

Comparaison de décompositions de la matrice homographique et essentielle pour l'estimation du mouvement de caméra

Martin Rebert^{1,2}

David Monnin¹

Sophie Kohler²

Christophe Cudel²

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

² Laboratoire MIPS (EA2332), Université de Haute-Alsace (UHA)

Institut franco-allemand de recherches de Saint-Louis
5 rue du Général Cassagnou
68300 Saint-Louis - France
martin.rebert@isl.eu

Résumé

Dans le but d'améliorer un système d'odométrie visuelle basé sur la décomposition d'une matrice d'homographie, l'utilisation de la géométrie épipolaire est envisagée. Elle présente l'avantage d'utiliser l'information 3D présente dans la scène. On se propose de comparer l'estimation de la transformation rigide liant deux positions de caméras, à partir de la décomposition de la matrice d'homographie et de la matrice essentielle. La comparaison est réalisée sur des scènes simulées ainsi que sur des scènes réelles. On évalue la précision de la transformation sur sa composante de translation et rotation. La précision est dans tous les cas au moins aussi bonne et la robustesse est améliorée en utilisant la matrice essentielle.

Mots Clef

Odométrie visuelle, matrice essentielle, homographie, transformation rigide

Abstract

The epipolar geometry is one of the candidates to improve a visual odometry system so far based on the decomposition of an homography matrix. This geometry enables the use of the 3D information present in the scene. The rigid transformation between the two camera positions can be obtained by the decomposition of the homography or the essential matrix, these decompositions are compared in this paper. The comparison is made on synthetic scenes and on real images as well. The precision of the transformation is evaluated on the translation and the rotation part. The decomposition of the essential matrix appears to be at least as precise and more robust than the one of the homography matrix.

Keywords

Visual odometry, essential matrix, homography, rigid transformation

1 Introduction

L'étude se fait dans le contexte d'un projet de détection de changement. Le système présenté dans [14] permet la détection d'IEDs (Improvised Explosive Devices) sur un itinéraire parcouru par un véhicule à deux dates distinctes. Lors du premier passage, l'itinéraire est établi comme étant sûr et sert de référence. Tout changement par rapport à cet état représente une menace potentielle. L'image actuelle et l'image de référence sont comparées par un opérateur humain après un recalage homographique. L'homographie a été choisie car les scènes sont majoritairement planes du fait de l'orientation de la caméra, et les menaces situées au niveau du plan du sol, voir Figure 1. De plus, elle permet un recalage rapide des images en temps réel.

Néanmoins, l'intervention humaine, notamment pour la conduite du véhicule, ne permet pas de reproduire au mieux la trajectoire de référence et donc de réaliser de la détection de changement sur toute l'image. En effet un écart par rapport à la trajectoire de référence donne des champs de vue différents pour une même scène. De plus, plus l'écart par rapport à la trajectoire de référence est important, plus le phénomène de parallaxe est visible pour les structures sortant du plan de l'homographie après le recalage, ce qui induit une fatigue supplémentaire chez l'opérateur humain. Ainsi pour pallier ces défauts, il a été proposé de rendre la conduite du véhicule autonome. Bien que les images soient géo-référencées, l'imprécision du capteur GPS ainsi qu'un potentiel brouillage en application militaire ne permettent pas d'utiliser cette information de façon fiable. La navigation par l'image est alors indiquée pour rendre le système autonome sans ajout de matériel supplémentaire et donc sans surcoût. Un des pré-requis à la navigation visuelle est l'odométrie visuelle, c'est-à-dire d'estimer la trajectoire du véhicule à partir du flux d'images de la caméra pour établir l'itinéraire de référence et ensuite naviguer par rapport à celui-ci. Il est donc nécessaire d'estimer, avec le moins d'erreur possible, le mouvement de la caméra entre deux



(a) UGV



(b)



(c)



(d)



(e)

FIGURE 1 – (a) drone de test équipé du système de détection de changement. (b), (c), (d) et (e) sont des exemples d’images utilisées par l’UGV pour la navigation. Dans (b) et (c) le plan horizontal de la route est majoritaire. Dans (d) et (e) d’autres plans sont très présents dans l’image.

prises de vue.

