



SLAM contraint pour la localisation dans des environnements partiellement connus

Sylvie Naudet, Vincent Gay-Bellile, Steve Bourgeois, Mohamed Tamaazousti,
Dorra Larnaout, Michel Dhome

► To cite this version:

Sylvie Naudet, Vincent Gay-Bellile, Steve Bourgeois, Mohamed Tamaazousti, Dorra Larnaout, et al..
SLAM contraint pour la localisation dans des environnements partiellement connus. Reconnaissance
de Formes et Intelligence Artificielle (RFIA) 2014, Jun 2014, France. hal-00988849

HAL Id: hal-00988849

<https://hal.science/hal-00988849>

Submitted on 9 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SLAM contraint pour la localisation dans des environnements partiellement connus

S. Naudet-Collette¹ V. Gay-Bellile¹ S. Bourgeois¹ M. Tamaazousti¹ D. Larnaout¹ M. Dhome²

¹ CEA, LIST, Laboratoire Vision et Ingénierie des Contenus, PC 173, F-91191 Gif-sur-Yvette, France

² Institut Pascal, UMR 6602 Université Blaise Pascal/CNRS/IFMA

¹prénom.nom@cea.fr, ²michel.dhome@univ-bpclermont.fr

Résumé

Cet article porte sur la localisation temps réel d'une caméra mobile dans un environnement partiellement connu. Ces travaux reposent sur une solution de type SLAM monoculaire et traitent la problématique de dérive inhérente à ce type de méthode en proposant un nouvel ajustement de faisceaux, dit ajustement de faisceaux contraint. Ce dernier permet d'intégrer des contraintes géométriques apportées par le modèle partiellement connu de la scène. Nous montrons l'apport de cette nouvelle méthode sur des applications de Réalité Augmentée sur de petits et grands environnements.

Mots Clef

SLAM, Ajustement de faisceaux contraint, réalité augmentée, SIG, grands environnements.

Abstract

This article addresses the challenging issue of real-time camera localization in a partially known environment. This work is based on a monocular SLAM algorithm and deals with inherent drifts characterizing such method, by proposing a new bundle adjustment, said constrained bundle adjustment. This allows to integrate geometric constraints provided by the partially known model of the scene. We demonstrate the advantages of this new method on augmented reality applications.

Keywords

SLAM, constrained bundle adjustment, augmented reality, GIS, large scale environments.

1 Introduction

Ce papier porte sur la localisation temps réel d'une caméra mobile pour les applications de Réalité Augmentée (RA). Le concept de réalité augmentée vise à enrichir notre perception du monde réel, en y ajoutant des éléments virtuels ou en modifiant l'apparence de la scène réelle. Afin de donner l'illusion que ces objets fictifs appartiennent au même monde, il est nécessaire de les recalibrer précisément par rapport aux objets réellement perçus. Ceci repose générale-

ment sur un calcul précis et temps réel de la pose de la caméra fixée sur le système mobile. Les travaux présentés ici, portent sur deux cadres applicatifs : la RA sur des objets 3D et la RA pour le guidage en environnement urbain. Outre l'aspect temps réel et la nécessité de fournir une localisation géo-référencée, ceci implique d'adresser deux défis majeurs :

- La stabilité de la localisation 6DOF indispensable pour la qualité du rendu visuel (éviter les effets de jittering¹).
- La précision au cours du temps.

L'ensemble de ces travaux reposent sur un algorithme de localisation et cartographie simultanées, SLAM visuel monoculaire introduit par [15]. Celui-ci s'appuie sur la notion d'images clés et d'ajustement de faisceaux local (AF local) dans lequel seules les poses des N dernières images clés et les M points 3D observés sont raffinés. Ce procédé présente l'avantage de fournir une pose stable grâce à l'exploitation des équations de projections multi-vues dans l'AF, cependant il ne permet qu'une localisation relative et il est sujet aux effets de dérive au cours du temps liés à l'accumulation d'erreur. L'objectif de ces travaux a alors été de trouver des alternatives pour pallier le problème de dérive et obtenir une localisation absolue. Nous verrons que pour les deux cadres applicatifs traités, ce défi a été adressé avec le même fil conducteur : l'intégration de données absolues au cœur de l'ajustement de faisceaux du SLAM. Ces données sont issues d'un modèle 3D partiel de la scène.

