

Faisceau de parallaxe : estimation du mouvement de caméra robuste aux scènes planes

Martin Rebert^{1,2} David Monnin¹ Stéphane Bazeille² Christophe Cudel²

¹ ISL (Institut franco-allemand de recherches de Saint-Louis), Saint Louis, France.

² IRIMAS (Institut de Recherche en Informatique, Mathématiques, Automatique et Signal), EA 7499, Université de Haute-Alsace, Mulhouse, France.

email: martin.rebert@uha.fr

Résumé

Aujourd'hui les véhicules autonomes sont équipés de nombreux capteurs extéroceptifs (caméra passive, caméra lumière structurée, Lidar, ultrasons...). Ces capteurs permettent de détecter les obstacles, de localiser le véhicule ou bien de cartographier son environnement. Malgré cela, le capteur GPS leur est souvent préféré pour la localisation car son utilisation est très simple, il ne nécessite aucun calcul à bord et ses données sont précises et robustes. Lorsque la localisation GPS est imprécise ou absente (forêt, tunnel, zone non couverte, problème électronique) une méthode alternative de localisation est pourtant primordiale pour continuer à naviguer en attendant de retrouver un signal GPS correct. Dans cet article, nous présentons une nouvelle méthode permettant l'estimation du déplacement d'un véhicule à partir d'une caméra monoculaire et robuste aux environnements plans et peu structurés.

Mots-Clefs

Mouvement de caméra, homographie, parallaxe, géométrie épipolaire, scènes planes.

Abstract

Autonomous vehicles are usually equipped with sensors allowing the perception of the vehicle surroundings (camera, structured light, LiDAR, ultrasonic sensor...). With these sensors the vehicle can detect obstacles, localize itself or map its environment, but the GPS receiver is often preferred for its simplicity of use and localization accuracy. However, when the GPS signal is absent or imprecise due to a forest, a tunnel or a jamming signal for example, another localization method is needed. In this paper, we present a new method for the estimation of the vehicle motion from a monocular camera. The method is designed to be robust to strong planar structures often present in scenes resulting from low structured environments.

Keywords

Camera motion, homography, parallax, epipolar geometry, planar scenes.

1 Introduction

1.1 Contexte

La navigation autonome des robots a été très étudiée ces dernières années et ses domaines d'applications sont de plus en plus nombreux. Dans des environnements bien contrôlés, les robots sont capables depuis quelques années déjà de tondre la pelouse ou de nettoyer l'intérieur d'une maison de manière complètement autonome. Plus récemment, des véhicules sont capables de naviguer de manière autonome en ville en Californie. Dans le domaine militaire, la navigation autonome est utilisée pour la logistique ou l'ouverture d'itinéraire par exemple. Dans cet article, nous nous intéressons à la navigation autonome d'un véhicule sur route en environnement ouvert, robuste aux défaillances ou imprécisions de localisation du système de localisation GPS. En effet, dans de nombreux cas (zone urbaine, forêt, défaut matériel) nous ne disposons pas d'une localisation GPS assez précise pour permettre au véhicule de naviguer de manière autonome. Il est donc nécessaire d'obtenir à partir d'un autre capteur (dans notre cas une caméra monoculaire), une estimation correcte du déplacement du robot (odométrie visuelle). Dans la littérature de nombreuses méthodes utilisent des données Lidar ou les données inertielles pour localiser le robot. Mais dans notre cas, puisque nous disposons dans la plupart des cas de scènes planes peu structurées (route dégagée) la localisation par Lidar n'est pas efficace.

Nous avons donc préféré l'utilisation d'une caméra pour son bas coût et sa richesse d'information. Notre caméra est inclinée vers le sol pour maximiser la vue sur le sol puisque les obstacles sont aussi détectés avec la caméra et sont en général placés au sol. Cette position de la caméra est adéquate pour l'odométrie visuelle puisque la majorité des informations visuelles se situe en dessous de la ligne d'horizon et que cette position évite de se retrouver éblouie par le soleil situé en face de la caméra.

L'odométrie visuelle est une technique permettant d'estimer le déplacement d'un véhicule en mouvement à partir d'images enregistrées par caméra embarquée sur le véhi-

culé. Pour résoudre ce problème de nombreuses méthodes existent : le Bundle ajustement est la technique la plus utilisée [1, 2, 10, 12]. Cette méthode repose sur une initialisation utilisant la géométrie épipolaire entre deux vues d'une même scène [2, 10, 12]. Ref. [2, 12] estime la matrice essentielle grâce à la méthode des 5-points [11]. Ref. [10] utilise quant à elle la méthode des 8 points pour calculer la matrice fondamentale [6]. Dans nos conditions, le plan du sol est souvent dominant dans les images (il représente plus de la moitié de l'image) nous cherchons donc à obtenir le déplacement de la caméra à partir d'une décomposition de la matrice d'homographie en faisant une hypothèse sur le plan dominant de la scène pour choisir la décomposition correcte. Cependant, quand il y a présence d'une structure verticale dans la scène, un mur par exemple, cette hypothèse est contredite et l'estimation est erronée. Des exemples d'images nous amenant à ce problème sont présentées dans la Figure 1 (deuxième ligne). Pour éviter ce problème, la géométrie épipolaire peut être utilisée. Dans ce cas les matrices essentielles ou fondamentales doivent être estimées. Dans ce cas particulier, une scène plane est problématique car elle n'apporte pas assez d'information pour obtenir de manière unique la géométrie épipolaire [6]. Nous avons donc cherché à développer de nouvelles méthodes adaptées aux scènes planes que nous rencontrons de manière récurrente dans nos cas d'études.

