Géolocalisation en milieu urbain par appariement entre une collection d'images et un SIG 2D.

Nicolas Bioret* — **Guillaume Moreau*** — **Myriam Servières***

* *Laboratoire CERMA UMR CNRS 1563*
rue Massenet, F-44319 Nantes Cedex 3
{prenom.nom}@ec-nantes.fr

RÉSUMÉ. *Nous proposons une approche de géolocalisation combinant d'une part de l'extraction d'information sur des images de façades et d'autre part des requêtes dans un Système d'Information Géographique 2D (SIG). À partir de photographies prises dans la rue et en utilisant l'hypothèse de façades verticales, nous calculons les angles et les rapports des largeurs entre les façades reconnues. Nous calculons aussi les incertitudes sur ces éléments géométriques en fonction des erreurs de mesure dans l'image, puis nous les transformons en tolérances dans des requêtes sur les données d'un SIG. Un ensemble de poses est alors généré (position 2D et orientation dans l'espace) et est ensuite filtré pour éliminer les poses incorrectes. Par exemple nous nous assurons de la cohérence avec les informations issues de la couche route du SIG, ou bien nous combinons les informations issues de plusieurs images prises avec des positions voisines. L'utilisation de tolérances dans les requêtes garantit que la pose correcte n'est jamais éliminée. Nous montrons, sous réserve que le SIG soit à jour, que l'on peut éliminer toutes les poses incorrectes pour obtenir une unique pose correcte pour la caméra. Nous présentons dans cet article nos expériences, basées sur des données SIG d'un quartier de la ville de Nantes, France.*

ABSTRACT. *We propose a localization approach in outdoor environment, which combines both information extraction from facade pictures and queries in a 2D Geographic Information System (GIS). Using photographs taken in the streets and based on the vertical facade assumption, we compute angle and width ratios between facades recognized in the images. We also compute the uncertainty associated to the angles and ratios, which depends of image measurements uncertainty. Then, all the information extracted from the images is transformed into GIS queries, which return a set of every possible poses (2D position and 3D orientation). In order to filter out every wrong pose, we check if each pose is consistent with other information in the GIS lay-*

ISI, pages 1 à 25

ers, for instance the road layer. We can also gather and compare poses resulting from different photographs taken in the same neighborhood. Use of tolerances in every query ensures that only wrong poses are filtered out. Our experiences, based on street photographs taken in the city of Nantes, France and using a correctly updated GIS, show that it is possible to filter out every wrong poses and get only the true pose for the camera.

MOTS-CLÉS : *géolocalisation, SIG, image, pose, orientation, façade, calcul d'angles, calcul de rapports de longueur*

KEYWORDS: *outdoor localization, GIS, image, pose, orientation, facade, angle computation, length ratios computation*

1. Introduction

L'utilisation des SIG sur le terrain fait l'objet d'une forte demande dans les milieux urbains. Au sein des collectivités territoriales, des opérateurs situés sur le terrain, travaillant par exemple sur les réseaux d'eau, de gaz ou de transport ont souvent besoin de données géographiques pour augmenter leur connaissance de leur environnement, ou à l'inverse, ils peuvent vouloir ajouter des connaissances sur leur environnement dans une base de données géographiques. Que ce soit pour retrouver ou mettre à jour de l'information géographique sur le terrain, une géolocalisation est nécessaire. Elle permet d'utiliser une interface spécifique pour visualiser et interagir avec les données. Par exemple une interface cartographique demande de connaître la position de l'utilisateur en 2D. On peut également imaginer des interfaces ayant recours à de la Réalité Augmentée qui permettraient de faire le lien entre des images d'une scène réelle et des données géographiques et de surcroit permettraient de visualiser des données 3D. Ce type d'interfaces demande de connaitre non seulement une position exacte, mais également une orientation dans l'espace, et ce dans un contexte temps réel.

C'est vers ce type d'interface de Réalité Augmentée en milieu urbain que se dirige notre recherche. Nous souhaitons réaliser une application permettant de faire l'acquisition de données 3D sur les bâtiments, parmi lesquelles les textures des façades verticales ou toutes les informations sémantiques pouvant en être extraites. Actuellement, de telles informations ne peuvent pas être facilement extraites de photographies aériennes. Par contre, à partir de photographies prises dans la rue, si elles peuvent être géolocalisées et orientées, nous sommes en mesure de produire des informations géographiques qui ensuite peuvent être intégrées au SIG. L'un des enjeux actuels est d'améliorer la qualité des données afin d'aider à la décision dans le domaine de l'aménagement du territoire et la politique urbaine. De nombreuses simulations architecturales, thermiques, aérauliques, sonores, lumineuses, s'appuient actuellement sur des données urbaines simplistes voire sur des cas théoriques pour justifier des aménagements urbains. L'accès à des données plus fiables, via un outil de mesure, d'acquisition ou de validation sur le terrain permettrait d'améliorer l'exactitude des données géographiques utilisées, tant sur le plan de la géométrie que des données attributaires, et par conséquent d'accroitre la fiabilité des simulations.

L'originalité de notre approche est l'utilisation du SIG pour calculer de manière exacte la pose des images. Un calcul de pose demande à la fois une géolocalisation dans l'espace et une orientation dans l'espace. C'est donc 6 degrés de liberté qu'il faut déterminer et non pas 2 comme dans les approches de géolocalisation pour des interfaces cartographiques. Deux degrés de liberté en rotation peuvent être déterminés en faisant l'hypothèse de verticalité des façades photographiées. Pour déterminer les 4 autres degrés de liberté, nous allons chercher à mettre en correspondance le SIG avec le contenu de l'image. Le SIG que nous utilisons ne contient à l'origine que des informations bidimensionnelles, à savoir une couche bâtiments 2D et une couche route. L'utilisation d'un GPS permet de réduire fortement la zone dans laquelle il est probable de trouver la position recherchée. Toutefois, l'utilisation d'un GPS en milieu urbain présente quelques inconvénients. Tout d'abord, du fait des nombreux

multitrajets subis par les signaux GPS, l'exactitude du positionnement par GPS est fortement dégradée par rapport à son fonctionnement en milieu parfaitement dégagé. Ensuite, dans certaines rues étroites des centres urbains denses, la faible portion de ciel visible fait que le récepteur GPS ne peut pas verrouiller les 4 satellites nécessaires à son fonctionnement. Enfin, un GPS qui n'est pas mis en mouvement ne permet pas d'enregistrer de données d'orientation sur les images. Nous proposons donc de partir d'une connaissance approximative de notre position, par exemple la connaissance du quartier dans lequel nous nous trouvons.

Un avantage d'une approche faiblement instrumentée, c'est-à-dire utilisant seulement des données images, est qu'elle s'applique à de la géolocalisation d'images lorsque l'on n'est plus sur le terrain. De nombreuses photographies prises en extérieures commencent à être géolocalisées pour offrir des outils de visualisation en milieu urbain (Snavely *et al.*, 2006).

Cet article est structuré de la manière suivante. La partie 2 présente un état de l'art sur la problématique du calcul de pose en milieu extérieur, et en particulier en ce qui concerne le milieu urbain. Un descriptif de notre approche est proposé en partie 3. Puis nous détaillons en parties 4 et 5 comment des informations pertinentes peuvent être extraites des images, et nous expliquons comment les combiner aux informations SIG dans la partie 6. La partie 7 présente les résultats obtenus sur un nombre significatif de photographies et la partie 8 propose une discussion sur les orientations futures de cette approche de localisation.

2. Etat de l'Art

Les approches de localisation en milieu extérieur à partir d'images sont principalement issues de deux domaines de recherche : celui de la robotique d'une part et celui de la réalité augmentée d'autre part. On peut les classifier selon le type de données géographiques utilisées pour géolocaliser une image : des données purement vectorielles peuvent être utilisées, ou bien des données images, ou encore des modèles texturés qui combinent des images et des données vectorielles.

2.1. *Localisation par base d'image*

D'excellents exemples de localisation basés essentiellement sur de l'appariement d'images sont proposés dans (Zhang *et al.*, 2006) et (Steinhoff *et al.*, 2007). Toutefois, la taille d'une base d'images peut augmenter très rapidement lorsque l'on cherche à couvrir une zone plus large, et par conséquent, les requêtes d'images dans la base risquent de se faire en un temps prohibitif. Plusieurs idées ont été proposées pour essayer de réduire la taille de la base d'images à parcourir, et pouvoir conserver un temps de calcul proche du temps réel. Par exemple (Johansson *et al.*, 2002) utilisent une base de textures simplifiées au lieu des photographies complètes des façades, et (Robertson *et al.*, 2004) ne conservent que les textures localisées au voisinage de points d'intérêts.