2 SLAM contraint pour le suivi d'objets 3D

Les techniques de localisation d'une caméra mobile peuvent être classées en deux grandes familles :

- les techniques de localisation et cartographie simultanée SLAM[9][15]
- les techniques s'appuyant sur un modèle 3D géométrique ou photométrique (Model Based tracking)[12].

Les techniques SLAM estiment simultanément le mouvement de la caméra et l'environnement, en exploitant les relations de projection multi-vues. De ce fait, ces tech-

1. tremblement dans l'image

niques fournissent une localisation relativement stable. Cependant, elles sont sujettes à trois limitations majeures : la localisation est relative, le facteur d'échelle est arbitrairement fixé et elles souffrent de dérive au cours du temps et d'accumulation d'erreurs. Les techniques s'appuyant sur un modèle 3D exploitent une connaissance *a priori* de la géométrie et/ou de l'apparence de l'objet pour estimer la pose à partir des correspondances des primitives 3D du modèle avec celles de l'image courante. Ces solutions fournissent donc une localisation absolue et relativement précise mais sont sujettes aux effets de tremblements et sont sensibles aux occultations, essentiellement parce qu'elles n'exploitent que les informations visuelles de l'objet d'intérêt et de l'image courante (La pose est déterminée indépendamment à chaque image). Pour un suivi stable, ce processus implique que l'objet d'intérêt soit visible de manière continue et tienne une place prépondérante dans les images durant l'ensemble de la séquence. Des travaux proposent de combiner ces deux familles d'approches. Bleser et al [2] utilisent une méthode basée modèle pour initialiser un SLAM mais ne corrigent pas la dérive. Kempter et al [8] proposent d'alterner ces deux familles d'approches pour exploiter les contraintes géométriques liées au modèle lorsque l'objet est proche et les contraintes multi-vue liées à la reconstruction lorsque celui-ci est loin. Cependant cette solution se limite à utiliser ces informations de manière alternées. Castle et al [3] proposent d'intégrer les coins d'objets plans connus dans un FKE-SLAM. Bien que cette méthode ne s'applique qu'à des points connus, elle montre l'apport en précision et stabilité de l'utilisation conjointe de la géométrie du modèle et des amers entourant l'objet.

Pour répondre aux critères de robustesse, précision et stabilité, recherchés pour la RA, nous avons proposé une nouvelle approche [17][16] reposant sur l'exploitation conjointe de la connaissance *a priori* sur la géométrie de l'objet observé et la richesse en terme d'amers visuels de l'environnement autour de cet objet. Cette méthode propose une unification, au sein d'un même formalisme, de méthodes de localisation et cartographie simultanée (SLAM) et de méthodes de calcul de pose à partir d'un modèle 3D (Model Based tracking). L'objectif étant de cumuler les avantages de ces deux familles de méthodes tout en limitant leurs inconvénients respectifs. Cette nouvelle approche, que nous avons appelée SLAM contraint, est un algorithme SLAM basé images clef dans lequel l'ajustement de faisceaux intègre d'une part des contraintes de projection multi-vue issues du SLAM et d'autre part des contraintes géométriques inhérentes au modèle 3D de l'objet. Cette intégration est faite via la minimisation d'une fonction d'erreur intégrant le résidu de projections de primitives de l'environnement $\mathcal{E}_E$ et le résidu des primitives liées au modèle $\mathcal{E}_M$.

$$\mathcal{E} = \mathcal{E}_E + \mathcal{E}_M \quad (1)$$

avec $\mathcal{E}_E$ la fonction de coût d'un ajustement de faisceaux classique.

$$\mathcal{E}_E \left(\{P_j\}_{j=1}^m, \{Q_i\}_{i=1}^N \right) = \sum_{i=1}^N \sum_{j \in \mathcal{A}_i} d^2(\mathbf{q}_{i,j}, P_j Q_i) \quad (2)$$

où $d^2(\mathbf{q}, \mathbf{q}') = \|\mathbf{q} - \mathbf{q}'\|^2$, $\mathbf{q}_{i,j}$ l'observation du point 3D Q_i dans la caméra C_j , P_j la matrice de projection associée définie par $P_j = K R_j^T (Z_3 | -\mathbf{t}_j)$. $\mathcal{A}_i$ l'ensemble des indices de caméra observant Q_i .

