



HAL
open science

Estimation des géométries planaire et épipolaire en environnement faiblement texturé basée sur la mise en correspondance de régions

Vincent Gaudillière, Gilles Simon, Marie-Odile Berger

► **To cite this version:**

Vincent Gaudillière, Gilles Simon, Marie-Odile Berger. Estimation des géométries planaire et épipolaire en environnement faiblement texturé basée sur la mise en correspondance de régions. RFIAP 2018 - Congrès Reconnaissance des Formes, Image, Apprentissage et Perception, Jun 2018, Marne-la-Vallée, France. hal-01802538

HAL Id: hal-01802538

<https://hal.science/hal-01802538>

Submitted on 29 May 2018

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.

Estimation des géométries planaire et épipolaire en environnement faiblement texturé basée sur la mise en correspondance de régions

Vincent Gaudillière¹

Gilles Simon¹

Marie-Odile Berger¹

¹ Inria Nancy - Grand Est / Loria, Nancy, France

vincent.gaudilliere@inria.fr

Résumé¹

Etant données deux vues d'une même scène, les techniques classiques de mise en correspondance géométrique s'appuient généralement sur l'efficacité maintes fois démontrée des descripteurs de points d'intérêt. Cependant, de tels indices sont mis en difficulté dans des environnements peu texturés, contenant possiblement des structures répétées et/ou des spécularités, comme les environnements industriels. Dans ce papier, nous proposons une nouvelle méthode pour l'estimation des géométries épipolaire et planaire entre deux vues, qui vise tout d'abord à détecter et mettre en correspondance les plans verticaux souvent présents dans ces environnements, avant d'estimer les homographies correspondantes. Les correspondances locales ainsi générées sont ensuite utilisées pour améliorer l'estimation de la matrice fondamentale. Le gain en précision a été démontré sur des images d'environnements industriel et urbain.

Mots Clef

Estimation de géométrie épipolaire, estimation d'homographie, détection de plan, environnements industriels faiblement texturés.

Abstract

Given two views of the same scene, usual correspondence geometry estimation techniques classically exploit the well-established effectiveness of keypoint descriptors. However, such features have a hard time in poorly textured man-made environments, possibly containing repetitive patterns and/or specularities, such as industrial places. In that paper, we propose a novel method for two-view epipolar and planar geometry estimation that first aims at detecting and matching physical vertical planes frequently present in these environments, before estimating corresponding homographies. Inferred local correspondences are finally used to improve fundamental matrix estimation. The gain in precision is demonstrated on industrial and urban environments.

Keywords

Epipolar geometry estimation, homography estimation, plane detection, low-textured industrial environments.

1 Introduction

Le problème de la mise en correspondance géométrique entre deux vues a été étudié de manière extensive au cours des dernières décennies. Les méthodes les plus employées consistent à mettre en correspondance, à partir de leurs descripteurs, des points d'intérêt préalablement extraits [1]. L'extraction et la description de points d'intérêt a connu son âge d'or à partir du moment où SIFT a été introduit [2]. Depuis lors, de nombreux travaux se sont intéressés à la conception de méthodes plus rapides [3, 4, 5], ou invariantes à davantage de transformations [6]. Cependant, ces méthodes fonctionnent dans des environnements bien texturés, mais échouent dans le cas inverse. De plus, le ratio de Lowe, qui demande, pour qu'une correspondance soit valable, que le rapport des distances au plus proche voisin et au deuxième plus proche voisin soit suffisamment éloigné de 1 [2], conduit à ne pas inclure les motifs répétés dans l'ensemble initial de correspondances.

Les segments sont très présents en milieux industriels et peuvent apparaître comme des indices appropriés, même si leur mise en correspondance reste un problème très difficile. Plusieurs stratégies de mise en correspondance ont été étudiées par le passé, et la plupart d'entre elles consiste à estimer itérativement un modèle géométrique parmi un ensemble de correspondances supposées (algorithmes de type RANSAC). La mise en correspondance initiale des segments peut se faire en étudiant la similarité de leurs voisinages [7, 8, 9]. Cependant, les segments présents dans les images d'environnements industriels correspondent souvent à des arêtes d'objets tridimensionnels. Autrement dit, leurs voisinages sont souvent constitués d'une partie pauvre en information (couleur unie par exemple), et d'une autre dont le contenu dépend du point de vue d'observation. Le même argument peut aussi être opposé aux méthodes dans lesquelles les segments sont décrits par les points d'intérêt présents dans leurs voisinages [10, 11]. A l'inverse, il est possible de créer l'ensemble de correspondances potentielles en se basant sur des critères géométriques.