(Lee *et al.*, 2002) utilisent des images sphériques, qui permettent d'utiliser significativement moins d'images pour couvrir la même zone, mais l'utilisation de caméras sphériques n'est pas encore très répandue. D'autres approches essaient non pas de réduire la taille de la base d'image mais plutôt de diminuer le temps de requête dans ces bases (Fritz *et al.*, 2005; Goedeme *et al.*, 2004; Shao *et al.*, 2003; Steinhoff *et al.*, 2007). Actuellement on arrive à des temps de réponse intéressants pour de la localisation en milieu extérieur en temps réel, et ce pour des zones couvertes assez larges. Par exemple (Robertson *et al.*, 2004) travaillent sur une rue comportant au total 2km de façades, alors que (Lee *et al.*, 2002) travaillent sur une zone plus restreinte de 60m par 15m. (Steinhoff *et al.*, 2007) travaillent dans une zone de 200m par 200m, et envisage de coupler son approche avec une localisation par GPS. Ce dernier permettrait de ne sélectionner qu'une certaine partie des images dans une base d'images assez large, améliorant le temps de réponse. Cependant, la question de la création d'une base d'images géolocalisée à grande échelle reste un souci majeur dans ce type d'approche (Teller *et al.*, 2001) même si des projets récents tels que Google Street View (`http://maps.google.com`) ont commencé la constitution de telles bases d'images.

2.2. *Localisation basée sur un modèle vectoriel*

Au contraire, les modèles vectoriels sont plus légers, mais ont tendance à trop simplifier la réalité qu'ils modélisent. On peut distinguer les modèles vectoriels selon leur étendue. Dans la plupart des approches rencontrées, en effet, on travaille au voisinage d'un seul bâtiment, dont on connait un modèle 3D. Le calcul de pose est alors basé sur du suivi optique du modèle 3D dans une séquence vidéo, ce qui permet d'obtenir une pose de manière exacte. Cependant une telle approche demande au préalable une initialisation suffisamment exacte de la pose. Celle-ci peut être obtenue à partir de correspondances établies manuellement (Behringer *et al.*, 2002; Jiang *et al.*, 2004; Ribo *et al.*, 2002) entre le modèle 3D et l'image, même si cela écarte la possibilité d'une application fonctionnant en temps réel. On peut noter que SoftPosit (David *et al.*, 2004) proposent une approche intéressante pour établir automatiquement des correspondances entre un modèle 3D et une image. Toutefois, aucun résultat convaincant n'a été présenté pour une utilisation en extérieur. L'utilisation d'un GPS associé à un capteur d'orientation (Haala *et al.*, 2003) demeure une solution très pratique pour réaliser l'initialisation de la pose de manière automatique et en temps réel. Par contre on n'obtient généralement pas le niveau d'exactitude requis pour une pose initiale. D'autres approches plus ambitieuses travaillent avec un modèle vectoriel à l'échelle d'un quartier ou d'une ville. (Sourimant *et al.*, 2007) utilisent un modèle 3D de plusieurs bâtiments, et avec l'utilisation d'un GPS, il peut se ramener aux cas précédents où l'on travaille avec un seul bâtiment. (Robertson *et al.*, 2002) travaillent avec un modèle vectoriel en 2D, comprenant l'emprise au sol de quelques bâtiments, mais il ne travaille pas sur du suivi optique. La pose est alors calculée à partir de correspondances établies manuellement entre les images et le modèle vectoriel 2D. Encore une fois, cela ne permet pas de travailler dans un contexte de temps réel, et cela suppose par ailleurs que l'utilisateur sache établir de telles correspondances.

2.3. *Localisation basée sur un modèle vectoriel texturé*

Les modèles vectoriels texturés sont les données les plus riches à partir desquelles on va pouvoir associer la précision du suivi optique à la robustesse de la corrélation d'images. Cela permet de travailler sur des étendues significatives (plusieurs bâtiments) dans un contexte temps réel. Comme pour le suivi optique basé sur des modèles géométriques, il existe un problème d'initialisation. Les travaux sur le sujet (Reitmayr *et al.*, 2006; Reitmayr *et al.*, 2007; Cappelle *et al.*, 2002; Coors *et al.*, 2000) utilisent tous un GPS, puis raffinent la position avec une approche de balayage. On peut noter que dans les travaux utilisant un GPS, la précision annoncée pour le GPS est inférieure à 10 m (Haala *et al.*, 2003; Sourimant *et al.*, 2007; Cappelle *et al.*, 2002; Reitmayr *et al.*, 2007) ce qui nous semble être une prévision trop optimiste en milieu urbain dense. En raison des nombreux multitrajets, la dégradation du signal peut être importante, et particulièrement pour les récepteurs très précis qui sont également très sensibles à ces phénomènes. De plus, la perte totale de signaux GPS est fréquente dans les rues très encaissées où la faible portion de ciel visible ne permet pas le verrouillage de 4 satellites.

2.4. *Localisation à partir d'une reconstruction de la structure 3D*

Une autre catégorie d'approche est celle qui utilise une reconstruction partielle de la scène, utilisant l'algorithme Structure from Motion (Moslah *et al.*, 2008; Snavely *et al.*, 2006; Bartczak *et al.*, 2007) ou des outils issus du domaine de la photogrammétrie (Schindler *et al.*, 2003) ou encore des approches utilisées en robotique (Wang *et al.*, 2007). Le modèle obtenu est alors un nuage de points 3D, qui peut être utilisé pour de la localisation, et qui peut se faire en temps réel (Bartczak *et al.*, 2007; Wang *et al.*, 2007). Toutefois, recaler automatiquement ce modèle local sur un modèle global est considéré comme une problématique difficile (Wang *et al.*, 2007). Des modèles vectoriels peuvent être utilisés pour faire un tel recalage : (Moslah *et al.*, 2008) utilisent un modèle 3D de bâtiments mais son approche reste locale. (Wang *et al.*, 2007) proposent un recalage sur un réseau routier. Des modèles rasters peuvent être utilisés : (Snavely *et al.*, 2006) effectuent un recalage sur une photographie aérienne et sur un modèle numérique de terrain. Les recalages proposés sont fait manuellement, or cette opération doit être automatisée pour pouvoir permettre une localisation en temps réel ou légèrement différé. Par ailleurs, il faut que la zone concernée dans le modèle global puisse être retrouvée manuellement, ce qui est parfois irréalisable car le modèle global peut être très étendu.

3. Approche proposée

L'état de l'art précédent nous amène à adopter l'approche suivante. Premièrement, la difficulté à construire une base d'images géolocalisées avec une large couverture invite à utiliser des données vectorielles comme données de base. Deuxièmement, l'uti-

lisation d'un GPS en milieu urbain dense peut se faire sous réserve de ne pas adopter une attitude trop optimiste sur sa précision. Eventuellement, la donnée GPS peut être remplacée par une indication de l'utilisateur sur sa position approximative. Troisièmement, nous pensons que la reconstruction d'une structure 3D suivie d'un recalage automatique sur un modèle vectoriel peut fournir une initialisation précise de la pose. C'est ce que nous allons particulièrement développer dans cet article. Quatrièmement, à partir d'une initialisation précise, une localisation par suivi optique d'un modèle vectoriel dans des images permettra une géolocalisation et une orientation dans l'espace de l'utilisateur.

Notre approche pour parvenir à une géolocalisation et une orientation est la suivante : tout d'abord, des éléments caractéristiques de la géométrie de la scène sont extraits, tels que des points de fuite et des arêtes verticales séparant des façades voisines. Des relations de voisinage sont également établies entre les différentes zones de façades identifiées. À partir de ces éléments, il est possible de calculer les rapports de longueur et les angles entre les différentes façades de la scène, ainsi que les incertitudes sur ces rapports de longueur et angles. Ces éléments caractéristiques de la scène vont permettre d'établir automatiquement des correspondances possibles entre le SIG et l'image et de calculer toutes les poses possibles résultantes. Généralement, de nombreuses hypothèses sont retenues, et c'est par une analyse plus fine de cohérence que l'on parvient à réduire ce nombre d'hypothèses à 1. L'utilisation de tolérances permet de s'assurer que l'hypothèse restante est la solution exacte, sauf dans les cas où le SIG n'est pas à jour.

4. Extraction de la géométrie à partir d'une image

Dans cette partie on se place dans un cadre géométrique idéal, ce qui revient à faire l'hypothèse qu'il n'y a aucune erreur dans les mesures. L'intégration des erreurs de mesure dans les formules est expliquée dans la partie suivante.