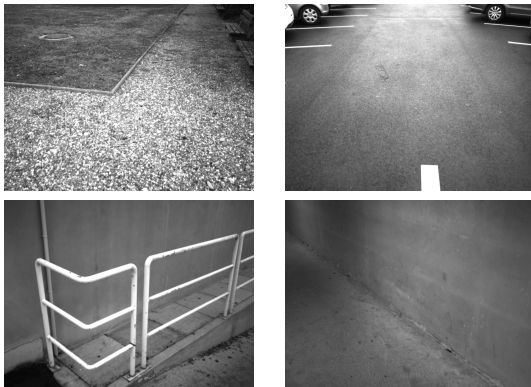


FIGURE 1 – Exemples de scènes rencontrées. **Ligne du haut** : scènes quasiment planes où une homographie est nécessaire à l'estimation du mouvement. **Ligne du bas** : scènes où des structures planes sont présentes mais une homographie ne suffit à décrire l'intégralité de la scène.

1.2 Travaux connexes

Plusieurs méthodes existent pour retrouver la géométrie épipolaire en exploitant la dégénérescence présente dans la scène. Ces méthodes se concentrent généralement sur les scènes contenant plusieurs plans dominants dans les images (voir Fig. 1 image en bas à droite). Parmi ces méthodes certaines basent leurs estimations sur une homographie et d'autres sur plusieurs. La méthode la plus simple a été proposée par [16] qui obtient une estimation robuste à partir de deux homographies extraites d'une même paire

d'image. Ensuite les inliers en commun sont utilisés pour estimer la matrice fondamentale. On peut également utiliser l'équation de compatibilité entre une homographie et la matrice fondamentale proposée par [6]. Cette dernière nécessite par contre une normalisation [17] pour être stable [7]. À noter que deux homographies indépendantes ne sont pas toujours compatibles avec la géométrie épipolaire [15] ce qui peut conduire à des estimations de mouvements erronées i.e. deux homographies peuvent faire apparaître différents groupes de lignes épipolaires. Toujours à partir de deux homographies, on peut construire une homologie qui transfère l'épipole sur lui-même [13]. Cet épipole correspond alors au vecteur propre associé à l'unique valeur propre de l'homologie, les deux autres étant égales. Cependant, avec des données réelles bruitées, rien ne garantit d'avoir des valeurs propres égales et choisir la valeur propre unique peut être un problème.

Dans le cas où il n'est pas possible d'obtenir deux homographies indépendantes, e.g. dans les cas où il n'y qu'un plan dans la scène, la géométrie épipolaire peut toujours être extraite à partir d'une unique homographie et deux points n'appartenant pas au plan de l'homographie. Dans Ref. [13], les auteurs détaillent les différentes méthodes pour obtenir la matrice fondamentale à partir d'une unique homographie. Cependant rien n'est dit sur les détections et la suppression des points aberrants et sur les valeurs initiales de minimisation. Pour conclure, l'épipole est nécessaire pour obtenir la matrice fondamentale à partir d'une unique homographie. L'épipole peut être obtenu en utilisant la méthode du *plane-and-parallax* [6, 14, 17], bien que cette méthode présente plusieurs défauts.

Dans la suite du papier, nous présenterons d'abord notre chaîne de traitement pour estimer le déplacement de notre robot et introduirons un peu plus en détails la méthode du *plane-and-parallax* car nous partons de ses défauts pour développer notre nouvelle méthode. Celle-ci sera présentée dans un deuxième temps. Finalement nous amènerons la méthode d'évaluation utilisée et décrirons nos résultats.

2 Une méthode hybride : le faisceau de parallaxe

2.1 Aperçu du système

La méthode présentée fait partie de la chaîne de traitement présentée dans la Figure 2. Le système prend en entrées deux images successives d'une même scène, la caméra ayant bougé entre les deux images. Des points caractéristiques sont ensuite détectés et mis en correspondance entre les deux images. La détection est faite par un détecteur de Harris modifié [4] qui permet d'obtenir une répartition uniforme des coins malgré des surfaces homogènes comme celles de la Figure 1. La mise en correspondance à l'aide de descripteurs BRIEF et d'une distance de Hamming. Les correspondances sont utilisées pour estimer une homographie grâce à un algorithme RANSAC modifié [9], qui évite les configurations aberrantes. Si une large majorité de cor-

respondances, autour de 90 %, sont incluses dans le support de l’homographie alors la scène est considérée comme plane. Le mouvement est alors estimé par la décomposition de la matrice d’homographie H [8]. Dans notre application, la désambiguïsation est faite en supposant le plan homographique comme étant celui de sol sur lequel roule le robot.