1. Ce travail est soutenu par le projet DGA/DGE RAPID EVORA.

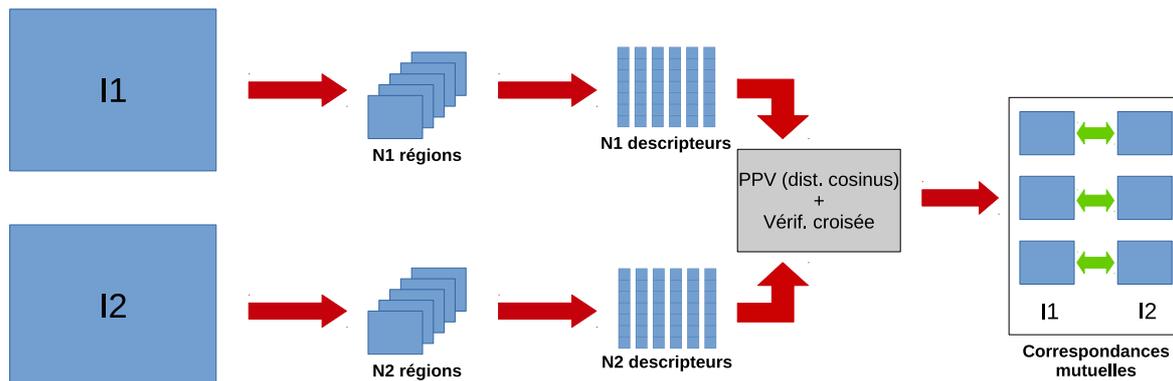


FIGURE 1 – Principe de notre algorithme de mise en correspondance des régions. Des centaines de régions d’intérêt sont extraites des deux images, puis un descripteur CNN est associé à chacune des régions. Pour chaque descripteur, une recherche du plus proche voisin (PPV) parmi les descripteurs de l’autre image est effectuée, avant qu’une étape de vérification croisée ne sélectionne que les correspondances mutuelles.

triques [12, 13, 14], mais ces méthodes centrées sur les invariants restent hautement sensibles au bruit, et se traduisent souvent par une explosion combinatoire. Enfin, la mise en correspondance de groupes de segments peut permettre de compenser le manque d’information associée à chaque segment, mais là encore au prix d’une combinatoire élevée [15, 16].

Dans notre travail, nous tirons profit de plusieurs propriétés globales des environnements industriels pour guider la mise en correspondance :

- Premièrement, les environnements industriels sont souvent constitués d’un nombre important de plans verticaux. Partant du constat que les informations à propos des plans existant dans une image sont d’une certaine manière contenues dans les points de fuite, la manière dont notre méthode tire profit de ces derniers est double. Tout d’abord, nous n’utilisons que les segments associés à des points de fuite, et ces associations sont ensuite utilisées pour contraindre la mise en correspondance des segments. Ensuite, les points de fuite sont utilisés comme primitives pour estimer les homographies locales, ce qui permet d’une part de réduire la combinatoire, et d’autre part de limiter le recours aux points d’intérêt visuels.
- Deuxièmement, il apparaît dans [17] que les régions (sous-parties) d’images sont des indices robustes qui peuvent être mis en correspondance même en présence de changements de conditions et/ou de points de vue. En effet, l’application de descripteurs globaux à des sous-images représente un bon compromis entre l’utilisation de descripteurs locaux, qui sont robustes aux changements de points de vue mais beaucoup moins aux changements de conditions, et l’utilisation de descripteurs globaux, qui présentent les performances inverses [18]. Dans notre méthode, les correspondances

entre régions sont utilisées comme hypothèses pour la détection et la mise en correspondance des plans verticaux. De cette façon, la mise en correspondance des segments est davantage contrainte, ce qui permet de contourner les difficultés inhérentes aux environnements industriels sans compromettre le temps de calcul.

Étant données deux vues différentes d’une même scène, notre algorithme de traitement se décompose de la manière suivante : (1) Des **correspondances entre régions** des deux images sont calculées (Section 2). (2) Les homographies locales qui peuvent exister entre les régions mises en correspondance sont détectées, puis fusionnées pour identifier les plans verticaux présents (Section 3). (3) Les correspondances entre segments générées par les homographies sont utilisées pour améliorer l’estimation de la géométrie épipolaire (Section 4).

2 Extraction et mise en correspondance des régions

Le première étape de notre méthode consiste à générer des hypothèses de correspondances entre régions, dans le but de calculer par la suite des homographies locales.