Afin de pouvoir gérer un large panel d'objets 3D, deux catégories de contraintes $\mathcal{E}_M$ ont été proposées. La première contraint le déplacement de la caméra via la projection de primitives du modèle 3D dans les images (section 2.1), tandis que la seconde contraint les primitives reconstruites à appartenir à la surface du modèle (section 2.2).

2.1 AF contraint aux primitives du modèle

Dans ce cas, le terme $\mathcal{E}_M$ minimise la distance entre les primitives du modèle projetées et les primitives dans l'image. Ainsi, disposant d'un modèle géométrique de l'objet (ex modèle CAO), $\mathcal{E}_M$ minimise la somme des distances orthogonales entre chaque segment projeté L_i et le point de contours $\mathbf{m}_{i,j}$ le plus proche dans l'image.

$$\mathcal{E}_M \left(\{P_j\}_{j=1}^m \right) = \sum_{i=1}^s \sum_{j \in \mathcal{S}_i} |\mathbf{n}_{i,j} \cdot (\mathbf{m}_{i,j} - P_j \mathbf{M}_i)| \quad (3)$$

où $\mathbf{n}_{i,j}$ est la normale au segment projeté et $\mathbf{M}_i$ le point milieu de L_i et $\mathcal{S}_i$ l'ensemble des indices de caméra observant les segments L_i .

Comparativement aux méthodes de type Model Based Tracking [4, 19], cette fonction de coût utilise sensiblement le même critère mais étendu au cas multi-vues sur $\mathcal{S}_i$. De manière identique, dans le cas de points 3D connus du modèle, notés $\bar{Q}_i$, la contrainte $\mathcal{E}_M$ s'écrit :

$$\mathcal{E}_M \left(\{P_j\}_{j=1}^m \right) = \sum_{i=1}^s \sum_{j \in \mathcal{S}_i} d^2(\mathbf{q}_{i,j}, P_j \bar{Q}_i) \quad (4)$$

2.2 AF contraint aux plans du modèle

Dans ce cas, le terme $\mathcal{E}_M$, contraint les points reconstruits à appartenir au modèle, ce dernier étant modélisé par un ensemble de plan π_i . Ainsi un point 3D Q_i associé à un plan π_i du modèle, noté $Q_i^{\pi_i}$, a uniquement deux degrés de liberté :

$$\mathcal{E}_M \left(\{P_j\}_{j=1}^m, \{Q_i^{\pi_i}\}_{i=1}^N \right) = \sum_{i=1}^N \sum_{j \in \mathcal{A}_i} d^2(\mathbf{q}_{i,j}, P_j M^{\pi_i} Q_i^{\pi_i}) \quad (5)$$

avec M^{π_i} supposée connue, la matrice de passage entre le repère du plan π_i et le repère du monde attaché à l'objet.

2.3 Intégration dans le SLAM et résultats

L'équ. 1 est minimisée dans l'AF du SLAM visuel au moyen de l'algorithme du Levenberg Marquardt dans lequel un M-estimateur de Geman-McClure permet de régler dynamiquement le poids des deux résidus $\mathcal{E}_E$ et $\mathcal{E}_M$ via

le seuil de rejet. L'association des primitives au modèle est également une phase cruciale. Dans le cas de l'équ. 5, celle-ci repose sur un vote majoritaire des re-projections des observations $q_{i,j}$ sur les plans π_i du modèle. Celle-ci est remise en cause après chaque minimisation de la fonction de coût : à savoir à chaque image clés dans le cas d'un AF local du SLAM ou dans un processus itératif, alternant la phase d'association primitives-modèle et la minimisation de la fonction de coût dans le cas d'un AF global (section 3.2). Pour plus de détail, le lecteur peut se référer à [17]. Ce système se comporte alors comme un SLAM classique lorsque l'objet n'est pas visible et comme un SLAM contraint dès que l'objet apparaît.