4.1. *Repères utilisés*

On définit tout d'abord un repère de référence lié au SIG (O_o, X_o, Y_o, Z_o). Ensuite, chaque façade verticale F peut être associée à un repère (O_f, X_f, Y_f, Z_f) avec Y_f vertical, vers le haut et Z_f orthogonal au plan de la façade et dirigé de la façade vers la camera. La caméra est associée à un repère (O_c, X_c, Y_c, Z_c) où X_c et Y_c sont dans le plan de l'image, Y_c est dirigé du haut vers le bas et Z_c est orthogonal au plan de l'image et dirigé de la caméra vers la scène. Enfin, le repère image est (O_i, X_i, Y_i), avec O_i le coin en haut à gauche de l'image. X_i est dans le plan de l'image et orienté de gauche à droite et Y_i est dans le plan de l'image et orienté du haut vers le bas.

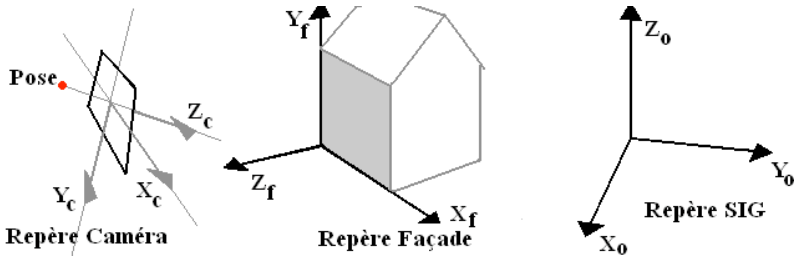


Figure 1 – Repères utilisés.

4.2. *Relations directes entre repères*

La relation entre le repère caméra et le repère façade est la composée d'une rotation et d'une translation. On note R_{FC} la matrice de rotation de la façade vers la caméra et $O_{fCamera}$ le vecteur des coordonnées de O_f dans le repère caméra. Pour un point $M = (M_{xf}, M_{yf}, M_{zf})$ dans le repère façade et (M_{xc}, M_{yc}, M_{zc}) dans le repère caméra on a la relation

$$\begin{bmatrix} M_{xc} \\ M_{yc} \\ M_{zc} \end{bmatrix} = R_{FC} \begin{bmatrix} M_{xf} \\ M_{yf} \\ M_{ff} \end{bmatrix} + O_{fCamera} \qquad [1]$$

La relation entre le repère caméra et le repère image est une projection de centre O_c caractérisée par une matrice K, appelée matrice de calibration.

$$K = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \qquad [2]$$

Entre un point m de l'image et un point correspondant M dans le repère façade, on a alors l'égalité projective suivante :

$$\begin{bmatrix} m_x \\ m_y \\ m_z \end{bmatrix} = K \left[R_{FC} \mid OfCamera \right] \begin{bmatrix} M_{xf} \\ M_{yf} \\ M_{zf} \\ M_{tf} \end{bmatrix} \qquad [3]$$

où $m = (m_x, m_y, m_z)$ en notation homogène et $M = (M_{xf}, M_{yf}, M_{zf}, M_{tf})$ en notation homogène.

4.3. *Relations inverses*

À partir des coordonnées d'un point dans le repère de l'image, on cherche à trouver les coordonnées correspondantes dans les autres repères. Il s'agit d'un problème inverse, avec a priori plusieurs solutions à moins de rajouter des contraintes 3D sur la géométrie de la scène. Or il existe des contraintes de parallélisme et d'orthogonalité dans une scène 3D présentant des bâtiments, et de plus on peut supposer que les points de fuite dominants associés à une façade sont respectivement verticaux et horizontaux. Pour une façade F_i on note p_i le point de fuite horizontal associé. Le point de fuite vertical, qui est commun à toutes les façades, est noté p_v. Dans le repère lié à l'image on a $p_i = (p_{ix}, p_{iy}, p_{iz})$ et $p_v = (p_{vx}, p_{vy}, p_{vz})$. p_v correspond dans le repère de la façade F_i à $(0, 1, 0, 0)$ et p_i correspond à $(1, 0, 0, 0)$.

On a donc :

$$p_i = KR_{F_iC_1} \quad [4]$$

$$p_v = KR_{F_iC_2} \quad [5]$$

où $R_{F_iC_j}$ est la $j^{\text{ème}}$ colonne de la matrice R_{F_iC}. Sous forme d'égalité euclidienne, il existe μ_i et μ_v réels non nuls tels que

$$\mu_i K^{-1} p_i = R_{F_iC_1} \quad [6]$$

$$\mu_v K^{-1} p_v = R_{F_iC_2} \quad [7]$$

avec les contraintes $\|R_{F_iC_1}\| = 1$ et $\|R_{F_iC_2}\| = 1$ (où $\|.\|$ représente la norme euclidienne dans l'espace). Cela permet de déterminer μ_i et μ_v au signe près. L'ambiguïté sur les signes est levée par des hypothèses supplémentaires sur la position de la caméra. On suppose que celle-ci est tenue à la main dans une position naturelle ce qui implique que $Y_c.Y_f < 0$. Cela amène à choisir μ_v tel que $R_{F_iC22} < 0$.

Le choix du signe des μ_i est effectué en fonction de la position d'une façade F_i par rapport à la droite $l_{i\infty} = p_v \wedge p_i$. Il s'agit de la droite verticale à l'infini associée à la façade F_i. Lorsque p_i est situé à l'infini, on fait l'hypothèse que la caméra est tenue dans une position naturelle avec $X_c.X_f > 0$, ce qui implique que $R_{F_iC_{11}} > 0$ et le signe de μ_i est alors celui de p_{ix}. Lorsque p_i n'est pas situé à l'infini, si F_i est située à gauche de $l_{i\infty}$ alors $\mu_i > 0$ et si F_i est située à droite de $l_{i\infty}$ alors $\mu_i < 0$. Ici F_i est située à gauche de p_i signifie que F_i est située dans la partie de l'image située à

gauche de $l_{i\infty}$ et F_i est située à droite de p_i signifie que F_i est située dans la partie de l'image située à droite de $l_{i\infty}$. Cette définition a toujours du sens lorsque l'on fait l'hypothèse $Y_c.Y_f < 0$.

On peut donc déterminer la matrice de rotation R_{F_iC} à partir de deux points de fuite p_i et p_v sous la forme :

$$R_{F_iC} = [R_{F_iC_1} \mid R_{F_iC_2} \mid R_{F_iC_1} \wedge R_{F_iC_2}] \quad [8]$$

4.4. *Calcul d'angle entre deux façades*

Lorsque l'on détecte deux façades F_1 et F_2 associées à deux points de fuite horizontaux p_1 et p_2, on obtient deux matrices de rotation R_{F_1C} et R_{F_2C}. La matrice de rotation entre le repère lié à R_{F_1C} et le repère lié à R_{F_1C} est alors

$$R_{F_1F_2} = R_{F_2C}^{-1}.R_{F_1C} \quad [9]$$

$R_{F_1F_2}$ est une matrice de rotation autour d'un axe vertical et donc elle doit avoir la forme suivante :

$$R_{F_1F_2} = \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix} \quad [10]$$

L'angle α entre deux façades F_1 et F_2, F_1 étant la façade de gauche, est alors défini comme l'angle entre X_{f2} et $-X_{f1}$, autour d'un axe vertical orienté vers le haut. On a alors :

$$\cos\alpha = -\mu_2 K^{-1} p_2 \cdot \mu_1 K^{-1} p_1 \quad [11]$$

On remarque que α est indépendant de p_v. Il reste une ambiguïté sur le signe de α, qui est résolue de la manière suivante. Soit $signe$ la fonction de R dans $\{-1, 1\}$ qui à x associe 1 si $x > 0$ et -1 si $x < 0$ et 0 si $x = 0$. Soit s défini par :

– lorsque p_1 et p_2 sont des points finis : $s = signe(\mu_1)signe(\mu_2)signe(p_{1x} - p_{2x})$,

– lorsque p_1 est à l'infini : $s = signe(\mu_1)signe(\mu_2)signe(p_{1x})$,

– lorsque p_2 est à l'infini : $s = signe(\mu_1)signe(\mu_2)signe(-p_{2x})$,

– lorsque p_1 et p_2 sont à l'infini : $s = 0$.

Si $s = 0$ on a $\alpha = \arccos(\cos\alpha)$ et si $s \neq 0$ on a $\alpha = s.\arccos(\cos\alpha)$.