Notre algorithme de mise en correspondance des régions est largement inspiré du travail de Sünderhauf *et al.* [17]. Dans ce travail, des correspondances entre régions sont établies dans le but de reconnaître des lieux de manière robuste, en présence de forts changements de conditions et de points de vue. Pour ce faire, des hypothèses de régions d’intérêt sont générées dans les deux images sous la forme de boîtes englobantes, avant qu’une carte de caractéristiques issue d’un CNN ne soit associée à chacune d’entre elles. Ainsi, chaque image est décrite par une collection de descripteurs CNN associés à différentes sous-images. Après une étape de réduction dimensionnelle, la mise en correspondance des descripteurs est réalisée sous la forme

d'une recherche du plus proche voisin dans l'autre image, suivie d'une étape de vérification croisée qui permet de ne conserver que les correspondances mutuelles.

Dans notre méthode, illustrée Fig. 1, nous suivons globalement la même approche, mis à part le fait que nous considérons directement des caractéristiques CNN de bas-niveau (c'est-à-dire la dernière couche avant les couches complètement connectées) en tant que descripteurs de régions, ce qui nous enlève la nécessité d'une étape de réduction dimensionnelle. Plus précisément, le CNN utilisé lors de nos expériences est celui décrit dans [24], mais un autre modèle peut bien sûr être exploité. En entrée du CNN, l'image est redimensionnée à 227×227 pixels, puis le descripteur extrait est un vecteur composé de 4096 éléments. Par ailleurs, la méthode *Edge Boxes* [19] est utilisée pour générer des hypothèses d'objets, car elle apparaît particulièrement adaptée aux scènes essentiellement constituées de contours, comme c'est le cas dans les environnements industriels.

3 Estimation des homographies locales

Une fois les sous-images mises en correspondance, notre méthode a pour objectif de détecter les homographies locales entre régions en correspondances. Pour assurer des estimations efficaces, nous avons développé un algorithme de type RANSAC dans lequel les hypothèses de modèle sont d'abord générées à partir des points d'intérêt visuels et des points de fuite, avant que ces mêmes hypothèses ne soit validées sur les points d'intérêt et les segments. Ce schéma nous permet de contourner les difficultés relatives aux images peu texturées (faible densité de points d'intérêt visuels et difficile mise en correspondance des segments), tout en tirant profit de l'abondance de segments et de points de fuite inhérente aux environnements industriels. Notre méthode est illustrée Fig. 2.

3.1 Extraction et mise en correspondance des indices

Les **points de fuite** (PF) sont détectés dans les deux images en utilisant la méthode décrite dans [20]. Cette méthode détecte le zénith ainsi que les points de fuite horizontaux présents dans les images, chacun de ces derniers étant donc associé à un plan vertical. Les zéniths sont ensuite directement mis en correspondance entre les deux images.

Les **segments** sont ensuite extraits par LSD [21]. Les segments associés aux PF préalablement extraits sont conservés, tandis que les autres sont abandonnés. L'association segment-PF se fait en considérant un seuil sur l'angle entre la droite à laquelle le segment appartient et la droite reliant le milieu du segment au PF. Ainsi, le segment est associé au PF si et seulement si cet angle est inférieur au seuil considéré.

Les **points d'intérêt** (PI) sont enfin détectés en utilisant une méthode appelée LIFT [6], qui est une version *apprise* de SIFT. Les correspondances entre PI des deux images

sont ensuite obtenues en appliquant le ratio de Lowe dans les deux sens, puis en ne gardant que les correspondances mutuelles, afin de maximiser le taux d'inliers au sein de l'ensemble de correspondances initial E_0 .