1.1 Travaux antérieurs

Afin de respecter la contrainte du temps réel, l’approche utilisée jusqu’à maintenant privilégie l’utilisation d’un élément déjà calculé par le système, la matrice d’homographie. Celle-ci est décomposée afin d’obtenir la transformation rigide reliant les deux plans de caméra. Néanmoins cette technique utilise l’a priori que le plan dominant de l’image et celui utilisé par l’homographie est celui de la route. Or pour des applications indoor ou urbaines le plan dominant peut être un mur occupant une grande partie du champ de vision. Il a également été constaté que même lorsque le plan dominant est physiquement celui de la route, ce n’est pas toujours exactement celui utilisé par l’homographie. Cela conduit alors à des estimations erronées de la position du véhicule.

Différentes approches peuvent alors être envisagées pour estimer le déplacement de la caméra à partir de l’image. Changer la caméra monoculaire pour une paire stéréo permettrait de recalculer deux ensembles de points 3D triangulés à partir des images stéréo, tel que décrit dans [12]. Néanmoins, cette technique demande un changement de matériel et l’incertitude sur la profondeur des points triangulés a un impact important sur l’estimation du mouvement [17]. D’autres techniques fondées sur des correspondances 3D-2D permettent de s’affranchir de cette incertitude et de l’utilisation d’une paire stéréo. Si la position de points 3D est connue ainsi que leurs correspondances dans l’image, on peut estimer la position de la caméra grâce à une méthode de *perspective from n points*, dont le problème minimal requiert trois points pour être résolu et un quatrième pour lever l’ambiguïté [9]. Si les points 3D sont inconnus, l’utilisation de techniques de *Bundle Adjustment* [21] permet d’accéder aux paramètres intrinsèques et extrinsèques des caméras ainsi qu’à la position des points 3D en minimisant l’erreur de reprojection. La prise en compte du caractère *sparse* du problème et les progrès matériels en permettent l’utilisation en temps réel au moins depuis [2]. Le *bundle adjustment* présente toutefois l’inconvénient de nécessiter de bonnes conditions initiales afin de converger rapidement et vers le minimum global, ce qui nécessite souvent une première estimation ou un a priori sur la solution. Finalement, l’utilisation de la géométrie épipolaire et la décomposition de matrice essentielle [7] permettent d’estimer le mouvement à partir uniquement de correspondances entre points dans les plans images (2D-2D). Cette méthode présente l’avantage de ne pas nécessiter d’a priori sur le mouvement ou sur la position des points 3D. Elle peut être également utilisée pour initialiser des méthodes de *bundle adjustment* ou de *perspective from n points* en estimant un premier mouvement de caméra permettant de trianguler un premier ensemble de points 3D.

1.2 Contribution

Cette étude va comparer la décomposition de la matrice essentielle à la décomposition homographique utilisée afin d’en évaluer l’apport potentiel. Cette décomposition permet d’utiliser l’information 3D de la scène contrairement à l’homographie. La décomposition utilisée est celle présentée dans [7] qui est donnée par [18] comme étant le standard pour faire l’estimation du mouvement à partir des correspondances dans les images. On présentera dans un premier temps les deux décompositions, ainsi que la méthode utilisée pour juger de la qualité de la décomposition par rapport à la vérité terrain. Les différents scénarios de tests et les résultats associés seront ensuite présentés. En dernière partie, la comparaison portera sur des images réelles et enfin nous concluons sur le bénéfice à utiliser la décomposition basée sur la géométrie épipolaire.

2 Estimer la transformation rigide

2.1 À partir de l'homographie

Pour un ensemble de points x_i dans un plan projectif 2D et son ensemble correspondant x'_i dans un autre plan projectif 2D, l'homographie est la transformation projective transformant x_i en x'_i . Dans le cadre de ce document, les plans projectifs sont les plans images et les points sont des pixels de l'image. Dans le système [14], l'homographie est estimée de façon robuste d'après [13].