4.5. *Calcul du rapport de longueur entre deux façades*

On considère deux façades F_1 et F_2 détectées dans l'image, et on fait également l'hypothèse que ces deux façades sont voisines, c'est-à-dire qu'elles sont séparées par une arête verticale commune dans la scène réelle. F_1 est la façade apparaissant à gauche sur l'image et F_2 la façade apparaissant à droite sur l'image. Le rapport des longueurs r entre les façades F_1 et F_2 est défini par : $r =$ largeur réelle de F_1 / largeur réelle de F_2.

Soit l_1 la droite verticale dans l'image limitant la façade F_1 à gauche, l_2 la droite verticale sur l'image séparant les façades F_1 et F_2, et soit l_3 la droite verticale dans l'image limitant la façade F_2 à droite.

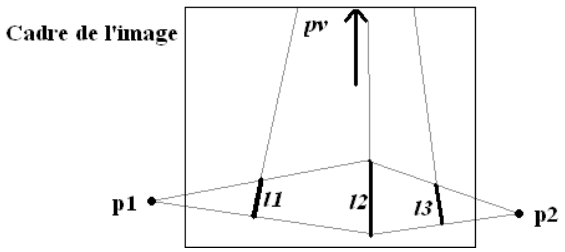


Figure 2 – Calcul de rapports de longueur.

En se plaçant dans un cadre de géométrie projective tel que présenté dans (Hartley *et al.*, 2004), on trouve :

$$r = rs\frac{d_1}{d_2} \quad [12]$$

$$d_1 = \frac{l_{2z}}{[KR_{F_1C_1}]^t.l_2} - \frac{l_{1z}}{[KR_{F_1C_1}]^t.l_1} \quad [13]$$

$$d_2 = \frac{l_{3z}}{[KR_{F_2C_1}]^t.l_3} - \frac{l_{2z}}{[KR_{F_2C_1}]^t.l_2} \quad [14]$$

$$rs = \frac{l_{2\infty z}p_1^t.l_2}{l_{1\infty z}p_2^t.l_2} \quad [15]$$

4.6. *Calcul de pose*

Connaissant des correspondances entre points du repère lié à la façade F et points dans l'image, et si les points de fuite horizontaux et verticaux de la façade sont connus, il est possible de calculer une pose. Plus précisément, il suffit de connaitre 2 correspondances entre le repère façade et l'image pour calculer la position de la caméra dans le repère façade. Chaque correspondance se fait entre une position horizontale M_i dans le repère façade et une droite verticale l_i (c'est-à-dire passant par le point de fuite vertical) qui correspond dans l'image. Tout d'abord, on sait calculer la matrice de rotation R_{FC} entre le repère façade et le repère caméra. Ensuite, on construit deux points m_1 et m_2 dans l'image de la manière suivante. Soit h une droite de l'image passant par le point de fuite horizontal mais pas par le point de fuite vertical. m_i est alors l'intersection de h et l_i. Les points M_i appartiennent à un même plan horizontal dans le repère façade, donc on sait qu'il existe Z tel que $\forall i \in \{1,2\}$ m_i corresponde à $M_i = (M_{ixf}, M_{iyf}, Z, 1)$ dans le repère façade. On va alors résoudre le système suivant :

$$\forall i \in \{1,2\} \mu_1 m_i = K[R_{FC} OfCamera] \begin{bmatrix} M_{ixf} \\ M_{iyf} \\ Z \\ 1 \end{bmatrix} \quad [16]$$

Il s'agit d'un système de 6 équations à 6 inconnues. $O_{fCamera}$ est le vecteur inconnu de l'origine du repère lié à la façade dans le repère lié à la caméra. Mais on montre que les équations ne sont pas linéairement indépendantes et donc il n'existe pas de solution unique. On adopte donc la convention $Z = 0$, qui revient à choisir l'origine des altitudes au niveau de la droite h sur l'image. Le système ci-dessus peut se reformuler sous la forme $A.OfCamera = B$ avec A une matrice connue de taille 4×3 et B un vecteur connu de R^4. Ce système linéaire surcontraint est généralement résolu par une méthode de type moindre carrés. La position de l'origine du repère caméra dans le repère façade est $O_{cfacade}$ et est liée à $O_{fCamera}$ par la relation :

$$O_{cfacade} = -R_{FC}^t . O_{fCamera} \quad [17]$$

Le calcul de la position de la caméra dans le repère SIG demande de connaitre 2 correspondances entre le repère SIG et le repère façade. L'altitude de la caméra dans le repère SIG n'est pas forcément recherchée, dans la mesure où l'on travaille avec un jeu de données bidimentionelles. Une estimation de cette altitude peut être éventuellement obtenue en utilisant un modèle numérique de terrain avec l'hypothèse que la caméra est tenue à une certaine hauteur au-dessus du sol.

Le calcul des angles de rotation entre les repères caméra et SIG se fait à partir de la matrice de rotation R_{FC} et également de 2 correspondances entre le repère SIG et

le repère façade. En particulier, il est intéressant de connaître l'azimut de la caméra. Dans le repère façade, cet azimut est défini ainsi : $\theta_{CF} = angle(X_f, Z'_c)$ avec Z'_c la projection de Z_c sur un plan (X_f, Y_f).

$$\theta_{CF} = \frac{\arccos(R_{FC_{31}})}{\sqrt{R^2_{FC_{31}} + R^2_{FC_{33}}}} \quad [18]$$

Dans le repère SIG, l'azimut de la caméra est l'angle θ_{CO} entre X_0 et Z'_c. En pratique on peut le calculer en passant par le repère de la façade de référence F, avec $\theta_{CO} = angle(X_o, X_f) + \theta_{CF}$.

4.7. *Orientation de façades*

L'orientation θ_{FO} d'une façade F dans le repère lié au SIG est définie par $\theta_{FO} = angle(X_o, X_f)$ et par conséquent nécessite de connaitre la matrice de rotation entre le SIG et la caméra. Si on ne connait pas cette matrice de rotation, on peut se contenter de calculer l'orientation θ_{FC} de la façade dans le repère caméra, définie par $\theta_{FC} = angle(Z'_c, X_f)$. On peut l'exprimer en fonction de la matrice de rotation R_{FC} car on montre que $\theta_{FC} = -\theta_{CF}$

Connaissant l'orientation θ_{CO} de la camera dans le SIG, on peut ensuite calculer l'orientation de la façade dans le SIG avec $\theta_{FO} = \theta_{FC} + \theta_{CO}$

4.8. *Azimut d'arêtes verticales*

Soit l une arête verticale dans l'image, soit M le point correspondant dans le SIG et soit P_0 la projection du centre de prise de vue sur le plan (M, X_o, Y_o). L'azimut θ_l de l est défini comme l'angle entre Z'_c et P_0M. Il s'agit donc d'un angle de rotation autour d'un axe vertical orienté vers le haut. De manière équivalente, on peut définir l'azimut d'un point q de l'image comme étant l'azimut de la droite verticale $l = q \wedge p_v$.

On peut calculer un azimut à partir d'une image : pour cela on a besoin de connaitre la droite horizontale sur l'image qui correspond au plan horizontal contenant M dans la scène réelle. Cela demande de connaitre par exemple deux points de fuite horizontaux distincts p_1 et p_2 et la ligne d'horizon est alors $l_H = p_1 \wedge p_2$. On a aussi besoin de connaitre la position du point de fuite vertical p_v. En effet on va construire la droite verticale l_{Vcam} passant le point principal $l_{Vcam} = (u_0, v_0, 1) \wedge p_v$. Puis on construit les intersections $p = l_H \wedge l$ entre l_H et l, et $p_{cam} = l_H \wedge l_{Vcam}$ entre l_H et l_{Vcam}. On pose $p = (x_p, y_p, 1)$ et $p_{cam} = (x_{p_{cam}}, y_{p_{cam}}, 1)$ puis on calcule

l'angle entre les vecteurs v_1 et v_2 définis par $v_1 = (x_{p_{cam}} - u_0, y_{p_{cam}} - v_0, f)$ et $v_2 = (x_p - u_0, y_p - v_0, f)$.

$$\theta_l = signe(x_p - x_{p_{cam}}) \frac{v_1.v_2}{\|v_1\|\|v_2\|} \quad [19]$$

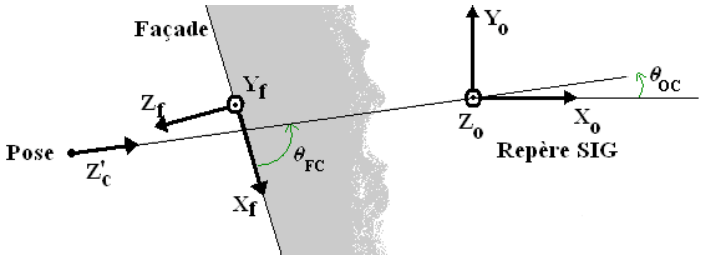


Figure 3 – Définition des angles utilisés.