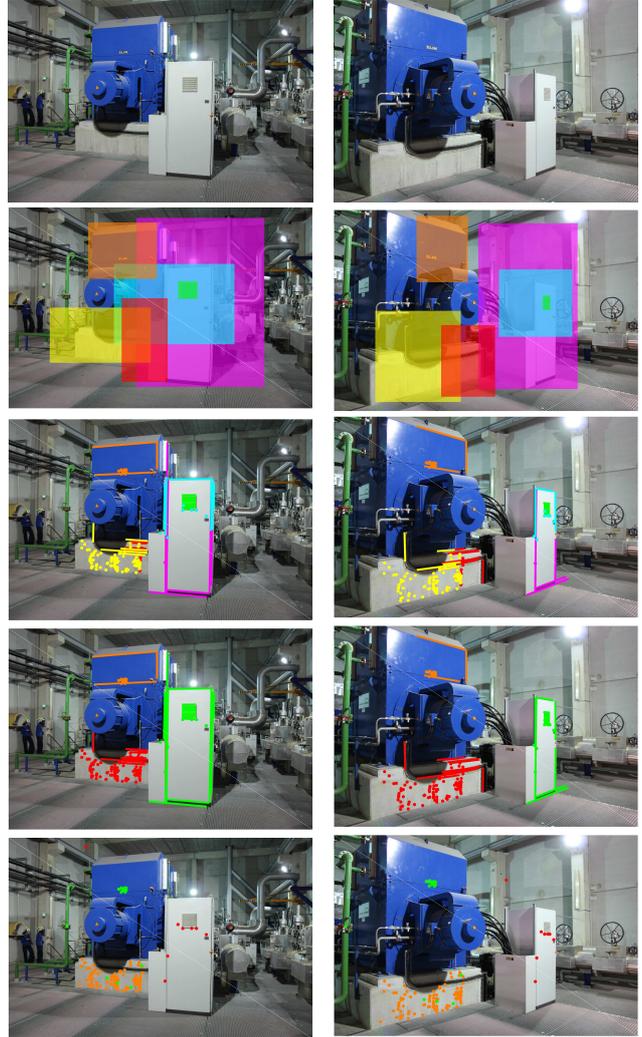


FIGURE 2 – Vue d'ensemble de notre méthode d'estimation des homographies locales. Ligne 1 : Deux vues d'un environnement industriel. Ligne 2 : Exemples de régions mises en correspondance (une couleur par correspondance). Ligne 3 : Inliers (points et segments) des homographies estimées entre régions mises en correspondance (une couleur par homographie). Ligne 4 : Inliers après fusion des homographies. Ligne 5 : Inliers des trois premiers plans obtenus par un algorithme multi-RANSAC 4 points.

3.2 Estimation des homographies

Sélection des régions En raison de la spécificité des environnements considérés, les régions extraites puis mises en correspondance lors de la première étape peuvent être considérées comme des hypothèses de plans verticaux. Cependant, il existe de nombreux recouvrements entre les différentes boîtes englobantes d'une même image, donc tes

ter l'ensemble des paires de boîtes pour y trouver des homographies ferait exploser inutilement le temps de calcul. Pour remédier à cela, les paires de boîtes sont d'abord classées grossièrement des plus petites vers les plus grandes, en considérant comme critère pour chaque paire l'aire moyenne des deux boîtes. L'idée principale est ensuite de traiter les paires de boîtes dans cet ordre, tout en tenant à jour deux cartes des pixels déjà visités (une par image). Ces cartes ont la même résolution que les images originales, et chaque pixel y prend la valeur 1 si une boîte le contenant a déjà été testée, 0 sinon. Une paire de boîtes est testée si et seulement si moins de 50% des pixels de chaque boîte ont déjà été visités.

Sélection des indices Le traitement d'une paire de régions consiste en premier lieu à sélectionner les indices contenus dans chaque région (PI, segments, et PF associés aux segments), puis à appliquer une étape de fusion des segments dans le but de corriger la sur-segmentation induite par LSD. En pratique, les segments à fusionner sont déterminés en utilisant des contraintes tangentielle et normale similaires à celles présentées dans [14]. Plus précisément, les segments associés au même PF sont comparés deux à deux, et les segments qui vérifient ces deux contraintes, i.e. qui sont suffisamment alignés, sont fusionnés. Finalement, seuls les segments les plus longs (i.e. dont la longueur est supérieure à un certain pourcentage du plus grand segment associé au même PF) sont conservés, ce qui permet de ne garder que les segments les plus significatifs et de réduire la combinatoire associée à leur mise en correspondance.

Estimation robuste d'homographie En raison des difficultés liées à la mise en correspondance des segments, les hypothèses du RANSAC sont générées uniquement à partir des PI et des PF. A chaque itération, nous tirons au hasard une paire de PF horizontaux, deux paires de PI parmi E_0 , et la paire de zéniths. Les 4 paires de points définissent ainsi une homographie induite par un plan vertical. Contrairement à la méthode présentée dans [9], les PI sont ici directement utilisés comme primitives pour la définition des modèles. Ce choix nous permet de réduire drastiquement la combinatoire ainsi que le recours aux PI visuels, en limitant à simplement deux le nombre de correspondances entre PI nécessaires pour définir une homographie.

Pour déterminer le nombre d'itérations de RANSAC, nous utilisons une version modifiée de l'algorithme adaptatif présenté dans [1] (Section 4.7.1). Etant donné $N_{total}^{p_{fh}}$ le nombre de paires possibles entre PF horizontaux des deux images, la probabilité de tirer la paire correcte à n'importe quelle itération est : $w_{p_{fh}} = 1/N_{total}^{p_{fh}}$. Etant donné N_{total}^p le nombre de correspondances entre PI contenues dans E_0 , la probabilité de tirer une paire correcte est $w_p = N_{inliers}^p/N_{total}^p$, avec $N_{inliers}^p$ le nombre d'inliers contenus dans E_0 à l'itération courante. Le nombre N_{iter} d'itérations nécessaires pour assurer, avec une probabilité p , qu'au moins un ensemble de quatre primitives correctes a

été tiré au sort est alors donné par la formule :

$$N_{iter} = \frac{\log(1-p)}{\log(1-w_{p_{fh}} \cdot w_p^2)} \quad (1)$$

A chaque itération, N_{iter} est mis à jour à partir de la valeur courante de w_p , et l'algorithme s'arrête si l'indice de l'itération courante est plus grand que N_{iter} .