Si la scène ne peut pas être considérée comme plane parce que le modèle homographique ne permet pas de la décrire dans sa globalité, on peut alors s’intéresser à la géométrie épipolaire. À ce stade, il est judicieux de réutiliser au maximum les efforts de calcul déjà faits et d’utiliser une méthode *plane-and-parallax* pour réutiliser l’homographie estimée. Cette méthode permet d’obtenir la position de l’épipole qui, combinée à un plan, permet d’obtenir la matrice fondamentale F . On propose ici un nouveau principe pour utiliser cette méthode. Une fois la matrice fondamentale obtenue, on peut estimer le mouvement de la caméra en décomposant la matrice essentielle E qui a été calculée à partir de F et la matrice K contenant les paramètres intrinsèques de la caméra, $E = K^T F K$. Lorsqu’une première estimation du mouvement est obtenue, on l’affine à l’aide d’une méthode des moindres carrés.

2.2 Faisceau de parallaxe

La formulation de la méthode *plane-and-parallax* présentée dans [6] permet d’estimer la matrice fondamentale à partir de correspondances de points. Les correspondances ont la particularité de correspondre en partie à des points coplanaires dans la scène. S’ils sont coplanaires, ils sont transférés d’une image à l’autre par une homographie H . On a donc pour des correspondances x_i et x'_i , la relation $x'_i = Hx_i$. S’ils ne correspondent pas à des points coplanaires dans la scène, on a la relation $x'_i = Hx_i + \lambda e'$, où $\lambda \in \mathbb{R}$ et e' est l’épipole de l’image. $\lambda e'$ est une parallaxe virtuelle due à l’erreur de la reprojection homographique. Les points x'_i , Hx_i et e' ont la propriété d’être alignés, on peut donc remonter à la position de e' en cherchant l’intersection de toutes les parallaxes. Une fois e' estimé, on peut calculer F [6],

$$F = [e']_{\times} H. \quad (1)$$

Approche standard : estimer e' à partir des lignes de parallaxe. L’application du principe précédemment énoncé à des données réelles n’est pas simple. Des imprécisions existent sur la position des points d’intérêt du détecteur de Harris et sur la matrice d’homographie estimée. Ces imprécisions ont pour effet de perturber l’intersection des lignes de parallaxe, l’intersection n’est plus commune, on a à la place une multitude d’intersections de paires de lignes dans un voisinage. La présence de correspondances aberrantes complique également. Une partie des correspondances, celles qui sont coplanaires, sont décrites par l’estimation de H . Les données générant de la parallaxe, c’est-à-dire les outliers de l’homographie. Ils représentent ici les points hors du plan homographique ainsi que les mauvaises correspondances. Trouver la position de l’épi-

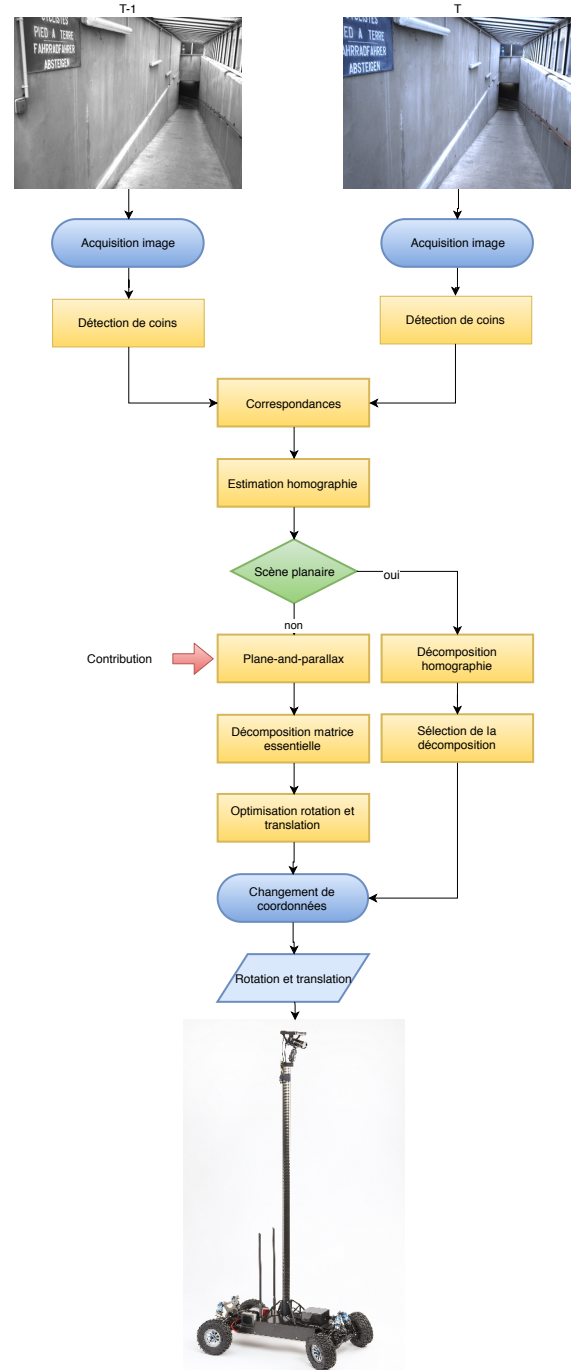


FIGURE 2 – Chaîne de traitement présente dans le système. Elle permet d’estimer le mouvement de la caméra de façon robuste lorsque peu d’information 3D sont présentes dans la scène.