5. Calcul d'incertitudes

En pratique, on ne peut pas négliger les erreurs de mesure dans les images, et il faut donc estimer l'effet de ces erreurs sur les calculs présentés précédemment. L'erreur dans la mesure d'un point sur l'image peut être modélisée par une distribution gaussienne, les différentes mesures étant considérées comme indépendantes les unes des autres (Criminisi *et al.*, 1997). Par contre on considère que les paramètres internes de calibration sont parfaitement connus. En pratique, en effet, il est possible de calibrer la caméra précisément en laboratoire, puis sur le terrain de garder les mêmes valeurs pour les paramètres internes de la caméra que sont u_0, v_0 et f. Par la suite, on notera avec un ˜ les valeurs bruitées, on notera σ_v l'écart type d'une variable aléatoire v modélisée avec une distribution gaussienne, et on notera Σ_e la matrice de covariance associée à un vecteur aléatoire e.

5.1. *Matrice de rotation*

Pour chaque façade F_i, on mesure un point de fuite horizontal $\tilde{p_i}$ et vertical $\tilde{p_v}$ qui sont bruités. Lors du calcul des éléments de R_{F_iC} on obtient une matrice $\tilde{R}_{F_iC}$ qui n'est pas exactement une matrice de rotation, même si elle en est assez proche. (Zhang, 1998) propose une méthode pour déterminer une matrice de rotation $\hat{R}_{F_iC}$ qui soit la plus proche possible de $\tilde{R}_{F_iC}$.

5.2. *Modélisation des erreurs sur les grandeurs mesurées*

On va modéliser α, r, θ_{FC} et θ_q en fonction des erreurs commises en entrée. On utilise alors le résultat suivant, dérivé de (Hartley *et al.*, 2004) (page 140). Pour toute fonction f_s de R^n dans R avec n>0, on note sa matrice jacobienne J_s, définie en tout point de R^n sur lequel f_s est dérivable. Soit $e = (e_1, \ldots, e_n)$ tel que f est dérivable en $\bar{e}$. Alors $s = f_s(e)$ est un vecteur aléatoire et on peut obtenir une valeur approchée au premier ordre de sa matrice de covariance par la formule :

$$\Sigma_s = J_s \Sigma_e J_s^t \qquad [20]$$

Cette approximation devient exacte lorsque f_s est linéaire. Les fonctions f_α, f_r, $f_{\theta_{FC}}$ et f_{θ_q} servant à calculer α, r, θ_{FC} et θ_q ne sont pas linéaires. Cependant, sous réserve d'une distribution homogène des erreurs de mesure dans l'image, nos expériences montrent que l'approximation réalisée est bonne et que les variables α, r, θ_{FC} et θ_q peuvent être modélisées comme des gaussiennes. De plus, les fonctions f_α, f_r, $f_{\theta_{FC}}$ et f_{θ_q} sont généralement dérivables dans des conditions normales de prises de vue : notamment les points de fuite considérés doivent être suffisamment éloignés du point principal.

5.3. *Diminution des incertitudes dans le cas du calcul du rapport de longueurs*

En pratique, le calcul du rapport de longueur entre deux façades ne part pas de la connaissance de 3 lignes verticales l_1, l_2 et l_3 mais de points appartenant à ces lignes verticales et du point de fuite vertical. Cela joue dans le calcule d'incertitude. Soient q_1, q_2 et q_3 trois points de l'image appartenant respectivement à l_1, l_2 et l_3 et distincts de p_v : alors $l_1 = q_1 \wedge p_v$, $l_2 = q_2 \wedge p_v$ et $l_3 = q_3 \wedge p_v$. Notons également que des simplifications sérieuses surviennent dans les formules dans le cas où l'angle entre les deux façades est de 180 degrés, c'est-à-dire quand $p_1 = p_2$.

5.4. *Diminution des incertitudes dans le cas du calcul de l'orientation d'une façade*

On remarque que dans ce calcul d'orientation, le résultat trouvé dépend de la position du point de fuite vertical. Une autre alternative pour calculer l'orientation d'une façade F est d'utiliser une façade de référence F_{ref} pour laquelle on connait l'orientation dans le repère SIG ainsi que le point de fuite horizontal associé dans l'image, puis de calculer les angles entre F_{ref} et F. Cette méthode entraine des incertitudes de calcul moins grandes car elle ne demande pas de connaitre le point de fuite vertical.

6. Formulation des requêtes dans le SIG

6.1. *Présentation des données à disposition*

Le SIG que nous utilisons couvre un quartier de la ville de Nantes, et les données peuvent être visualisées sur `http://www.geonantes.fr`. Ces données sont celles utilisées par la communauté urbaine de Nantes, et nous estimons qu'elles ont une précision submétrique. En plus de la couche bâtiments, de type multipolygones et qui représente l'emprise au sol des bâtiments, nous avons à disposition une couche route de type multilignes qui représente l'axe des voies routières. Nous avons simplifié la couche bâtiments de façon à supprimer toutes les arêtes réputées non visibles depuis une rue. Pour cela, nous calculons l'union de tous les polygones ayant une intersection de dimension non nulle, et nous supprimons les trous des polygones résultants. Cette opération de fusion de modifie pas la géométrie des façades visibles, et donc n'altère pas la qualité des données. Cependant, on peut noter que déterminer des arêtes invisibles depuis une rue uniquement à partir de données 2D revient à faire l'hypothèse que les bâtiments ont une hauteur infinie. Bien que cette hypothèse semble peu réaliste, nos expériences montrent que des situations ambiguës sont peu fréquentes.



Figure 4 – Simplification du SIG.

6.2. *Requêtes utilisant les angles et les rapports de longueur*

Les données géométriques issues des images, associées à leurs incertitudes de mesure, sont transformées en requêtes sur la couche bâtiments d'un SIG. On souhaite utiliser des tolérances assez larges dans les requêtes SIG, qui permettent de ne pas écarter les vrais résultats. Actuellement, nous avons choisi la création des tolérances D_α, D_r à partir des écarts types σ_α, σ_r en multipliant par un facteur k. Dans nos expériences nous avons choisi $k = 2$. Le principe des requêtes est de trouver les couples de segments (s_1, s_2) dans le SIG tels que :

– s_1 et s_2 sont deux segments voisins, c'est-à-dire qu'ils partagent une extrémité en commun,

– s_1 correspond à la façade de gauche et s_2 correspond à la façade de droite, c'est-à-dire que pour des polygones dont les points sont enregistrés dans le sens horaire, s_1 est situé après s_2,

– l'angle entre s_2 et s_1 situé à l'intérieur du polygone est compris dans $[\alpha - D_\alpha; \alpha + D_\alpha]$,

– le rapport longueur(s_1)/longueur(s_2) est compris dans $[r - D_r; r + D_r]$.

De manière plus générale, on peut également rechercher des n-uplets de segments $(s_1, \ldots, s_n)$ où chaque couple de segments (s_i, s_{i+1}) vérifie les 4 contraintes mentionnées ci-dessus.

Une fois tous les n-uplets de segments sélectionnés, on sait associer chaque segment de chaque n-uplet à une façade sur l'image. On sélectionne alors la façade par rapport à laquelle on va calculer la pose, de façon à minimiser l'incertitude dans le calcul de pose. Cette façade de référence F_{ref} peut être celle telle que l'angle entre X_{fref} et Z'_c est le plus proche possible de 90°, et correspond donc dans le SIG au segment s_{ref}. On dispose alors de deux correspondances entre des arêtes verticales sur l'image limitant F_{ref}, et des points SIG limitant s_{Ref}. Le calcul de la position horizontale s'effectue ensuite par résolution d'un système d'équations linéaires surcontraint et le calcul de l'orientation de la caméra s'effectue à partir de la matrice de rotation $R_{F_{Ref}C}$, tel qu'expliqué au paragraphe 4.6. Les deux autres angles d'orientation de la caméra sont calculés directement à partir de $R_{F_{Ref}C}$.

On obtient en résultat de cette première requête l'ensemble des poses (positions horizontales et orientations dans l'espace) possibles pour la caméra, sachant que cet ensemble contient la vraie pose. Il existe par ailleurs une incertitude sur la position et l'orientation de la caméra, qu'il est difficile d'évaluer car elle dépend de l'exactitude des données SIG.