A titre de comparaison, il est important de noter que lorsqu'un seul type de primitive est utilisé, la formule donnant N_{iter} devient de la forme :

$$N_{iter} = \frac{\log(1-p)}{\log(1-w^4)} \quad (2)$$

Si on suppose une configuration avec deux PF horizontaux par image, 40% d'inliers points, et 1% d'inliers segments (segments non mis en correspondance), notre méthode nécessite 113 itérations, tandis que la méthode 4-points en nécessite 178, et la méthode 4-segments 4.6×10^8 . Si la paire de régions considérée est incorrecte, le gain de performance entre notre méthode et la méthode 4-points est beaucoup plus important.

Validation de modèle (RANSAC) La validation est basée sur les PI et les segments. Les inliers PI sont déterminés de manière classique, tandis que toutes les paires de segments possibles (entre les segments associés aux zéniths d'une part, et entre les segments associés aux deux PF horizontaux tirés au sort d'autre part) sont testées. Les paires segment transféré/segment qui satisfont à la fois les contraintes tangentielle et normale présentées dans [14] sont retenues. Pour éviter les configurations dégénérées, un segment de l'image originale ne peut être mis en correspondance qu'avec un seul segment de l'image cible (le plus proche au sens de la contrainte normale). A partir de là, le nombre de segments de l'image cible impliqués dans des paires qui satisfont au modèle d'homographie calculé ($N_{inliers}^s$) est ajouté au nombre de PI inliers, définissant ainsi un score pour le modèle H : $Score(H) = N_{inliers}^p + N_{inliers}^s$. Comme seuls les segments les plus significatifs ont été conservés, les correspondances entre segments ont de grandes chances d'être de même importance. De plus, le fait d'ajouter les contributions des points et des segments est justifié par le fait que cela permet de se reposer sur n'importe lequel de ces indices lorsque l'autre se fait rare.

Décision finale Les sous-images mises en correspondance pendant la première étape peuvent ne pas être liées par une homographie, soit parce qu'elles ne contiennent pas d'objets plans, soit parce que la correspondance est incorrecte. Ainsi, pour limiter la présence de fausses homographies, on ne considère que les homographies qui possèdent plus de 10 inliers.

3.3 Fusion des homographies

A ce stade, plusieurs homographies locales peuvent lier différentes parties d'un même plan physique. On applique

donc une étape de fusion qui consiste, pour chaque homographie, à tester les inliers des autres homographies. Si plus de 50% des inliers PI d’une homographie A sont aussi inliers d’une homographie B, et si la même performance est atteinte pour les inliers segments, alors A et B sont fusionnées.

4 Estimation de la géométrie épipolaire

Les paires de segments contenues dans les ensembles de consensus des homographies calculées à l’étape précédente peuvent désormais être utilisées pour améliorer l’estimation de la géométrie épipolaire. L’idée principale, inspirée par [8], est d’ajouter les intersections de segments à l’ensemble initial de correspondances entre PI E_0 , puis de d’utiliser ce nouvel ensemble de correspondances en entrée d’un algorithme d’estimation de la géométrie épipolaire appelé ORSA [22].

Dans notre méthode, les segments sont d’abord convertis en droites dans chacune des images. Pour chaque homographie trouvée à l’étape précédente, les intersections entre droites verticales (celles générées par les segments associés au zénith) et droites horizontales (celles générées par les segments associés au point de fuite horizontal) sont calculées. Ensuite, comme plusieurs segments peuvent générer approximativement la même droite (à cause de l’étape de fusion des homographies), et comme des correspondances quasi-identiques répétées peuvent perturber le bon fonctionnement de l’algorithme d’estimation de la géométrie épipolaire, nous avons fait le choix de diviser les images selon une grille régulière, et de ne garder au plus qu’un point d’intersection par case de la grille. Finalement, ces nouvelles correspondances entre points d’intersection (déduites des correspondances entre segments) sont ajoutées à l’ensemble initial de correspondances entre PI E_0 .