Les résultats comparatifs entre le SLAM contraint, le SLAM classique et une méthode de type model-based [4] sont présentés dans [18] et montrent l'apport de la méthode en terme de précision, de robustesse et stabilité. Des expérimentations sur différents types d'objets complexes et peu texturés (pièce industrielle, automobile, mobiliers etc.) nous ont permis de démontrer l'intérêt de l'approche et de mettre en évidence sa robustesse aux occultations, aux mouvements brusques et aux fortes variations de distance par rapport à l'objet. Cette dernière peut s'appliquer à des objets texturés ou non texturés. La figure 1 présente quelques résultats. D'autres sont également présentées dans [1][5][18].



FIGURE 1 – RA sur un véhicule et sur des équipements industriels.

Cette technologie est actuellement utilisée par la société DiotaSoft qui commercialise des systèmes de RA mobiles. La principale limitation de cette solution est qu'elle ne fonctionne aujourd'hui que sur des objets statiques. Nos travaux futurs visent à proposer des solutions sur objets mobiles. Le défi est alors de trouver un critère permettant de détecter lorsque la contrainte de rigidité de la scène n'est plus respectée et de dissocier dans l'AF les déplacements de la camera et ceux des d'objets mobiles.

3 Localisation dans de grands environnements

L'idée conductrice de ces travaux, a été comme précédemment d'intégrer l'information géométrique, ici issue des modèles 3D des Systèmes d'Information Géographique SIG. Cependant comparativement au cadre applicatif précédent, nous sommes confrontés aux problèmes majeurs suivants :

- de longues trajectoires sur lesquelles l'effet de dérive est plus important,
- des modèles géométriques plus pauvres et imprécis,
- des occultations plus importantes (véhicules mobiles, végétation ...)

Les modèles 3D SIG utilisés sont des modèles de bâtiments, reconstruits par photogrammétrie aérienne et calés sur les bords de gouttière, avec une précision est de l'ordre de $1m$. D'autre part, en ligne droite la géométrie apportée par les façades est relativement pauvre (un plan vertical de part et d'autre de la chaussée) et ne contraint que 3 DOF de la caméra (t_x, t_y et l'angle de lacet)². Seuls les carrefours et les virages apportent potentiellement des contraintes géométriques. Cependant, très souvent à chaque virage, le SLAM visuel a trop dérivé pour intégrer dans l'AF la contrainte au modèle 3D SIG (équ. 1 avec 2 et 5) : la dérive est si importante que les associations points-modèle sont pour la plupart erronées et compromettent la convergence de l'AF contraint (voir fig 3).

Une de nos premières approches [14] a consisté à corriger la dérive du SLAM par un facteur d'échelle déduit des homographies calculées sur les points du sol, connaissant la position et l'orientation de la caméra par rapport au sol, puis à recalculer après chaque virage, le nuage de points reconstruit sur les modèles 3D par une technique d'ICP rigide. L'ICP est moins sensible aux erreurs d'association point-plan, comparativement à un AF, car elle ne calcule qu'une transformation rigide entre le nuage de points reconstruit et les bâtiments. Cependant ce calcul étant réalisé avec un retard temporel (*a posteriori* après chaque virage), il ne répond pas au critère temps réel exigé par la RA.

D'autre part, les modèles 3D SIG n'apportent aucune contrainte sur le plan horizontal. Sans prise en compte d'une telle contrainte, l'altitude de la caméra dérive et altère les associations point-modèle, rendant la méthode de plus en plus fragile au cours du temps.