Par la suite l'homographie est décomposée selon [22]. La décomposition distingue trois cas en fonction de la multiplicité des valeurs singulières de la matrice d'homographie. Dans les cas où toutes les valeurs singulières n'ont pas une multiplicité égale à 1 (i.e. 2 ou 3), la décomposition est unique, mais dépendante du plan dominant de l'homographie. Les mouvements correspondants sont une rotation autour de l'origine puis une translation suivant la normale au plan pour une multiplicité égale à 2 et à une rotation pure pour une multiplicité égale à 1. Dans le cas général (multiplicité 1), deux solutions existent et correspondent à deux plans différents. En faisant l'hypothèse qu'un des plans est celui de la route, donc horizontal, on choisit la solution ayant la normale au plan la plus verticale possible. Cette hypothèse est en pratique parfois fautive et conduit à une estimation erronée, parce que la mauvaise solution a été choisie ou bien parce qu'aucune des deux solutions ne correspond à un plan horizontal.

2.2 À partir de la géométrie épipolaire

La géométrie épipolaire est représentée algébriquement par la matrice essentielle E , et par la matrice fondamentale F , respectivement dans le cas calibré et non-calibré. La notion de calibré correspond au fait que la matrice intrinsèque, K , de la caméra est connue.

$$K = \begin{bmatrix} f_x & s & p_x \\ 0 & f_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

Dans le cas de la géométrie à deux vues [7], E ou F transforme un point dans une image en une ligne dans une autre, sur laquelle doit se trouver le point correspondant. La matrice essentielle est directement le produit vectoriel des quantités que l'on cherche à obtenir, à savoir la translation t et la rotation R entre les deux vues.

$$E = [t]_{\times} R \quad (2)$$

où $[t]_{\times}$ est la matrice antisymétrique telle que, $[t]_{\times} x = t \times x$ pour tout x . Le lien entre E et F étant le suivant $E = K^T F K$.

Différentes méthodes ont été proposées pour estimer F ou E , le cas calibré permet d'après [16] d'augmenter la précision et la robustesse de l'estimation notamment en présence de scènes planaires ou quasi-planaires. Si tous les points de l'image sont coplanaires, il y a une infinité de

possibilités pour choisir F en revanche E est alors restreinte à deux solutions possibles [15] sauf si le mouvement est perpendiculaire au plan. On parle alors de cas dégénéré car plusieurs solutions existent.

Bien que la matrice essentielle puisse être estimée à partir de cinq et six points, seules les méthodes à cinq points présentent cette robustesse au cas planaire. La solution pour estimer le mouvement relatif de la caméra entre deux vues à partir de cinq points est connue depuis longtemps [4], en revanche ce n'est que depuis Nister [16] qu'une implémentation est utilisable en temps réel. D'autres méthodes ont par la suite vu le jour avec notamment le souci d'augmenter la vitesse de calcul, [20] [10] [1]. D'autres encore sont basées sur des processus itératifs [8] [23]. Certains proposent de s'affranchir de la géométrie épipolaire pour estimer la pose à partir de cinq points [3], néanmoins leur méthode nécessite trois vues et un mouvement connu pour résoudre le problème. Les méthodes itératives annoncent diviser le temps de calcul par deux néanmoins elles nécessitent de bonnes conditions d'initialisation et n'annoncent pas de gain significatif en robustesse et/ou en précision. C'est pourquoi la méthode utilisée, dans cette comparaison, pour estimer la matrice essentielle est celle présentée dans [10], qui est une méthode directe. La méthode est imbriquée dans une boucle RANSAC [5]. Le support des hypothèses générées par RANSAC sont calculés en utilisant la distance de Sampson.