5 Résultats et conclusion

Les expériences que nous avons menées l’ont été sur deux ensembles d’images. Le premier est constitué de 46 paires d’images prises dans une usine d’électricité, tandis que le second est formé de 14 paires d’images d’environnements urbains tirées de la base présentée dans [23]. Ce choix d’images s’explique par le contexte applicatif de ce travail, qui est celui de la Réalité Augmentée. Le scénario classique est celui d’un opérateur, évoluant dans un environnement connu, qui chercherait à s’y positionner. Pour chaque paire d’images, la vérité terrain est constituée d’une vingtaine de paires de points placés à la main. Ces points ont été sélectionnés avec l’objectif de couvrir chaque image le plus largement possible, tout en assurant une certaine homogénéité de leur distribution. Les images industrielles ont une résolution de 1280×1920 pixels, tandis que les images urbaines ont une résolution de 640×640 . Il convient de noter que la méthode présentée dans [9] échoue à détecter des correspondances correctes entre segments dans la plupart

	Nb d’inliers moyen		Taux d’inliers moyen (%)	
	indus.	urban	indus.	urban
SIFT	198.6	102.8	30.02	52.3
LIFT	207.4	100.3	42.24	58.4
Nous	260.7	126.0	45.84	62.44

TABLE 1 – Performances moyennes sur les jeux de test, en termes de nombre d’inliers et de taux d’inliers.

des exemples considérés. Or ces correspondances sont nécessaires pour calculer la rotation entre les deux caméras. En leur absence, la méthode se repose alors uniquement sur les correspondances entre PI pour estimer la matrice essentielle, ce qui n’est pas pertinent dans les environnements considérés ici.

Dans la suite, notre méthode est comparée à la méthode classique qui consiste à (i) extraire et mettre en correspondance les PI des deux images en appliquant le ratio de Lowe dans les deux sens avant de ne garder que les correspondances mutuelles (ii) calculer la géométrie épipolaire à partir de ces correspondances en utilisant ORSA. SIFT et LIFT ont été utilisées comme méthode d’extraction de PI dans nos comparaisons. Les géométries épipolaires estimées ont finalement été comparées en termes d’inliers (nombre et taux), et de précision (erreur vis-à-vis de la vérité terrain).

Le précision des matrices fondamentales estimées est évaluée en Fig. 3. Pour chaque paire d’images, l’écart entre les points de la vérité terrain et les droites épipolaires estimées a été mesuré par l’intermédiaire de l’erreur quadratique moyenne (RMSE) et de l’erreur maximale. Ces mesures ont ensuite été moyennées sur l’ensemble des images de chaque base de données. En moyenne, notre méthode présente la plus grande précision, quel que soit le critère d’erreur considéré.

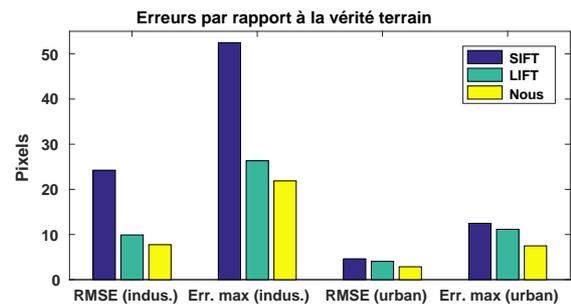


FIGURE 3 – Erreurs moyennes sur les bases d’images industrielles et urbaines (désignées par *indus.* et *urban*), par rapport à la vérité terrain.

Le tableau 1 présente la qualité des modèles estimés en termes d’inliers (ici encore, les résultats ont été moyennés sur les ensembles d’images). Notre méthode présente le plus grand nombre d’inliers, ainsi que le meilleur taux d’inliers, ce qui peut permettre d’améliorer la qualité des résultats d’une éventuelle étape suivante de reconstruction ou de

calcul de pose. La Fig. 4 illustre le plus grand nombre d'inliers obtenu grâce à notre méthode. Par ailleurs, il convient de noter qu'il existe 6 paires d'images industrielles pour lesquelles ORSA échoue à estimer un modèle à partir des correspondances entre PI SIFT, alors que cette même méthode y parvient avec nos correspondances.

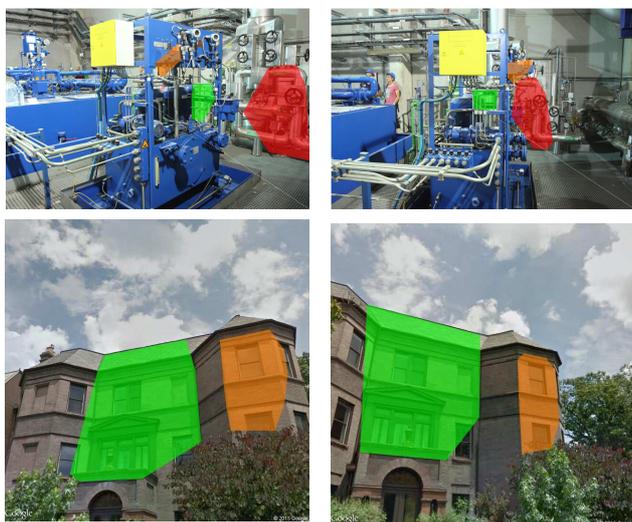


FIGURE 5 – Enveloppes convexes des inliers (points et segments) des homographies estimées par notre méthode, après l'étape de fusion, pour deux paires d'images de test (une couleur par homographie).