Ceci nous a conduits à proposer des solutions de fusion avec d'autres données pour réduire la dérive et d'autres approches d'association point-modèle pour atteindre le bassin de convergence de l'ajustement de faisceau contraint. Ce problème a alors été abordé à travers deux pistes :

- la localisation en ligne par un SLAM contraint au modèle 3D SIG et au modèle numérique d'élévation de terrain MET,
- la localisation en ligne combinant SLAM visuel et re-

2. Le facteur d'échelle et l'angle de roulis sont quant à eux très peu contraints.

connaissance de points de vue, en nous focalisant particulièrement sur le problème de construction préalable d'une base de données grande échelle géo-référencées. Ces deux contributions sont présentées ci-après.

3.1 SLAM contraint pour la localisation en ligne

Cette solution présentée dans [10] s'appuie sur la notion d'ajustement de faisceaux local contraint (équ. 1 avec 2 et 5) et étend celle-ci au contexte urbain en s'attachant à résoudre le problème difficile de dérive en altitude du SLAM et d'association points-modèle. Comme vu précédemment, les modèles 3D SIG n'apportent aucune contrainte sur l'altitude la caméra. Nous avons donc proposé dans [10] un SLAM visuel reposant sur un ajustement de faisceaux contraint au MET et aux modèles 3D des bâtiments (SIG). Une solution immédiate pour intégrer la contrainte au plan horizontal, apportée par le MET, dans l'AF aurait pu être l'équ. 5, consistant à contraindre les points 3D reconstitués sur le sol à appartenir au MET. Cependant, dans un contexte urbain en condition de trafic normal, nous constatons que la majorité des points extraits ne sont pas sur le sol et que les quelques points détectés sur le plan horizontal sont plutôt des points sur les véhicules ou autres objets de la scène. Nous avons donc proposé une autre fonction de coût pour l'AF contraint. Celle-ci intègre d'une part, une contrainte au modèle des bâtiments pour les points reconstitués associés aux façades et d'autre part une contrainte sur les poses des caméras limitant ces dernières à rester à une altitude constante et connue par rapport au MET.

$$\mathcal{E} \left(\left\{ \mathbf{P}_j^{\psi_i} \right\}, \left\{ \mathbf{Q}_i \right\}_u, \left\{ \mathbf{Q}_i^{\pi_i} \right\}_{\mathcal{M}} \right) = \mathcal{E}_E + \mathcal{E}_M \quad (6)$$

avec

$$\mathcal{E}_E \left(\left\{ \mathbf{P}_j^{\psi} \right\}, \left\{ \mathbf{Q}_i \right\}_u \right) = \sum_{i \in u} \sum_{j \in \mathcal{A}_i} d^2(\mathbf{q}_{i,j}, \mathbf{P}_j^{\psi} \mathbf{L}_j \mathbf{Q}_i)$$

$$\mathcal{E}_M \left(\left\{ \mathbf{P}_j^{\psi} \right\}, \left\{ \mathbf{Q}_i^{\pi_i} \right\}_{\mathcal{M}} \right) = \sum_{i \in \mathcal{M}} \sum_{j \in \mathcal{A}_i} d^2(\mathbf{q}_{i,j}, \mathbf{P}_j^{\psi} \mathbf{L}_j \mathbf{M}^{\pi_i} \mathbf{Q}_i^{\pi_i})$$

où $\mathbf{P}_j^{\psi}$ est la matrice de projection de la caméra j dont la pose, exprimée dans le plan de la route, est définie par 5 degrés de liberté. $\mathbf{L}_j$ est la matrice de passage entre le plan du MET associé à la caméra j et le repère monde.

La segmentation du nuage de points, entre ceux associés au modèle et ceux associés à l'environnement, se basant uniquement sur un lancer de rayon comme dans le cas de suivi d'objets 3D n'est pas suffisant dans le contexte véhicule, en raison d'un trop grand nombre d'occultations. Nous avons donc proposé dans [10] une amélioration de la segmentation reposant sur une analyse statistique et temporelle des distances séparant chaque point $\mathbf{Q}_i$ à son plan le plus proche dans le modèle 3D SIG.