pole e' revient alors à chercher le point où la plupart des lignes semblent converger. Le bruit sur les bonnes correspondances et les mauvaises correspondances ne permettent pas d’identifier rapidement un point d’intersection unique. Mis à part la description du principe donné dans [6], nous n’avons pas trouvé d’algorithme détaillé pour résoudre le

Algorithm 1 Epipole à partir des lignes de parallaxe

```
while  $iteration < iterationMax$  do
  Sélectionner 2 lignes
  Calculer leur point d'intersection
  Sélectionner tous les lignes passant dans le
  voisinage du point d'intersection
  Associer au point d'intersection un score corres-
  pondant au nombre de lignes dans son voisinage
  if  $score > plusGrosScoreTrouvé$  then
    Retenir le point d'intersection et les lignes dans
    son voisinage
  end if
end while
Appliquer la méthode des moindres-carrés pour trouver
le point le plus proche de toutes les lignes retenues en
utilisant le point d'intersection retenu comme point de
départ
```

problème. On en propose un dans l'algorithme 1, qui fait appel à des techniques standards.

Après avoir essayé la méthode, elle présente plusieurs limites. Premièrement, l'épipole peut ne pas être estimé correctement lorsqu'il y a un nombre pair de lignes qui semblent se recouper dans deux points distincts. Deuxièmement, le bruit lié à la détection des points d'intérêt et à l'estimation de H n'influence pas de la même façon toutes les lignes de parallaxe. Des lignes générées à partir de points x'_i et Hx_i très proches vont avoir leur orientation très perturbée par le bruit. À l'inverse pour des points plus éloignés, le bruit va avoir moins d'influence sur l'orientation de la ligne. En revanche toutes les lignes sont considérées de façon identique par la méthode. Troisièmement, la méthode nécessite une notion de voisinage autour du point d'intersection. À notre expérience, les seuils de voisinage sont généralement entre 10 pixels et 100 pixels. Affiner entre ces deux limites est dépendant de la scène.

Faisceau de parallaxe : estimation robuste de l'épipole. Pour palier les limitations énoncées, on propose un modèle de faisceau de parallaxe permettant de considérer que

1. la position des points d'intérêt x'_i est imprécise,
2. l'estimation de l'homographie est imprécise donc la position de Hx_i aussi,
3. les points proches du plan homographique, donc de faible parallaxe, sont moins fiables que ceux ayant une parallaxe plus grande.

Si la position des points d'intérêt est corrompue par du bruit, on peut considérer que la vraie position se situe dans le voisinage du point détecté. On considère une zone uniforme autour du point détecté dans laquelle il est probable que le vrai point soit. La zone est représentée dans la Figure 3(a). Elle est définie par son centre et son rayon r . La méthode fonctionnant sur des images proches en termes de point de vue, on considère que l'homographie déforme suffisamment peu pour que la zone autour de Hx_i soit tou-

jours circulaire. Les zones d'incertitude autour de x'_i et de Hx_i définissent un ensemble de lignes de parallaxe possibles, Figure 3. L'orientation de toutes les lignes de parallaxe est contenue entre deux orientations possibles. Ces orientations sont définies par les tangentes croisées aux deux cercles. La zone entre les deux tangentes contenant les deux cercles est appelée *faisceau de parallaxe*. Le faisceau de parallaxe contient toutes les lignes de parallaxe qui auraient pu être obtenues à partir des points d'intérêt bruités. En comparant la Figure 3(a) et la Figure 3(b) on remarque que la taille de la parallaxe influence l'ouverture du faisceau de parallaxe. À bruit constant, des points proches du plan homographique contiennent donc une information moins fiable, qui se traduit par une ouverture plus importante du faisceau.

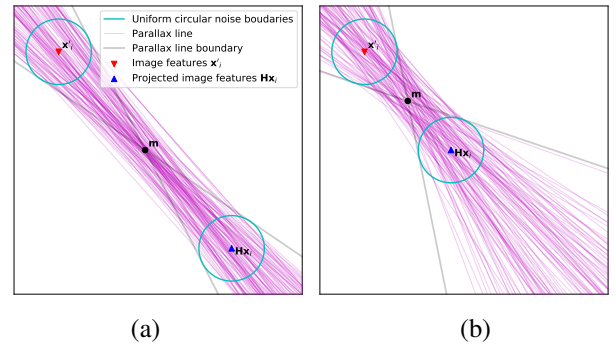


FIGURE 3 – Un ensemble de lignes de parallaxe générées à partir de points pris à l'intérieur des cercles. Les lignes noires représentent les limites d'orientation des lignes de parallaxe. (a) La distance entre x'_i et Hx_i est de 4 pixels. (b) La distance est de 2 pixels, c'est-à-dire que les points sont plus proches du plan homographique.