Par la suite, on utilise la méthode décrite dans [16] et [7] pour décomposer E en R et t . La méthode est basée sur une décomposition en valeurs singulières de E , qui aboutit à quatre solutions possibles (deux rotations et deux translations). Cependant un seul couple, rotation-translation, permet de trianguler les points devant les deux caméras, de cette façon on discrimine la bonne solution. À noter que l'ambiguïté sur la matrice essentielle dans le cas planaire est également levée de cette façon, sauf si les points de la scène sont tous plus proches d'une des positions de caméra. Ce qui est par exemple le cas lors d'un mouvement majoritairement vers l'avant. La rotation et la translation ainsi obtenues décrivent l'orientation et la position de la première caméra exprimées dans le repère de la seconde.

2.3 Limitations

Dans le cas d'une caméra monoculaire, comme c'est le cas ici, il n'est possible de retrouver le mouvement qu'à un facteur d'échelle près. La rotation est indépendante de l'échelle et est donc estimée correctement, en revanche la translation présente une ambiguïté d'échelle. On considèrera donc uniquement le vecteur unitaire pour la translation ainsi seule sa direction sera évaluée.

$$(R, t) \in \text{SO}(3) \times \mathbb{S}^3, \text{ où } \mathbb{S}^3 \text{ est la sphère unité dans } \mathbb{R}^3 \quad (3)$$

3 Méthode de comparaison

3.1 Vérité terrain à partir des matrices de caméra

Pour chacune des deux caméras, on définit une matrice de projection P de la forme

$$P_i = K [R_i | t_i], \text{ où } i \text{ indexe la caméra, } i \in 1, 2 \quad (4)$$

où K est la matrice intrinsèque, R_i traduit l'orientation du monde par rapport à la caméra i et t_i est la position de l'origine du monde dans le repère de la caméra i . On cherche R décrivant l'orientation du repère de la première caméra par rapport au repère de la seconde et t la position de la première dans le repère de la seconde. On a donc

$$R_{vt} = R_2 R_1^T \quad (5)$$

$$t_{vt} = R_2 (R_2^T t_2 - R_1^T t_1) \quad (6)$$

où vt est pour "vérité terrain". Ces deux valeurs vont servir de référence pour évaluer les erreurs dans l'estimation de la transformation rigide.

3.2 Évaluer l'erreur

Pour le vecteur de translation, on évalue uniquement la direction car la norme est ambiguë. Pour cela on évalue donc l'angle entre deux vecteurs en 3D en calculant leur produit scalaire. En notant $\Delta\theta_t$ l'erreur angulaire sur la direction de la translation, on a

$$\Delta\theta_t = \arccos \left(\frac{t \cdot t_{vt}}{\|t\| \|t_{vt}\|} \right) \quad (7)$$

Pour comparer les matrices de rotation R et R_{vt} , on évalue les différences entre les éléments physiques associés, à savoir l'axe de rotation et l'angle de rotation. On utilise pour cela un résultat de [11] stipulant que toute matrice de rotation peut être décomposée en un axe de rotation u et un angle de rotation α autour de l'axe. En exprimant

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix},$$

on a

$$\alpha = \arccos \left(\frac{\text{trace}(R) - 1}{2} \right) \quad (8)$$

et

$$u = \frac{1}{2 \sin(\alpha)} \begin{pmatrix} r_{32} - r_{23} \\ r_{13} - r_{31} \\ r_{21} - r_{12} \end{pmatrix}. \quad (9)$$

On compare les axes de rotation en utilisant la même méthode que pour les vecteurs de translation.

Ces formules ne sont en revanche pas valables si $R = I_{3 \times 3}$, car l'axe de rotation n'est alors pas défini. On se contentera dans ce cas là de vérifier que l'angle de rotation soit bien égal à 0.

4 Simulations

Dans un premier temps, le modèle de caméra sténopé utilisé pour projeter la scène dans le plan image est présenté. Par la suite les mouvements possibles de caméras sont décrits ainsi que la génération des images correspondant à deux vues.