6.3. *Requêtes de cohérence*

Pour chaque pose calculée, on peut vérifier si elle est cohérente avec les informations du SIG qu'on n'a pas encore utilisées. Premièrement, on peut estimer d'après le SIG s'il existe une route entre la pose et les façades reconnues. Si une route est reconnue sur l'image alors que le SIG met en évidence l'absence de route pour une pose, alors cette pose est déclarée incorrecte. Ensuite, on peut vérifier une contrainte de non-obstruction, qui veut que pour toute pose, il n'existe pas de bâtiment situé entre la pose et les façades visualisées. On exclut ainsi, au passage, toutes les poses situées dans un bâtiment. On utilise une tolérance lors de ces requêtes, afin de minimiser l'effet des erreurs dans le calcul de pose. Pour cela, la pose calculée est remplacée temporairement par une pose secondaire, légèrement plus proche des façades.

À partir de chaque pose, on peut ensuite calculer un polygone de visibilité. Ce polygone est défini comme l'ensemble des points du SIG visibles depuis la pose, dans un domaine angulaire donné et avec une portée donnée. Pour nos expériences, nous avons considéré une portée de 100 m, étant donné que la résolution des images rend difficile la reconnaissance de bâtiments très éloignés de la pose. Ces polygones de visibilité nous donnent des informations sur les orientations des façades et les azimuts des arêtes verticales visibles. On peut alors comparer ces informations extraites du SIG avec celles extraites de l'image, tel que mentionné dans les paragraphes 4.7 et 4.8, et si une incohérence apparait manifestement, la pose est déclarée incorrecte. Cette requête utilise également des tolérances sur les azimuts d'arêtes verticales et les orientations, lesquelles doivent tenir compte des incertitudes de mesures dans l'image et dans le calcul de pose.

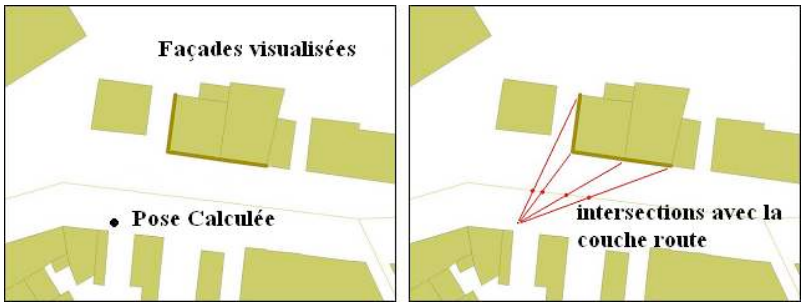


Figure 5 – Resultat de la requête sur les segments et du calcul de pose (à gauche) et principe de la requête sur la couche route (à droite). On vérifie qu'il y a bien une route qui passe entre les façades et la pose calculée



Figure 6 – Illustration de la requête de non-obstruction (à gauche) et du calcul des polygones de visibilité (à droite). Il s'agit dans les deux cas de déterminer un polygone n'ayant pas d'intersection avec les polygones du SIG

6.4. *Fusion de plusieurs poses*

Les requêtes de cohérence permettent généralement de réduire de manière importante le nombre de poses possibles, mais elles ne permettent pas systématiquement de le réduire à 1. On peut alors utiliser les informations de cohérence entre des vues séparées, qui permettent d'augmenter la quantité d'informations extraites sur la géométrie de la scène. S'il est possible de travailler avec plusieurs images orientées vers une même scène afin de pouvoir travailler dans un cadre de géométrie épipolaire, nous avons ici travaillé, au contraire, avec des images orientées dans plusieurs directions différentes. On peut ensuite calculer une couche de pose différente pour chaque image. On fait pour cela l'hypothèse que les différents centres de prise vue sont peu distants les uns des autres, c'est-à-dire compris dans un cercle C avec un rayon r de quelques mètres. On peut alors supprimer des couches de poses toutes les poses qui n'entrent pas dans un tel voisinage.

7. Expériences

Notre outil de géolocalisation a été programmé en java, basé sur la librairie GDMS (Leduc *et al.*, 2009). Cette librairie est une des composantes de base du logiciel OrbisGIS, et permet notamment un accès générique à différentes sources de données géographiques vectorielles.

7.1. *Préparation du SIG*

La zone urbaine choisie comprend 7138 bâtiments, ce qui correspond à une zone de 1 km de rayon centrée sur un point. Cela découle d'une hypothèse que nous faisons, que l'utilisateur sait dans quel quartier il se trouve, soit d'après ses connaissances, soit par utilisation d'un récepteur GPS peu exact. Suite à la simplification dans la couche bâtiments décrite au paragraphe 6.1, nous obtenons une couche comprenant 2715 polygones, avec une baisse significative du nombre de segments à explorer dans le SIG.

7.2. *Préparation des images*

32 photographies ont été acquises sur 9 sites différents. La distance focale a été fixée pendant la campagne d'acquisition, et ainsi tous les paramètres internes de la caméra sont considérés comme connus. Sur chaque site, les photographies ont été acquises avec des centres de prises de vue séparés de moins de 20 m. L'analyse des images peut être réalisée de manière semi automatique : d'une part la détection des points de fuite peut se faire automatiquement (Kalantari *et al.*, 2008; Kosecka *et al.*, 2002) et d'autre part la détection des arêtes verticales et les relations de voisinage entre les façades peuvent être établies rapidement par l'opérateur.

Par ailleurs nous avons écarté de nos expériences les cas dans lequel le SIG n'était pas à jour. En effet, sur 4 des images, un bâtiment récent a été construit et ne figure pas dans le SIG, empêchant la mise en correspondance correcte du SIG et des images. La figure 7 montre un exemple de photographie pour lequel le SIG n'est manifestement pas à jour. La base d'images utilisée pour les tests contient donc 28 images seulement.



Figure 7 – Situation dans laquelle le SIG n'est pas à jour. Le SIG (à droite) ne correspond plus à la photographie (à gauche) en raison d'un bâtiment qui a été construit récemment.

7.3. *Résultats de la géolocalisation par images*

8 images sur 28 ont été localisées directement avec succès. L'exactitude dans le calcul de la position de la caméra est de l'ordre du mètre.

Pour 7 images le nombre de poses possibles n'a pas pu être ramené à 1.

Pour les 13 images restantes, la méthode a conduit à des échecs, parmi lesquels on a les suivants :

– 4 images n'ont pas pu être localisées à cause d'une mauvaise modélisation dans le SIG. Plus précisément, on a observé que 3 échecs sont dus à une mauvaise modélisation des angles entre façades et qu'un échec est causé par une structure 3D n'appartenant pas au bâti mais modélisée dans le SIG. La figure 8 présente un exemple de mauvaise modélisation des angles dans le SIG.

– 7 erreurs sont dues à des situations ambiguës dans laquelle il faut utiliser une approche multihypothèse afin de retenir la correspondance correcte entre le SIG et l'image. Il s'agit de requêtes portant sur des façades séparées par un angle de 180 degrés. Un exemple de ce genre de situation est présenté sur la figure 9. Cet figure illustre le fait que pour une arête verticale identifiée dans l'image, on ne peut pas savoir s'il existe une arête verticale correspondante dans le SIG, et réciproquement pour un point du SIG, il n'est pas forcément possible de détecter une arête verticale correspondante dans l'image.

– 1 erreur est due à un nombre trop faible d'informations dans l'image (impossible de voir une façade en entier).

Dans la plupart des cas heureusement, on peut contourner le problème, comme expliqué dans la partie suivante.

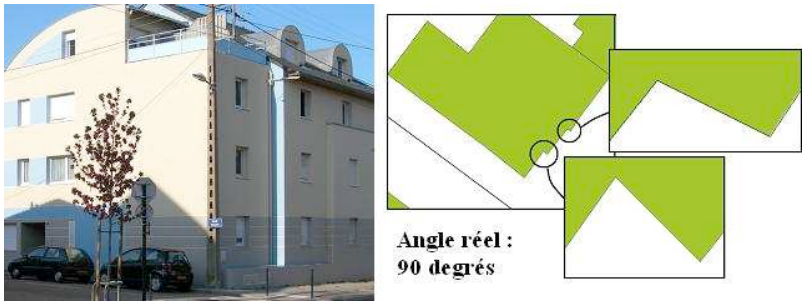


Figure 8 – Situation dans laquelle le SIG comporte des anomalies de modélisation. Deux angles du SIG (à droite) ne correspondent pas à la réalité, tel que visible sur la photographie (à gauche). Les angles sont censés être très voisins de 90 degrés.