Le faisceau de parallaxe peut être complètement défini par les points m , t_1 et t_2 de la Figure 4(a). m est le milieu du segment reliant x'_i et Hx_i , t_1 et t_2 sont les points tangents au cercle dont les tangentes passent par m . Pour obtenir t_1 et t_2 , on déplace l'origine en Hx_i . Dans ce système de coordonnées, on a $m = (x_m, y_m)$ et $t_1 = (x, y)$. Comme t_1 se trouve sur le cercle de centre Hx_i et de rayon r , on peut écrire

$$r^2 = x^2 + y^2. \quad (2)$$

En utilisant la relation de Pythagore $l^2 = r^2 + (x - x_m)^2 + (y - y_m)^2$, la distance de m à l'origine Hx_i $x_m^2 + y_m^2 = l^2$ et Équation 2, on obtient :

$$x = \frac{r^2 - y_m y}{x_m}. \quad (3)$$

On substitue Équation 3 dans Équation 2 pour obtenir un polynôme du second degré en y et dont les racines sont les coordonnées y de t_1 et t_2 si le point m se situe en dehors

du cercle, i.e. $l > r$. On a donc

$$0 = y^2 l^2 - y(2r^2 y_m) + r^4 - x_m^2 r^2$$

$$y = \frac{r \left(r y_m \pm \sqrt{(r y_m)^2 + (l x_m)^2 - (l r)^2} \right)}{l^2}. \quad (4)$$

La coordonnée x de t_1 et t_2 peut être obtenue grâce à Équation 3. Finalement t_1 et t_2 sont replacés dans le repère global en leur ajoutant Hx_i . Les lignes (m, t_1) et (m, t_2) forment l’enveloppe du faisceau de parallaxe.

Pour estimer la position de l’épipole à partir d’un ensemble de faisceaux de parallaxe, on cherche la zone où la majorité des faisceaux se recouvrent. On fait donc l’intersection de toutes les enveloppes de faisceau, ce qui génère un ensemble de points d’intersection, Figure 5. Les points d’intersection sont placés sur les sommets de polygones contenant à chaque fois un nombre fixe de faisceaux. On cherche donc le polygone contenant le plus de faisceaux, ce qui revient à chercher le polygone contenu dans le plus de faisceaux et donc les sommets contenus dans le plus de faisceaux. Pour tester si un point quelconque de l’image p_i est contenu dans le faisceau, il suffit d’effectuer le changement de repère vers le repère formé par (m, \vec{u}, \vec{v}) . Dans ce repère, un point dans le faisceau aura des coordonnées de même signe, et de signes opposés s’il se trouve à l’extérieur du faisceau. Une fois le polygone identifié, la position de l’épipole est obtenue en prenant le barycentre des sommets le composant.

Jusqu’ici nous avons défini notre modèle de faisceau de parallaxe et la façon de retrouver l’épipole à partir de ce dernier. La méthode fonctionne également en présence d’outliers, on retiendra alors comme inliers les faisceaux qui contiennent l’épipole, les autres étant considérés comme outliers. Par ailleurs, la Figure 5 a été générée en utilisant différentes tailles de parallaxe pour illustrer l’intérêt du faisceau de parallaxe. De petites parallaxes vont générer des faisceaux relativement ouverts et donc être moins discriminants sur la position de l’épipole. À l’inverse des grandes parallaxes vont contribuer de façon plus significative à l’estimation de l’épipole du fait de leur faisceau plus étroit. Finalement, la Figure 6 propose une comparaison entre l’approche classique qui utilise les lignes de parallaxe et l’approche utilisant le faisceau de parallaxe pour estimer la position de l’épipole. On voit que l’estimation par le faisceau s’en sort mieux que l’intersection de lignes. Une façon d’améliorer l’intersection de lignes consisterait à enlever les parallaxes trop faibles, mais ce procédé enlèverait des correctes également, là où notre méthode les utilise toutes mais avec des influences différentes.

3 Évaluation de la méthode proposée

3.1 Conditions d’évaluation

On évalue la méthode du faisceau de parallaxe dans des scènes synthétiques présentant des parties planes importantes et dans quatre configurations différentes. Les scènes

comportent deux plans perpendiculaires, un plan horizontal correspondant au sol et un plan vertical correspondant à un mur. On se place dans les cas suivants :

- à 2.5 m du mur, l’image est quasiment entièrement occupée par le mur, seul un morceau du sol est visible dans le bas de l’image.
- à 5 m du mur, le mur et le sol sont représentés équitablement dans l’image.
- à 10 m du mur, le sol devient le plan dominant de l’image, mais le mur est toujours visible à hauteur de 25 %.
- à 15 m du mur, le sol devient très dominant dans l’image et seule une petite partie du mur est visible.

Pour chaque configuration, on réalise 500 mouvements. Les mouvements ont été pris des fichiers de vérité terrain du KITTI Vision Benchmark Suite [3], de sorte à ce qu’ils soient cohérents avec le mouvement d’un véhicule à quatre roues dont deux directrices. Nous plaçons dans les conditions de notre véhicule, Figure 2, avec une caméra placée à 1.6 m de haut, orientée de 20 degrés vers le bas, ayant une optique 8.5 mm et une image de 1280×960. La distance entre deux images varie entre 30 et 50 cm. Finalement les données d’entrée de l’estimation de mouvement sont bruitées avec un bruit blanc gaussien et corrompues avec 20 % d’outliers équivalents à de mauvaises correspondances. On travaille sur un total de 256 correspondances présentant une répartition uniforme dans l’image, telles qu’obtenues par [4].