La figure 2 présente deux trajectoires de 1500m et 1000m dans un quartier urbain, parcourues par un véhicule équipé du système et la figure 3 présente les résultats obtenus (d'autres expérimentations sont également présentées dans

[10]). L'intégration directe des contraintes dans le processus d'optimisation du SLAM permet d'améliorer considérablement la précision de la localisation. Notons que cette solution fournit la pose de la caméra, en ligne et en temps réel (à 30 Hz sur un PC standard). Cependant, cette approche manque de robustesse lorsque peu de contraintes géométriques sont disponibles (i.e. quand peu de bâtiments sont observés). Elles ne s'appliquent donc qu'en contexte urbain dense. En vue de lever cette limite, nos travaux actuels portent sur l'intégration de données GPS.



FIGURE 2 – Les deux trajectoires parcourues par le véhicule en environnement urbain et images acquises.

3.2 SLAM contraint pour la création d'une base d'amers géo-référencés

Une autre approche pour résoudre le problème de localisation d'une caméra dans un grand environnement consiste à exploiter une base d'amers géo-référencés. Dans [6][7], nous avons proposé une solution combinant un SLAM visuel avec un système de reconnaissance de points de vue. Cette méthode fait cohabiter un SLAM visuel avec un algorithme de relocalisation calculant une pose absolue par la mise en correspondance de l'image courante avec une base d'amers géo-référencés. L'intérêt principal de cette approche repose sur son fonctionnement en mode SLAM lorsque aucune correspondance avec la base de données n'est possible (changement d'illumination, base incomplète ou non conforme avec la scène) et dans sa possibilité de corriger ponctuellement la dérive du SLAM dès qu'une image de la base est reconnue. Une des problématiques majeures de cette solution repose sur la création préalable d'une base précise d'amers géo-référencés à grande échelle.

Tandis que les solutions existantes exigent des matériels coûteux et/ou un temps d'exécution très important, notre contribution porte sur un processus qui crée cette base, automatiquement et en quelques minutes, en utilisant uniquement une caméra standard, un GPS bas coût et des modèles SIG [11]. Le but ultime visé est de proposer un système collaboratif dans lequel l'utilisateur lui-même participera à l'enrichissement et la mise à jour de cette base.

Cette solution [11] repose un processus fusionnant les



FIGURE 3 – Localisation en milieu urbain avec un SLAM contraint : (a) SLAM classique, (b) SLAM contraint au modèle des bâtiments SIG, (c) SLAM contraint au modèle SIG et MET.

contraintes multi-vues du SLAM, les mesures GPS et les contraintes géométriques apportées par les modèles 3D des bâtiments et le MET. Étant donné qu'il n'est pas trivial de fusionner simultanément toutes ces contraintes sans perturber la convergence du processus d'optimisation, nous proposons une solution qui se focalise dans un premier temps sur une correction grossière de la dérive du SLAM avant d'optimiser plus finement la base d'amers résultante dans un ajustement de faisceaux global.

La première étape repose sur un SLAM visuel intégrant en ligne les mesures de GPS et les données MET dans un même processus d'optimisation. Afin de garantir une certaine robustesse face aux données aberrantes du GPS et aux incertitudes du MET, le processus d'optimisation s'appuie sur un ajustement de faisceaux avec contrainte d'inégalité inspirée de la méthode introduite dans [13]. Cette approche permet de créer en ligne et automatiquement une reconstruction initiale géo-référencée et cohérente. Cependant, sa précision reste limitée à l'incertitude du GPS. Pour cette raison, la base obtenue est raffinée, *a posteriori*, à travers deux ajustements de faisceaux globaux contraints au modèle SIG complet de la scène (i.e les modèles 3D des bâti-

ments et le MET). Le premier AF global minimise la fonction de coût définie par : l'équ. 1 avec 2 et 5, prenant ainsi en compte la contrainte aux bâtiments sur l'ensemble de la structure reconstruite. Pour le deuxième AF global, la fonction de coût est composée d'un terme de régularisation calculé à partir de l'erreur de re-projection intégrant la contrainte aux bâtiments (equ. 1 avec 2 et 5) et d'un terme de pénalité reposant sur la contrainte MET. Cette pénalité représente la distance entre l'altitude de chaque pose de la caméra et l'altitude souhaitée, supposée connue.