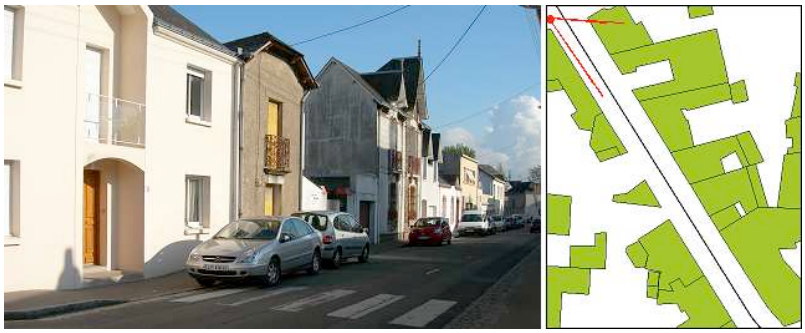


Figure 9 – Situation dans laquelle il est difficile de faire le lien entre le SIG et l'image. Les façades sont toutes alignées et il est difficile de distinguer sur l'image les arêtes verticales limitant deux façades distinctes.

8. Discussion

8.1. *Analyse des erreurs*

Parmi les causes d'échec dans la localisation d'une image, certaines peuvent être corrigées. Par exemple les structures mal modélisées dans le SIG sont souvent des arêtes plus courtes que la moyenne. Une étude plus fine de la manière dont le SIG a été

construit permet de prédire en partie de telles erreurs de modélisation. De nombreuses erreurs sont liées à la reconnaissance de façades séparées par des angles de 180°. Il n'est alors possible de séparer deux façades que par une analyse de la texture, car aucun indice géométrique ne permet de le faire. Ce type de requête est ambigu par nature, car la correspondance entre le SIG et l'image n'est pas définie, et cela résulte en de nombreux échecs. En gardant plusieurs arêtes verticales possibles sur l'image, et la possibilité de ne pas prendre en compte certains points du SIG, on génère plus de poses possibles mais la probabilité d'écarter la pose correcte diminue.

8.2. *Utilisation des tolérances*

Comme on ne souhaite pas éliminer la pose correcte, on doit utiliser des tolérances dans toutes les requêtes. En particulier, les requêtes de cohérence sont sensibles à l'erreur dans le calcul de pose, laquelle dépend de l'erreur dans la mesure sur l'image mais aussi des erreurs dans les points du SIG. Or, obtenir un modèle de précision fin associé au SIG demande une bonne connaissance de la manière dont ont été produites les données de base, et nous l'envisageons dans nos travaux futurs. On se contente actuellement d'un modèle empirique grossier avec des tolérances souvent surévaluées dans les requêtes de cohérence, permettant de tenir en compte les éventuelles erreurs dans les données SIG. Les erreurs de calcul de pose sont particulièrement critiques au moment où on calcule les polygones de visibilité, lesquels peuvent changer d'allure si des façades proches de la pose, mais non visibles sur l'image, entrent dans le champ de vision calculé. Augmenter les tolérances utilisées dans les requêtes a pour effet de générer moins d'échecs, mais plus d'hypothèses possibles. Le nombre de poses possibles générées par les requêtes varie également beaucoup en fonction de l'environnement. Il dépend du nombre de bâtiments considérés dans la zone de recherche initiale. Il dépend aussi du tissu urbain, et par exemple de l'orientation relative des différentes façades. Les angles atypiques entre les façades offrent des indices précieux pour réduire significativement le nombre de poses possibles. Nous proposons une discussion plus approfondie sur ce point dans (Bioret *et al.*, 2008).

8.3. *Utilisation de capteurs externes*

L'utilisation d'un GPS permet de réduire très fortement la zone de recherche initiale. On peut faire l'hypothèse qu'avec un GPS de qualité moyenne on peut réduire la zone de recherche à quelques dizaines de mètres carrés, sauf en centre urbain très dense où les signaux GPS sont rarement accessibles. On peut imaginer que dans certains cas, une indication donnée par l'utilisateur peut être également intéressante : par exemple le nom d'une rue ou un élément remarquable du paysage. Un capteur d'orientation absolue permet, sur des tissus urbains globalement isotropes, de réduire le nombre de poses possibles, même si ce capteur est entaché d'une erreur de mesure d'une dizaine de degrés. On peut imaginer plusieurs capteurs tels que des inclino-

mètres, qui permettent de faciliter le calcul des matrices de rotations entre façades et caméra.

8.4. *Utilisation de données 3D*

L'utilisation d'une couche bâtiments 3D dans le SIG au lieu d'une couche bâtiments 2D permettrait de réduire très fortement le nombre de poses possibles, comme nous l'illustrons sur un exemple dans (Bioret *et al.*, 2009). On pourrait ainsi reformuler les requêtes présentées dans le paragraphe 6.2 en utilisant également les rapports hauteur/largeur calculés à partir de l'image.

Toutefois, les SIG 3D à l'échelle d'une ville entière sont actuellement difficilement disponibles. De plus, la modélisation 3D obtenue au final pour les bâtiments n'est pas toujours fiable, et se résume souvent à une hauteur moyenne sur le bâtiment. En effet, la reconstruction de modèles 3D urbains est encore un domaine de recherche à part entière (Lafarge, 2007; Müller *et al.*, 2007). Certains travaux visent une couverture large et se basent sur des données LIDAR ou des reconstructions manuelles assistées par ordinateur (Evans *et al.*, 2005), d'autres proposent des reconstructions plus locales mais plus précises à partir de photographies terrestres (Cornou *et al.*, 2003).

Enfin, la détection des structures 3D dans l'image est délicate, et par exemple il n'est pas toujours possible de définir la hauteur d'une façade dans une image. L'appariement entre SIG 3D et images demande donc un effort de recherche sérieux.

9. Conclusion

Nous avons présenté une approche de géolocalisation à partir d'images, en milieu urbain, basée sur l'utilisation d'un Système d'Information Géographique. Nous avons tout d'abord présenté comment les angles, rapports de longueur entre façades, orientations de façades et azimuts d'arêtes verticales peuvent être extraits des images. Nous avons ensuite présenté le calcul d'incertitude sur ces éléments géométriques et leur transformation en requêtes dans un SIG. Cette méthode de localisation permettra la production de modèles urbains plus fins, et en particulier des modèles de bâtiments texturés, ces données pouvant être directement intégrées dans un SIG. Elle permettra avant tout de présenter les informations géographiques 3D avec une interface de Réalité Augmentée, si nécessaire sur site. Une application possible en est la visualisation de projets d'aménagements urbains dans un environment réel.

Nos résultats expérimentaux montrent la pertinence de l'approche dans certains cas, mais également ses limites actuelles. La correspondance entre images et SIG est en effet ambiguë, mais peut être résolue par l'utilisation de plusieurs images au lieu d'une seule, apportant une information plus redondante. Nos travaux futurs comprennent l'identification et la résolution de ces ambiguïtés dans les correspondances entre SIG 2D et image, ainsi que la mise en correspondance d'images et de jeux de

données tridimentionnelles. L'automatisation dans la reconnaissance des façades sur l'image est également envisagée pour le futur.

10. Bibliographie

Bartczak B., Koeser K., Woelk F., Koch R., « Extraction of 3D freeform surfaces as visual landmarks for real-time tracking », *Journal of Real-Time Image Processing*, vol. 2, n° 2-3, p. 81-101, November, 2007.

Behringer R., Park J., Sundareswaran V., « Model-Based Visual Tracking for Outdoor Augmented Reality Applications », *ISMAR, Darmstadt, Germany, September 30-October 1*, p. 277-278, 2002.

Bioret N., Moreau G., Servières M., « Localization Approach Based on Image/GIS Correspondence Using a Simple 2D Building Layer », *REGARD, Québec, Canada, August 28-29*, 2008.

Bioret N., Moreau G., Servières M., « Pose computation based on Correspondence between 2D GIS and Street Images », *CISA, Zarzis, Tunisia, March 23-25*, 2009.

Cappelle C., Najjar M. E. B. E., Pomorski D., Charpillet F., « Localisation in urban environment using GPS and INS aided by monocular vision system and 3D geographical model », *IEEE Intelligent Vehicles Symposium, Istanbul, Turkey, June 13-15*, p. 811 - 816, 2002.

Coors V., Huch T., Kretschmer U., « Matching buildings : Pose estimation in an rban environment », *IEEE and ACM International Symposium on Augmented Reality, Munich, Germany, October 5-6*, p. 89-92, 2000.

