michael P. Finn & Michael Strarbuck. *Integrating data layers to support the national map of the United States.* International Cartographic Conference, 2005.

[Valentin 09] ROBERT Valentin. *Reconnaissance de formes - Transformée de Hough.* Rapport technique, ENSEIRB-MATMECA à l'Institut Polytechnique de Bordeaux, Bordeaux, 2009.

[Van Der Drift 09] Marcel Van Der Drift, Stephen Boyd Davis, Rob Van Kranenburg, sophie Hope & Tom Stanforf. Emotional cartography technologies of the self. CHRISTIAN NOLD, London, April 2009.

[Volz 05] Steffen Volz. *Data-Driven Matching of Geospatial Schemas.* In Anthony Cohn & David Mark, editeurs, Spatial Information Theory, volume 3693 of *Lecture Notes in Computer Science*, pages 115–132. Springer Berlin / Heidelberg, 2005.

[von Gioi 10] Rafael Grompone von Gioi, Jeremie Jakubowicz, Jean-Michel Morel & Gregory Randall. *LSD : A Fast Line Segment Detector with a False Detection Control.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, pages 722–732, 2010.

[Wan 09] Minghua Wan, Zhihui Lai, Jie Shao & Zhong Jin. *Two-dimensional local graph embedding discriminant analysis (2DLGEDA) with its application to face and palm biometrics.* Neurocomputing, vol. 73, no. 1-3, pages 197 – 203, 2009.

[Wimmer 99] Michael Wimmer, Markus Giegl & Dieter Schmalstieg. *Fast Walkthroughs with Image Caches and Ray Casting.* Computers & Graphics, vol. 23, no. 6, pages 831–838, 1999.

[Wladyslaw 94] Skarbek Wladyslaw & Koschan Andreas. *Color image segmentation, a survey.* Technical report 94-32, Institute for Technical Informatics, Technical University of Berlin, Berlin, Germany, October 1994.

[Woloszyn 11] Philippe Woloszyn & Thomas Leduc. *A landscape potential characterization : spatial template of pedestrian ambient fields within the urban fabric.* In 1e conférence intercontinentale en intelligence territoriale (IT-Gatineau 2011), Gatineau, Canada, October 2011.

[Wu 09] Mei-Yi Wu, Chuan-Chi Hsu & Jia-Hong Lee. *A GeoTagging Scheme Using Image Steganography and GPS Information Authentication.* pages 1245–1248, September 2009.

[Yaegashi 09] Keita Yaegashi & Keiji Yanai. *Can Geotags Help Image Recognition ?* In Toshikazu Wada, Fay Huang & Stephen Lin, editeurs, Advances in Image and Video Technology, volume 5414, pages 361–373. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[Zampelli 06] Stéphane Zampelli, Yves Deville & Pierre Dupont. *Elimination des symétries pour l'appariement de graphes.* In Deuxièmes Journées Francophones de Programmation par Contraintes (JFPC06), Nîmes - Ecole des Mines d'Alès / France, 2006.

Titre : ***Interfaçage de bases de données photographiques et géographiques par appariement de lignes.***

Résumé :

Interfacer les photographies et les bases de données géographiques est une démarche utile pour les architectes, les historiens, les géographes mais aussi pour le grand public. Affecter une position géographique à une photographie permet en effet de décrire son contenu potentiel en s'appuyant sur les informations contenues dans une base de données géographique.

L'objectif de la thèse est de proposer une méthode permettant de localiser automatiquement une photographie urbaine en la comparant avec des images de synthèse en 3D générées de manière systématique à partir d'un modèle virtuel des bâtiments. Le principe est que la photographie partage avec les images des caractéristiques géométriques qui permettent de les rapprocher et donc de déduire la localisation de la photographie à partir de la position connue de l'image.

La méthode utilisée pour retrouver les images de synthèse correspondantes est un appariement entre les lignes présentes dans la photographie et les lignes détectées dans les images de synthèse par la transformation de Hough. Cet appariement est suivi par une analyse statistique permettant de proposer une localisation probable avec une valeur d'approximation associée.

Malgré les obstacles présents dans les photographies, cette approche utilisant uniquement les lignes est une solution simple et potentiellement efficace pour le positionnement de photographies. Plusieurs scénarios d'usage sont proposés : la géolocalisation d'une image, la validation du positionnement d'une image localisée et l'utilisation de photographies pour repérer des changements dans le paysage urbain.