4.1 Modèle sténopé

On utilise ce modèle pour projeter les points de la scène dans le plan image. On ne tient compte d'aucun effet de distorsion et tous les points de la scène se trouvant dans le champ de vue sont considérés comme étant dans la zone de focus, car on cherche à évaluer l'estimation du mouvement dans différentes situations et non l'influence de la distorsion sur l'estimation. Chaque point 3D, X de la scène est projeté dans le plan image sur x en utilisant la matrice de projection P_i (4).

$$x = P_i X, \text{ où } x \text{ et } X \text{ sont des points homogènes.} \quad (10)$$

Les paramètres utilisés pour la caméra sont les suivants :

- distance focale : 8.5 mm,
- définition de l'image : 1280 × 960,
- taille d'un pixel : 6.45 μm ,
- point principal situé à (640, 480) dans le plan image.

4.2 Mouvements

Deux positions de caméra sont considérées pour chaque simulation afin de se placer dans le cas de la géométrie à deux vues. Les positions sont décrites par le repère qui leur est associé $C_i = [x \ y \ z \ \alpha_x \ \alpha_y \ \alpha_z]$, où x , y et z représentent l'origine du repère caméra dans le repère monde et α_k les rotations du repère caméra par rapport à l'axe k du repère monde. Le repère C_1 de la première position est le même au cours de toutes les simulations, la caméra est placée deux mètres en hauteur et est inclinée de 21 degrés vers le bas selon l'axe x , $C_1 = [0 \ 2 \ 0 \ 21 \ 0 \ 0]$. Cette position correspond approximativement au placement de la caméra sur l'UGV par rapport au sol. On génère ensuite la deuxième position aléatoirement et en ayant suffisamment de recouvrement. Ainsi C_2 a pour origine $[\mathcal{U}(-1, 1) \ 2 \ \mathcal{U}(0, 1)]$ et pour orientation $[21 + \mathcal{N}(0, 2) \ \mathcal{U}(-10, 10) \ \mathcal{U}(-5, 5)]$ où $\mathcal{U}(a, b)$ est la distribution uniforme entre a et b . Afin de ne pas avoir de déplacements trop petits ou trop grands on limite la norme de la translation de C_1 à C_2 entre 10 cm et 1 m.

4.3 Scènes et images synthétiques

On considère des images ayant des points répartis sur toute la surface de l'image. Les images sont toujours constituées de 256 points visibles depuis les deux positions de caméra afin de ne pas avoir de variabilité sur l'estimation de la matrice d'homographie ou essentielle en fonction du nombre de points utilisés. Les scènes présentent toujours un plan dominant horizontal $y = 0$ correspondant au sol.

Le nombre de points hors de ce plan est augmenté progressivement entre 0 % et 20 % des points vus par le couple de caméras. Avoir 2 % de points hors plan signifie que sur les 256 points observés, cinq ne sont pas dans le plan.

4.4 Erreur de calibrage

L'influence des erreurs de calibrage est également étudiée. Pour cela, des perturbations sont ajoutées sur les éléments de la matrice K . L'élément s sera toutefois toujours forcé à 0 dans les simulations. [19] le justifie en annonçant que ce degré de liberté supplémentaire ne sert souvent qu'à adapter le modèle au bruit et le calibrage résultant s'en trouve dégradé. La bibliothèque OpenCV force également ce paramètre à 0. Les perturbations sur f_x , f_y , p_x et p_y sont des bruits gaussien de moyenne nulle et de variance σ^2 ,

$$K = \begin{bmatrix} f_x + \mathcal{N}(0, \sigma^2) & 0 & p_x + \mathcal{N}(0, \sigma^2) \\ 0 & f_y + \mathcal{N}(0, \sigma^2) & p_y + \mathcal{N}(0, \sigma^2) \\ 0 & 0 & 1 \end{bmatrix}.$$

La valeur choisie pour les simulations est $\sigma^2 = 100$, ce qui correspond à des erreurs de calibrage obtenue lorsque celui-ci n'est pas réalisé de manière minutieuse.

4.5 Comparaison des résultats

On génère 100 couples d'images pour chaque pourcentage de points hors du plan et on mesure l'erreur par rapport à la vérité terrain tel que décrit précédemment.