La Fig. 2 illustre la capacité de notre méthode à détecter et mettre en correspondance les plans verticaux présents dans les images (ligne 4), tandis qu'une méthode multi-RANSAC basée sur les correspondances entre PI ne parvient pas à séparer les différents plans physiques (cf inliers orange et verts), ni à en saisir les contours (cf inliers rouges et verts) (ligne 5). La Fig. 5 montre d'autres exemples de plans détectés par notre méthode (enveloppes convexes des inliers points et segments) dans des images de test d'environnements industriel et urbain.

Ces expériences montrent que notre méthode permet d'améliorer l'estimation de la géométrie épipolaire dans des environnements multiplanaires complexes, et ce même en l'absence d'un contenu visuel riche.

Références

- [1] R.I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, second edition, 2004.
- [2] David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, *Int. J. Comput. Vision*, vol. 60, pp. 91-110, Nov. 2004.
- [3] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, Speeded-Up Robust Features (SURF), *Comput. Vis. Image Underst.*, vol. 110, pp. 346-359, Jun. 2008.
- [4] Edward Rosten, and Tom Drummond, Machine Learning for High-Speed Corner Detection, *European Conference on Computer Vision*, 2006.
- [5] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha, and P. Fua, BRIEF : Computing a Local Binary Descriptor Very Fast, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 1281-1298, 2012.
- [6] Kwang Moo Yi, Eduard Trulls, Vincent Lepetit, and Pascal Fua, LIFT : Learned Invariant Feature Transform, *European Conference on Computer Vision*, 2016.
- [7] Lilian Zhang and Reinhard Koch, An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency, *Journal of Visual Communication and Image Representation*, vol. 24, pp. 794-805, 2013.
- [8] H. Bay, V. Ferrari, and L. Van Gool, Wide-Baseline Stereo Matching with Line Segments, *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- [9] Y. Salaün, R. Marlet, and P. Monasse, Robust and Accurate Line- and/or Point-Based Pose Estimation without Manhattan Assumptions, *European Conference on Computer Vision*, 2016.
- [10] Bin Fan, Fuchao Wu, and Zhanyi Hu, Line matching leveraged by point correspondences, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010.
- [11] H. Li, D. Song, Y. Lu, and J. Liu, A two-view based multilayer feature graph for robot navigation, *IEEE International Conference on Robotics and Automation*, 2012.
- [12] Taemin Kim, Jihwan Woo, and In-So Kweon, Probabilistic matching of lines for their homography, *International Conference on Image Processing*, 2009.
- [13] Bin Fan, Fuchao Wu, and Zhanyi Hu Robust line matching through line-point invariants, *Pattern Recognition*, vol.45, pp. 794-805, 2012.
- [14] Qi Jia, Xinkai Gao, Xin Fan, Zhongxuan Luo, Haojie Li, and Ziyao Chen, Novel Coplanar Line-Points Invariants for Robust Line Matching Across Views, *European Conference on Computer Vision*, 2016.
- [15] Lu Wang, Ulrich Neumann, and Suya You, Wide-baseline image matching using Line Signatures, *International Conference on Computer Vision*, 2009.
- [16] J. López, R. Santos, X.R. Fdez-Vidal, X.M. Pardo, Two-view Line Matching Algorithm Based on Context and Appearance in Low-textured Images, *Pattern Recogn.*, vol. 48, pp. 2164-1484, Jul. 2015.
- [17] Niko Sunderhauf, Sareh Shirazi, Adam Jacobson, Feras Dayoub, Edward Pepperell, Ben UPIroft, and Michael Milford, Place recognition with ConvNet land-