Mots clés : SIG, photo, appariement de lignes, lignes de Hough, reconstitution 3D, étiquettes, ville

Title: ***Interfacing a photographic database and a geographical database by line matching.***

Abstract:

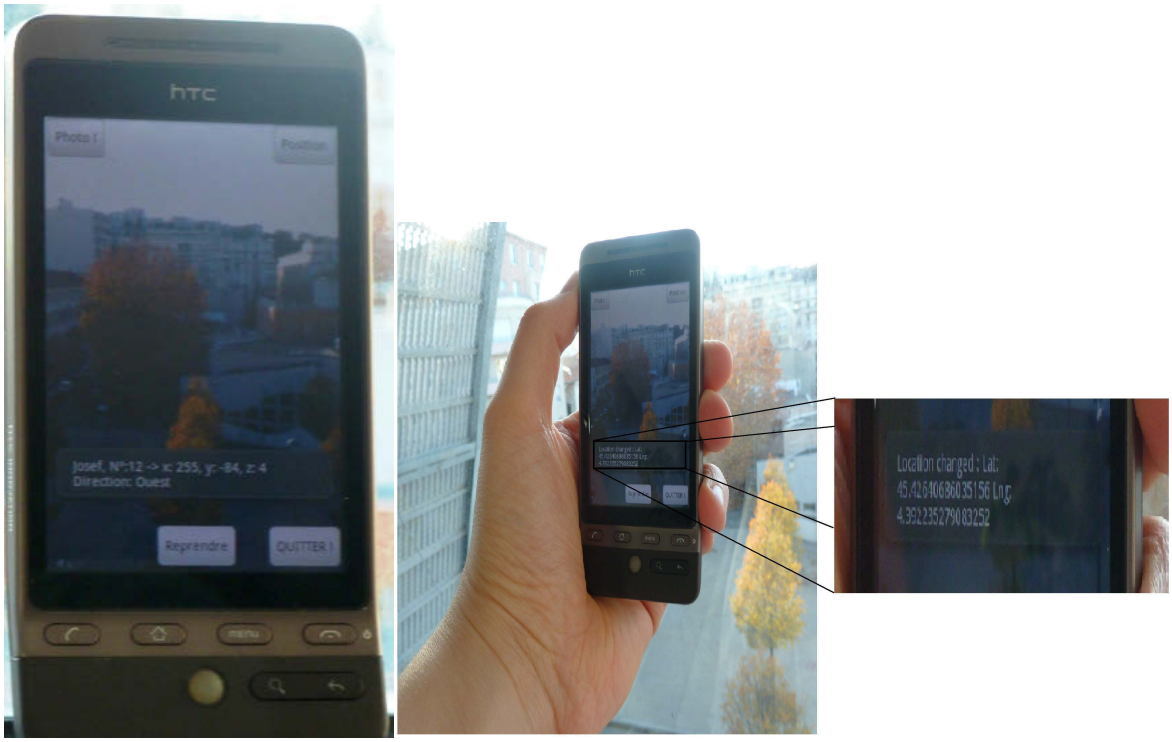
Interfacing photographs with geographic databases is a useful approach not only for architects, historians, geographers but also for a general audience. By assigning a geographical position to a picture it becomes possible to describe its potential content based on the information contained in the geographic database.

The aim of the thesis is to propose a method to locate automatically an urban photograph by comparing this photograph with 3D images generated systematically from a virtual model of the buildings. The principle is to put in relation a photograph with the synthetic images that are sharing analog geometric characteristics and therefore to deduce the location of the photograph from the known position of the image.

In order to find synthetic images related to a photograph the method used is a matching between detected lines in the photograph and the images based on Hough transform. This matching is followed by a statistical analysis to propose a probable location with an estimation of accuracy.

Despite the obstacles present in the photographs, this approach using only lines is simple and potentially effective for positioning photographs in urban space. Several scenarios of use are proposed: first location of an image, validation of the position of an image broadly located and use of photographs to identify change in the urban landscape.

Keywords: GIS, photo, matching lines, lines Hough, 3D reconstruction, tag, city



(a) Les degrés de rotation (b) Latitude et longitude

FIGURE B.4 – Application d'ajout des données de positionnement sur les photographies prises par le HTC Hero.



FIGURE B.5 – Imprimé écran de la prise N°2 avec les degrés de rotation sur les axes X, Y et Z. (Source : capture d'écran de l'XPERIA X10)

géographique et avec une boussole analogique pour préciser la direction de visée pendant la prise de la photographie.