On compare les erreurs obtenues par chacune des décompositions suivant 5 valeurs de statistique descriptive (Five-point summary en anglais). Le deuxième centile, le premier quartile, la médiane, le troisième quartile et le 98^e centile. On pourra ainsi comparer la justesse de l'estimation en comparant les valeurs médianes et la fidélité grâce à la dispersion des erreurs. Si l'erreur médiane est faible, l'estimation est alors juste. Si les quartiles et les centiles sont proches de la médiane, l'estimation est alors fidèle. Si les deux conditions vraies alors l'estimation sera jugée exacte.

5 Résultats des simulations

5.1 Cas idéal : calibrage parfait

Les résultats obtenus avec un calibrage parfait se divisent en deux cas. Un cas où les points utilisés ont été donnés en coordonnées flottantes, tels que projetés par le modèle de caméra, ce qui représenterait un détecteur de coins sub-pixelique idéal. L'autre cas correspond à des coins exprimés en valeurs entières de pixels, les valeurs flottantes ayant été arrondies aux entiers les plus proches. Les résultats sont donnés Figure 2 et Figure 3. Dans les deux cas l'ambiguïté sur la matrice essentielle est présente lorsque la scène est parfaitement planaire et à moins de lever celle-ci avec une troisième vue, la décomposition résultante est fautive. L'erreur médiane étant supérieure à 10 degrés. Parmi les trois éléments de la transformation rigide, l'angle de rotation est systématiquement estimé avec l'erreur la plus faible.

Dans le cas sub-pixelique, l'erreur sur la décomposition de E est inférieure à celle de la matrice d'homographie dès lors que la scène présente quelques points hors du plan et à partir de 4% de points hors du plan, l'erreur est divisée par dix. La matrice d'homographie garde une erreur médiane faible pour la translation en revanche des erreurs de l'ordre du degré sont possibles et une tendance croissante est observée en fonction du nombre de points hors du plan pour la translation et l'angle de rotation. L'axe de rotation présente une erreur relativement importante par rapport aux autres éléments.

Dans le cas avec des coordonnées entières, les performances de la décomposition de la matrice essentielle sont dégradées et même inférieures à celle de la matrice d'homographie. On remarque une plus grande dispersion des erreurs et une erreur médiane légèrement supérieure. De plus des erreurs de l'ordre de la dizaine de degrés sur la translation et l'axe de rotation ne sont pas à exclure, contrairement à l'homographie.

Avec la Figure 4 et Figure 5, deux plans dominants sont présents et le pourcentage de points hors du plan correspond au nombre de points dans le plan qui n'est pas horizontal. Le plan utilisé par l'homographie n'est pas toujours horizontal, comme le montre les valeurs élevées du troisième quartile et du 98^e centile. Les erreurs maximum sont de l'ordre du degré voir de la dizaine de degré pour la translation et l'axe de rotation pour la matrice homographique. Cette différence n'est en revanche pas visible pour des points données en coordonnées entières.

5.2 Erreurs de calibrage

En perturbant le calibrage par un bruit $\mathcal{N}(0, 100)$, les conclusions du cas idéal restent valables, on a en revanche une différence moins importante entre les deux décompositions dans le cas sub-pixelique. Les résultats sont donnés Figure 6 et Figure 7.

6 Avec des images réelles

6.1 Expérience

Afin de compléter l'étude, les décompositions de la matrice homographique et essentielle sont également comparées sur une séquence d'images réelles provenant du *KITTI Vision Benchmark Suite* [6]. L'utilisation de cette base de données permet d'avoir accès à la position des caméras lors de chaque prise de vue et donc d'avoir la vérité terrain. Les images utilisées sont issues de la séquence 03 et les images utilisées vont de l'indexe 000000 à l'indexe 000150 en utilisant une image sur deux. La facteur d'échelle pour la translation entre deux images est obtenu à partir de la vérité terrain. Bien que dans la séquence la caméra ne soit pas orientée vers le bas comme sur l'UGV, les images correspondent néanmoins à des types de scènes observables par celui-ci. Les images du benchmark, contiennent un plan dominant, la route, qui est plus présent que dans les Figures 1d et 1e.