Le but étant d’estimer le mouvement de la caméra entre deux prises de vues, on mesure l’erreur d’estimation sur la translation et la rotation obtenues après avoir décomposé la matrice essentielle. En notant R_{vt} , la rotation issue de la vérité terrain et R la rotation issue de l’estimation, on calcule l’erreur sur la rotation ϵ_R en prenant l’angle de la rotation résiduelle $R_{diff} = R^T R_{vt}$.

$$\epsilon_R = \arccos \left(\frac{\text{tr}(R_{diff}) - 1}{2} \right)$$

Sans information supplémentaire, la translation obtenue à partir de deux vues n’est pas à l’échelle, on comparera donc les orientations des translations entre celles estimées et la vérité terrain. Ainsi pour t la translation estimée et t_{vt} la translation issue de la vérité terrain on a une erreur ϵ_t

$$\epsilon_t = \arccos \left(\frac{t \cdot t_{vt}}{\|t\| \|t_{vt}\|} \right),$$

qui correspond à l’angle entre les deux vecteurs de translation.

La méthode est comparée avec l’algorithme des 5-points [11] et des 8-points [5] qui sont des méthodes classiques d’estimation de la géométrie épipolaire. On compare également à la méthode utilisant les lignes de parallaxe donnée dans Algorithm 1.

3.2 Résultats

Les résultats obtenus pour les quatre méthodes dans les conditions décrites à la section précédente sont présentés

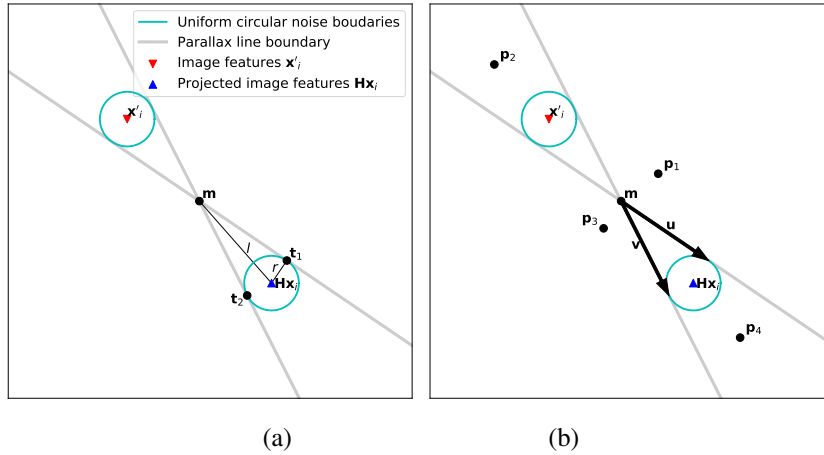


FIGURE 4 – (a) Calcul des lignes définissant le faisceau de parallaxe. (b) Les points p_i correspondent aux cas possibles pour déterminer si un point se trouve dans le faisceau.

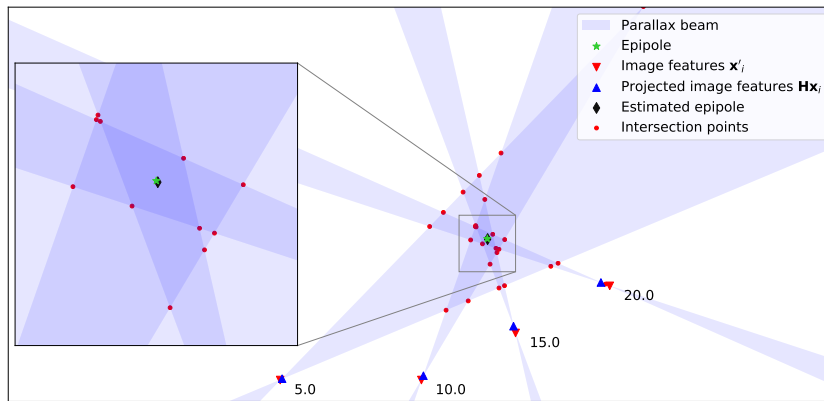


FIGURE 5 – L'intersection des enveloppes de faisceau génère un maillage de points d'intersection. L'épipole est estimé comme étant le point moyen des sommets définissant la zone de recouvrement maximum. Les valeurs indiquées sur la figure correspondent à la taille de la parallaxe donnée en pixels.

dans la Figure 7 et la Figure 8. Nous proposons l'erreur moyenne et l'erreur maximum comme critères de comparaison. L'erreur moyenne permet d'obtenir une performance globale des méthodes comparées et l'erreur maximum permet d'attester la robustesse des méthodes dans les conditions d'évaluation.

En regardant l'erreur moyenne on se rend compte que l'algorithme des 8-points est sensible à la dégénérescence de la scène et il n'estime correctement le mouvement que dans le cas où deux plans se partagent l'image, ce qui représente la dégénérescence la plus faible. La méthode utilisant les lignes de parallaxe résiste jusqu'à une dégénérescence plane plus présente (à 10 m) du mur. En cas de forte présence plane, la méthode échoue. L'algorithme des 5-points et la méthode des faisceaux de parallaxe présentent des erreurs moyennes comparables avec un léger avantage pour la méthode des 5-points sur l'estimation de la translation et inversement pour l'estimation de la rotation.