Des évaluations sur différents sites urbains sont présentées dans [11]. La figure 4 montre la base d'amers obtenues sur une des trajectoires de la figure 2 et illustre le gain en précision pour chacune des étapes. Cette base d'amers est ensuite exploitée pour une localisation en ligne combinant SLAM et relocalisation. La figure 5 présente un exemple de reprojection des modèles SIG dans la séquence et un résultat de guidage par RA.

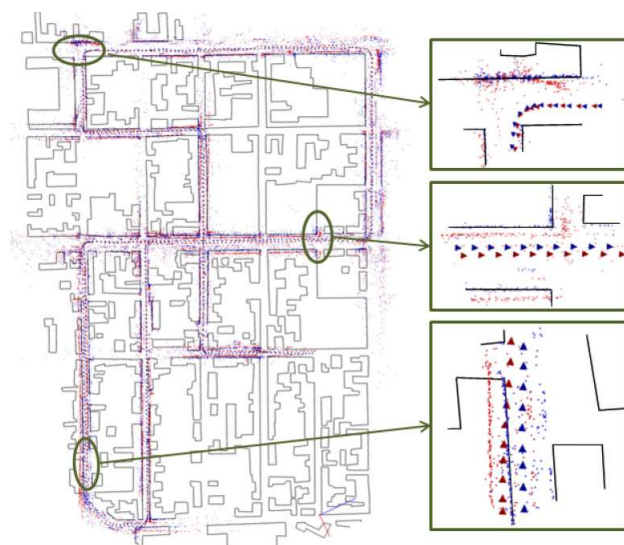


FIGURE 4 – Création d'un nuage de points géo-référencés : résultat avec le SLAM contraint aux données GPS et MET (rouge) puis avec ajustement de faisceaux global contraint au modèle SIG complet (bleu).

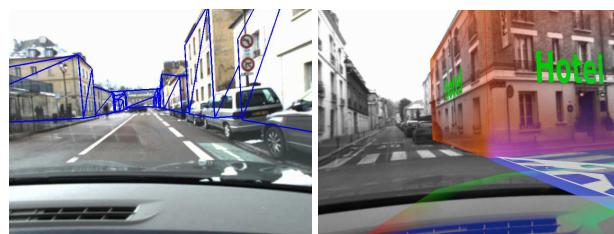


FIGURE 5 – Projection des modèles SIG dans l'image (erreur moyenne de 6 pixels) et guidage par réalité augmentée

3.3 Conclusion

Nous avons présenté une nouvelle approche, appelée SLAM contraint, pour résoudre le problème de localisation d'une caméra mobile pour la Réalité Augmentée : RA sur des objets 3D et RA pour le guidage dans des environnements urbains. Ces deux applications requièrent une estimation temps réel, précise et stable de la pose de la caméra. Le SLAM contraint, repose sur un ajustement de faisceaux local intégrant des informations absolues inhérentes à la scène (modèle 3D, GPS). Nous avons montré que le SLAM contraint répond aux critères recherchés pour la RA sur des objets 3D statiques. Pour les applications de RA en milieu urbain, les effets de dérives et d'association primitives-modèles sont plus difficiles à gérer. De ce fait, deux solutions ont été investiguées. La première repose sur un SLAM contraint au modèle SIG et MET de la scène. Cette méthode permet de localiser en ligne (à 30Hz) la caméra en milieu urbain dense. Elle reste cependant sensible lorsque la géométrie de la scène est pauvre. Pour pallier cette limite, nous étudions une approche combinant un SLAM visuel avec des données GPS, le MET et les modèles de bâtiments du SIG.

La deuxième solution, consiste à combiner un SLAM avec un algorithme de relocalisation s'appuyant sur une base d'amers géo-référencés grande échelle. L'ajustement de faisceaux contraint est alors mis à profit pour construire cette base d'amers de manière automatique et en quelques minutes. L'approche visée est de proposer à terme un système mettant à jour la base le plus fréquemment possible grâce aux usagers et aux véhicules connectés. Un des principaux verrous à lever dans ce dernier cas, concerne le critère de qualité de la base reconstruite et la manière de la fusionner avec la base existante. Enfin, ces travaux en milieu urbain, se sont focalisés jusqu'à présent au contexte automobile. Nos travaux actuels s'intéressent également au guidage de piéton pour lequel le mouvement est plus irrégulier et où l'hypothèse de connaissance *a priori* de l'altitude de la caméra devra être levée.