Cornou S., Dhome M., Sayd P., « Architectural reconstruction with multiple views and geometric constraints », *BMVC, Norwich, UK, September 9-11*, 2003.

Criminisi A., Reid I., Zisserman A., « A plane Measuring Device », *Image and Vision Computing*, vol. 17, n° 8, p. 625-634, September, 1997.

David P., DeMenthon D., Duraiswami R., Samet H., « SoftPOSIT : Simultaneous Pose and Correspondence Determination », *IJCV*, vol. 59, n° 3, p. 259-284, September, 2004.

Evans S., Hudson-Smith A., Batty M., « 3-D GIS : Virtual London and beyond », *SAGEO, Avignon, France, June 20-23*, 2005.

Fritz G., Seifert C., Paletta L., « Urban Object Recognition from Informative Local Features », *International Conference on Robotics and Automation, Barcelona, Spain, April 18-22*, p. 131-137, 2005.

Goedeme T., Tuytelaars T., Gool L. V., « Fast Wide Baseline Matching for Visual Navigation », *CVPR, Washington, DC, USA, June 27 - July 2*, p. 24-29, 2004.

Haala N., Böhm J., « A multi-sensor system for positioning in urban environments », *ISPRS Journal of Photogrammetry & Remote Sensing*, vol. 58, p. 31-42, January, 2003.

Hartley R., Zisserman A., *Multiple View Geometry In Computer Vision*, Cambridge University Press, 2004.

Jiang B., Neumann U., You S., « A Robust Hybrid Tracking System for Outdoor Augmented Reality », *IEEE Virtual Reality, Chicago, Illinois, USA, March 27-31*, p. 3-10, 2004.

Johansson B., Cipolla R., « A System for Automatic Pose-Estimation from a Single Image in a City Scene », *Signal Processing, Pattern Recognition, and Applications, Crete, Greece, June 25-28*, 2002.

Kalantari M., Jung F., Guédon J.-P., Paparoditis. N., « Détection automatique des points de fuite et calcul de leur incertitude à l'aide de la géométrie projective », *RFIA, Amiens, France, January 22-25*, 2008.

Kosecka J., Zhang W., « Video Compass », *7th European Conference on Computer Vision, Part IV, Copenhagen, Denmark, May 28-31*, p. 476-490, 2002.

Lafarge F., Modèles Stochastiques pour la Reconstruction Tridimensionnelle d'Environnements Urbains, PhD thesis, Ecole des Mines de Paris, 2007.

Leduc T., Bocher E., Gonzaléz-Cortés F., Moreau G., « GDMS-R : A mixed SQL to manage raster and vector data », *GIS 2009, Ostrava, Czech Republic, January 25-58*, 2009.

Lee J. W., You S., Neumann U., « Tracking with Omni-directional Vision for Outdoor AR Systems », *ISMAR, Darmstadt, Germany, September 30-October 1*, p. 47-58, 2002.

Moslah O., Guitteny V., Couvet S., « Fitting Uncalibrated Photographs to Geo-Referenced Urban 3D Models », *Journées de l'AFRV, Bordeaux, France, 30-31 october*, p. 57-61, 2008.

Müller P., Zeng G., Wonka P., Gool L. V., « Image-based Procedural Modeling of Facades », *Proceedings of ACM SIGGRAPH 2007 / ACM Transactions on Graphics*, vol. 26, n° 3, p. 85-93, July, 2007.

Reitmayr G., Drummond T. W., « Going out : Robust Model-based Tracking for Outdoor Augmented Reality », *ISMAR, Santa Barbara, California, USA, October 22-25*, p. 109-118, 2006.

Reitmayr G., Drummond T. W., « Initialisation for Visual Tracking in Urban Environments », *ISMAR, Nara, Japan, November 13-16*, p. 1-9, 2007.

Ribo M., Lang P., Ganster H., Brandner M., Stock C., Pinz A., « Hybrid Tracking for Outdoor Augmented Reality Applications », *IEEE Computer Graphics and Applications*, vol. 22, n° 6, p. 54-63, November, 2002.

Robertson D., Cipolla R., « An Image-Based System for Urban Navigation », *BMVC, London, UK, September 7-9*, 2004.

Robertson D. P., Cipolla R., « Building Architectural Models from Many Views Using Map Constraints », *7th European Conference on Computer Vision, Part II, Copenhagen, Denmark, May 28-31*, 2002.

Schindler K., Bauer J., « Towards Feature-Based Building Reconstruction From Images », *Computer Graphics, Visualization and Computer Vision, Plzen, Czech Republic, February 3-7*, 2003.

Shao H., Svoboda T., Tuytelaars T., Gool L. V., « HPAT indexing for fast object/scene recognition based on local appearance », *International Conference on Image and Video Retrieval, University of Illinois at Urbana-Champaign, USA, July 24-25*, p. 71-80, 2003.

Snavely N., Seitz S. M., Szeliski R., « Photo tourism : exploring photo collections in 3D », *ACM Transaction on Graphics*, vol. 25, n° 3, p. 835-846, July, 2006.

Sourimant G., Morin L., Bouatouch K., « Gps, Gis and Video fusion for urban modeling », *Computer Graphics International, Petropolis, Brazil, May 30-June 2*, p. 79-88, 2007.

Steinhoff U., Omercevic D., Perko R., Schiele B., Leonardis A., « How computer Vision can Help in Outdoor Positionning », *Ambient Intelligence, Darmstadt, Germany, November 7-10*, p. 124-141, 2007.

Teller S., Antone M., Bodnar Z., Bosse M., Coorg S., Jethwa M., Master N., « Calibrated, Registered Images of an Extended Urban Area », *IEEE Computer Vision and Pattern Recognition, Kauai Marriott, Hawaii, USA, December 9-14*, p. 813-820, 2001.

Wang C.-C., Thorpe C., Hebert M., Thrun S., Durrant-Whyte H., « Simultaneous Localization, Mapping and Moving Object Tracking », *The International Journal of Robotics Research*, vol. 26, n° 6, p. 889-916, June, 2007.

Zhang W., Kosecka J., « Image Based Localization in Urban Environments », *International Symposium on 3D Data Processing, Visualization and Transmission, North Carolina, Chapel Hill, June 14-16*, p. 33-40, 2006.

Zhang Z., A Flexible new Technique for Camera Calibration, Rapport de recherche n° MSR-TR-98-71, Microsoft Research, December, 1998.

Article reçu le 18/12/2008.
Version révisée le 18/12/2008.
Rédacteur responsable : NICOLAS BIORET

SERVICE ÉDITORIAL – HERMES-LAVOISIER
14 rue de Provigny, F-94236 Cachan cedex
Tél. : 01-47-40-67-67
E-mail : revues@lavoisier.fr
Serveur web : http://www.revuesonline.com

ANNEXE POUR LE SERVICE FABRICATION
A FOURNIR PAR LES AUTEURS AVEC UN EXEMPLAIRE PAPIER
DE LEUR ARTICLE ET LE COPYRIGHT SIGNE PAR COURRIER
LE FICHIER PDF CORRESPONDANT SERA ENVOYE PAR E-MAIL

1. ARTICLE POUR LA REVUE :
ISI

2. AUTEURS :
Nicolas Bioret[*] *— Guillaume Moreau*[*] *— Myriam Servières*[*]

3. TITRE DE L'ARTICLE :
Géolocalisation en milieu urbain par appariement entre une collection d'images et un SIG 2D.

4. TITRE ABRÉGÉ POUR LE HAUT DE PAGE MOINS DE 40 SIGNES :
géolocalisation en milieu urbain

5. DATE DE CETTE VERSION :
18 mai 2009

6. COORDONNÉES DES AUTEURS :

– adresse postale :
[*] Laboratoire CERMA UMR CNRS 1563
rue Massenet, F-44319 Nantes Cedex 3
{prenom.nom}@ec-nantes.fr

– téléphone : 02 40 16 01 37
– télécopie :
– e-mail : nicolas.bioret@ec-nantes.fr

7. LOGICIEL UTILISÉ POUR LA PRÉPARATION DE CET ARTICLE :
LaTeX, avec le fichier de style `article-hermes.cls`,
version 1.23 du 17/11/2005.

8. FORMULAIRE DE COPYRIGHT :
Retourner le formulaire de copyright signé par les auteurs, téléchargé sur :
`http://www.revuesonline.com`

SERVICE ÉDITORIAL – HERMES-LAVOISIER
14 rue de Provigny, F-94236 Cachan cedex
Tél. : 01-47-40-67-67
E-mail : revues@lavoisier.fr
Serveur web : http://www.revuesonline.com