L'erreur maximum sur l'estimation du mouvement permet

en revanche une meilleure discrimination des méthodes. La méthode des 8-points présente toujours les mêmes cas de réussite et d'échec. La méthode des lignes de parallaxe échoue parfois complètement à estimer la translation correctement en la présence de deux plans équitablement distribués dans l'image. Ce cas avait déjà relevé dans [17]. L'algorithme des 5-points est pour sa part robuste lorsque deux plans sont équitablement présents et lorsque le mouvement de caméra est perpendiculaire à la surface plane [11]. Il présente des cas d'échec lorsque le mouvement n'est pas perpendiculaire à la surface plane comme c'est le cas à 10 m et à 15 m.

Finalement seule la méthode faisant appel aux faisceaux de parallaxe permet d'obtenir une estimation du mouvement de la caméra dans tous les cas. Les 30 degrés d'erreur maximum obtenus sur la translation à 15 m sont à relativiser car dans ce cas l'homographie aurait suffi à estimer le mouvement d'après le procédé de la Figure 2. On notera que pour toutes les méthodes l'estimation de la translation

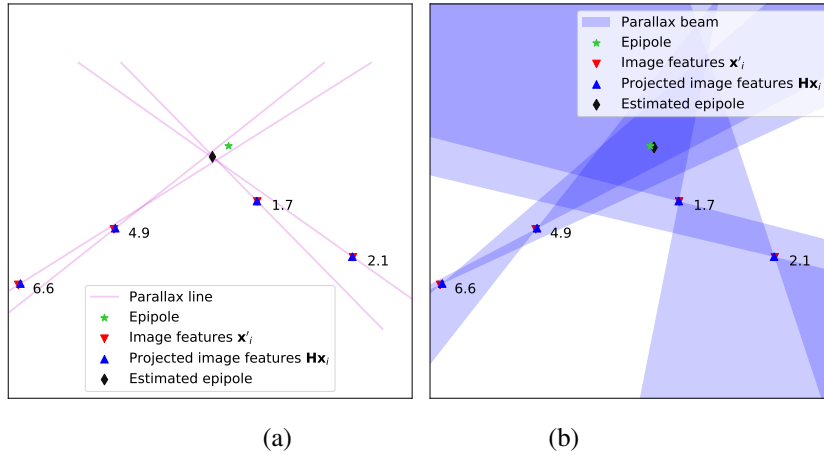


FIGURE 6 – Les points d’intérêt sont perturbés par un bruit blanc gaussien d’écart-type $\sigma = 0.17$. (a) L’approche utilisant l’intersection des lignes de parallaxe ne permet pas d’estimer l’épipole correctement. (b) L’approche utilisant le modèle du faisceau de parallaxe permet d’obtenir une estimation plus précise de la position de l’épipole. Les valeurs sur la figure correspondent à la taille des parallaxes en pixels.

est moins précise que la rotation.

4 Conclusion

La localisation grâce à la vision dans des environnements peu structurés et les environnements où les structures planes sont fréquentes est compliquée comme le montrent les nombreux travaux de recherches dans ce domaine. L’estimation de mouvement peut se faire à l’aide de deux modèles, la géométrie épipolaire ou l’homographie. Le premier est utilisé dans des scènes génériques et le second est dédié aux structures planes. En utilisation réelle les scènes peuvent ne pas être complètement planes, ni parfaitement génériques on parle de scènes quasi-planes. Ces scènes peuvent mettre en difficulté l’un ou l’autre des modèles. Nous avons donc proposé un modèle hybride qui utilise l’homographie tant qu’il le peut et qui est capable de remonter de façon robuste à la géométrie épipolaire si c’est nécessaire pour estimer de manière robuste les déplacements d’un véhicule. Inspirée de la méthode *plane-and-parallax*, la méthode du faisceau de parallaxe est adaptée aux données bruitées et corrompues. Son efficacité a été illustrée sur des scènes où la dégénérescence était faible, moyenne et forte.

Des premiers essais prometteurs sur le benchmark KITTI laissent entrevoir une méthode utilisable également sans la présence d’une structure plane dominante. Néanmoins, ces résultats seront à approfondir et à confirmer par la suite. Cette nouvelle méthode présente une limite d’utilisation si le mouvement de la caméra est parallèle au plan image car on viendrait alors chercher un épipole, qui serait à l’infini, en superposant des faisceaux *dans le voisinage de l’infini*. Néanmoins, ce cas est peu réaliste dans le cas d’un véhicule à roues avec l’essieu avant directeur, car c’est la translation perpendiculaire au plan image qui prédomine. Par la suite nous avons prévu de tester cet algorithme en

situation réelle et en embarqué sur le véhicule autonome de l’IRIMAS : ARTEMIPS.