Références

- [1] B. Besbes, S. Naudet-Collette, M. Tamaazousti, S. Bourgeois, and V. Gay-Bellile. An interactive augmented reality system : a prototype for industrial maintenance training applications. In *ISMAR*, 2012.
- [2] G. Bleser, H. Wuest, and D. Stricker. Online camera pose estimation in partially known and dynamic scenes. In *ISMAR*, 2006.
- [3] R. O. Castle, D. J. Gawley, G. Klein, and D. W. Murray. Video-rate recognition and localization for wearable cameras. In *BMVC*, 2007.
- [4] T. Drummond and R. Cipolla. Real-time visual tracking of complex structures. *PAMI*, 24(7) :932–946, 2002.
- [5] G. Gay-Bellile, S. Bourgeois, M. Tamaazousti, and S. Naudet-Collette. A mobile markerless augmented reality system for the automotive field. In *ISMAR Workshop*, 2012.
- [6] V. Gay-Bellile, Pierre Lothe, S. Bourgeois, E. Royer, and S. Naudet-Collette. Augmented reality in large environments : Application to aided navigation in urban context. In *ISMAR*, 2010.
- [7] V. Gay-Bellile, M. Tamaazousti, R. Dupont, and S. Naudet-Collette. A vision-based hybrid system for real-time accurate localization in an indoor environment. In *VISAPP*, 2010.
- [8] T. Kempter, A. Wendel, and H. Bischof. Online model-based multi-scale pose estimation. In *Computer Vision Winter Workshop, CVWW*, 2012.
- [9] G. Klein and D. Murray. Parallel tracking and mapping for small AR workspaces. In *ISMAR*, 2007.
- [10] D. Larnaout, S. Bourgeois, V. Gay-Bellile, and M. Dhome. Towards bundle adjustment with gis constraints for online geo-localization of a vehicle in urban center. In *3DIMPVT*, pages 348–355, 2012.
- [11] D. Larnaout, V. Gay-Bellile, S. Bourgeois, B. Labbé, and M. Dhome. Driving in an augmented-city : from fast and automatic large scale environment modeling to on-line 6dof vehicle localization. In *VRCAI*, 2013.
- [12] V. Lepetit and P. Fua. Monocular model-based 3d tracking of rigid objects : A survey. *Foundations and Trends in Computer Graphics and Vision*, 2005.
- [13] M. Lhuillier. Incremental fusion of structure-from-motion and gps using constrained bundle adjustments. *IEEE PAMI*, 34(12) :2489–2495, 2012.
- [14] P. Lothe, S. Bourgeois, E. Royer, M. Dhome, and S. Naudet-Collette. Real-time vehicle global localization with a single camera in dense urban areas : Exploitation of coarse 3d city models. In *CVPR*, 2010.
- [15] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd. Real time localization and 3d reconstruction. In *CVPR*, 2006.
- [16] M. Tamaazousti, V. Gay-Bellile, S. Naudet-Collette, and M. Dhome. Localisation précise et temps réel dans un environnement partiellement connu : application au suivi d'objets 3d. In *RFIA*, 2012.
- [17] M. Tamaazousti, V. Gay-Bellile, S.N. Naudet-Collette, S. Bourgeois, and M. Dhome. Nonlinear refinement of structure from motion reconstruction by taking advantage of a partial knowledge of the environment. In *CVPR*, 2011.
- [18] M. Tamaazousti, V. Gay-Bellile, S.N. Naudet-Collette, S. Bourgeois, and M. Dhome. Real-time accurate localization in a partially known environment : application to augmented reality on textureless 3d objects. In *ISMAR workshop*, 2011.
- [19] H. Wuest, D. Stricker, and J. Herder. Tracking of industrial objects by using cad models. *Journal of Virtual Reality and Broadcasting*, 4(1), 2007.