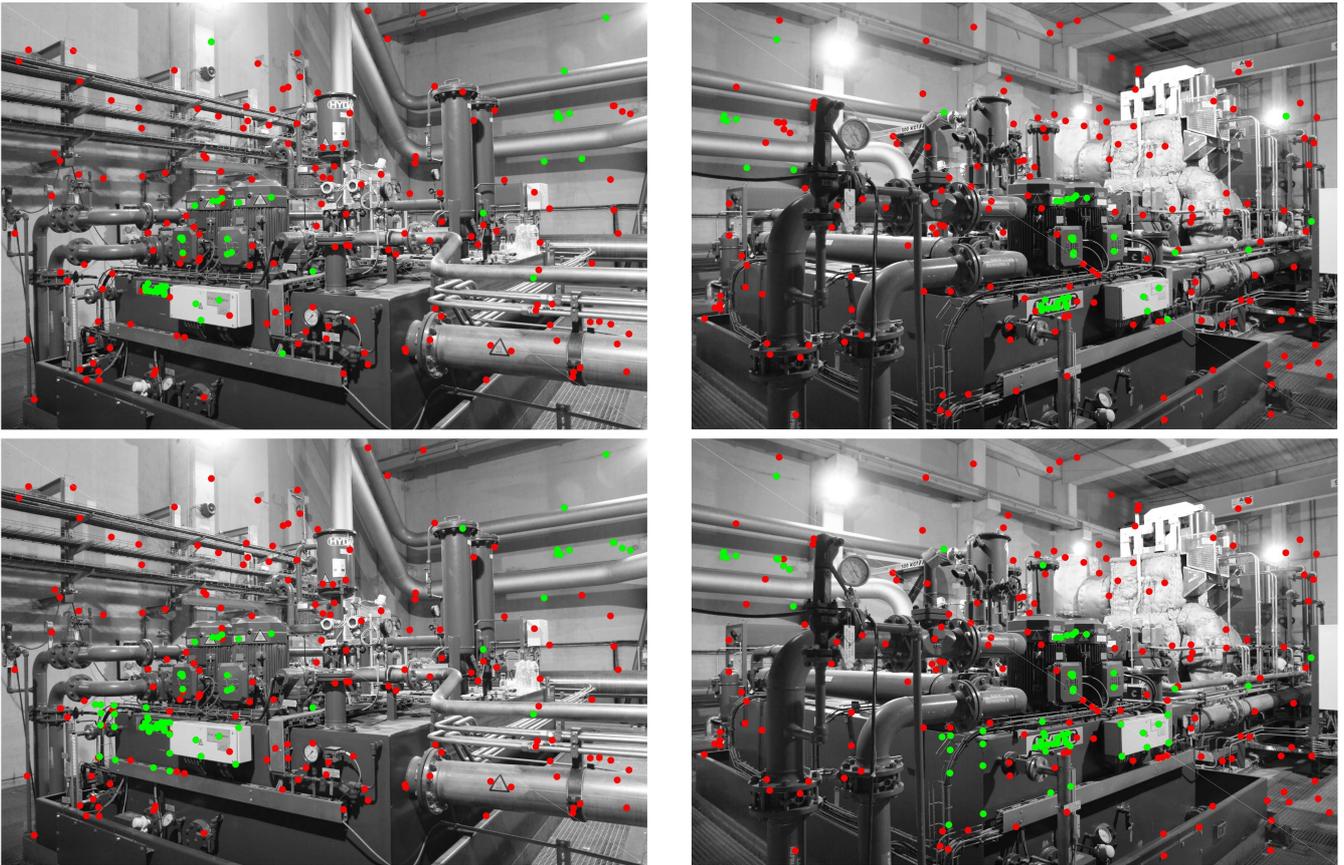


FIGURE 4 – Exemple d’images de test pour lesquelles les inliers définis par ORSA (en vert) sont plus nombreux parmi nos correspondances (Ligne 2) que parmi les correspondances entre LIFT (Ligne 1).

marks : Viewpoint-robust, condition-robust, training-free, *Robotics : Science and Systems*, 2015.

- [18] S. M. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. D. Cox, P. I. Corke, M. J. Milford, Visual Place Recognition : A Survey, *IEEE Trans. Robotics*, vol. 32, pp. 1-19, 2016.
- [19] Larry Zitnick, and Piotr Dollar, Edge Boxes : Locating Object Proposals from Edges, *European Conference on Computer Vision*, 2014.
- [20] G. Simon, A. Fond, and M.-O. Berger, A Simple and Effective Method to Detect Orthogonal Vanishing Points in Uncalibrated Images of Man-Made Environments, *Eurographics*, 2016.
- [21] Rafael Grompone von Gioi, Jérémie Jakubowicz, Jean-Michel Morel, and Gregory Randall, LSD : a Line Segment Detector, *Image Processing On Line*, vol. 2, pp. 35-55, 2012.
- [22] Lionel Moisan, Pierre Moulon, and Pascal Monasse, Fundamental Matrix of a Stereo Pair, with A Contrario Elimination of Outliers, *Image Processing On Line*, vol. 6, pp. 89-113, 2016.
- [23] Amir R Zamir, Tilman Wekel, Pulkit Agrawal, Colin Wei, Jitendra Malik, and Silvio Savarese, Generic 3d

representation via pose estimation and matching, *European Conference on Computer Vision*, 2016.

- [24] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, *International Conference on Neural Information Processing Systems*, 2012.