Références

- [1] J. Engel, V. Koltun, and D. Cremers. Direct Sparse Odometry. *arXiv :1607.02565 [cs]*, July 2016. arXiv : 1607.02565.
- [2] C. Forster, M. Pizzoli, and D. Scaramuzza. SVO : Fast semi-direct monocular visual odometry. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 15–22, May 2014.
- [3] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for Autonomous Driving ? The KITTI Vision Benchmark Suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [4] L. Gond, D. Monnin, and A. Schneider. Optimized feature-detection for on-board vision-based surveillance. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XVII*, volume 8357, page 83571N. International Society for Optics and Photonics, May 2012.
- [5] R. I. Hartley. In defence of the 8-point algorithm. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1064–1070, June 1995.
- [6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN : 0521540518, second edition, 2004.
- [7] Q. T. Luong and O. D. Faugeras. Determining the fundamental matrix with planes : instability and new algorithms. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 489–494, June 1993.

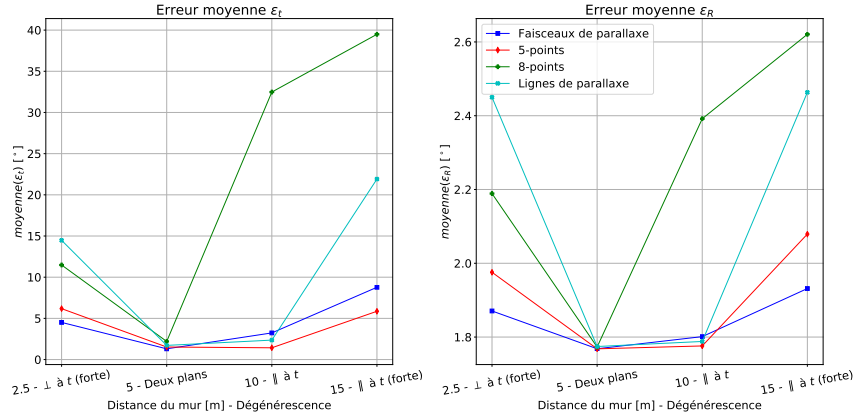


FIGURE 7 – Erreurs moyennes obtenues pour l’estimation de la translation et de la rotation effectuée par la caméra.

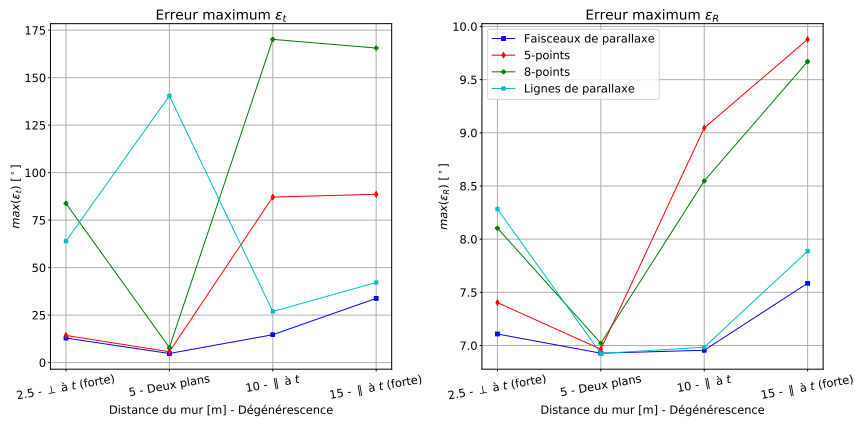


FIGURE 8 – Erreurs maximum obtenues pour l’estimation de la translation et de la rotation effectuée par la caméra.

[8] E. Malis and M. Vargas. Deeper understanding of the homography decomposition for vision-based control. Research Report RR-6303, INRIA, 2007.

[9] D. Monnin, E. Bieber, G. Schmitt, and A. Schneider. An effective rigidity constraint for improving RANSAC in homography estimation. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 203–214. Springer, 2010.

[10] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos. ORB-SLAM : a Versatile and Accurate Monocular SLAM System. *IEEE Transactions on Robotics*, 31(5) :1147–1163, Oct. 2015. arXiv : 1502.00956.

[11] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE transactions on pattern analysis and machine intelligence*, 26(6) :756–770, 2004.

[12] E. Royer, M. Lhuillier, M. Dhôme, and J.-M. Lavest. Monocular Vision for Mobile Robot Localization and Autonomous Navigation. *International Journal of Computer Vision*, 74(3) :237–260, 2007.

[13] D. Sinclair, H. Christensen, and C. Rothwell. Using the relation between a plane projectivity and the fun-

damental matrix. In *Proc. SCIA*, pages 181–188, 1995.

[14] R. Szeliski and P. H. Torr. Geometrically constrained structure from motion : Points on planes. In *European Workshop on 3D Structure from Multiple Images of Large-Scale Environments*, pages 171–186. Springer, 1998.

[15] Z. L. Szpak, W. Chojnacki, and A. v. d. Hengel. Robust multiple homography estimation : An ill-solved problem. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2132–2141, June 2015.

[16] E. Vincent and R. Laganière. Detecting planar homographies in an image pair. In *Image and Signal Processing and Analysis, 2001. ISPA 2001. Proceedings of the 2nd International Symposium on*, pages 182–187. IEEE, 2001.

[17] Y. Zhou, L. Kneip, and H. Li. A Revisit of Methods for Determining the Fundamental Matrix with Planes. In *Digital Image Computing : Techniques and Applications (DICTA), 2015 International Conference on*, pages 1–7. IEEE, 2015.