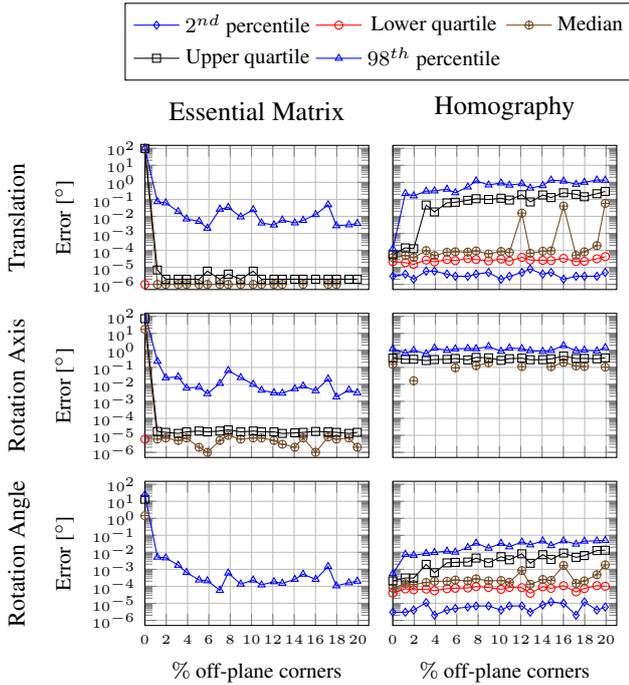


FIGURE 2 – Erreurs obtenues lorsque le calibrage est parfait et les points donnés en coordonnées flottantes

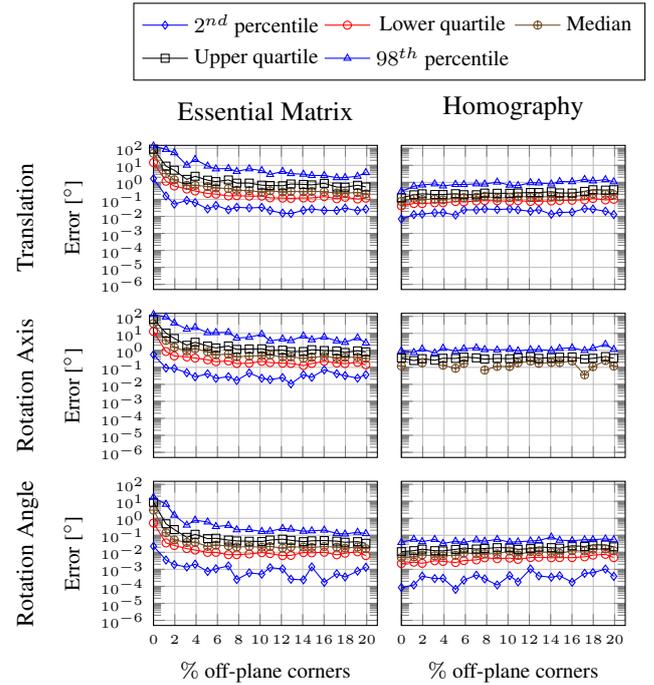


FIGURE 3 – Erreurs obtenues lorsque le calibrage est parfait et les points donnés en coordonnées entières

6.2 Résultats

Les trajectoires obtenues sont données Figure 8. La trajectoire utilisant la décomposition de la matrice essentielle est significativement plus précise que celle issue de la décomposition de la matrice homographique. Il y a 2.47 m de dérive dans le plan $x - z$ contre 49.81 m. La décomposition homographique souffre de la non-validité de l'hypothèse que le plan de la route soit le plan choisi par l'homographie et cela donne parfois un virage à angle droit entre deux prises de vues ce qui est aberrant dans le cas d'un véhicule terrestre à roues.

7 Conclusion

L'estimation du mouvement de caméra utilisée dans le système de détection de changement est basée sur la décomposition de la matrice d'homographie liant deux plans images. Cette méthode repose sur l'apriori que le plan utilisé par l'homographie est celui de la route, qui est souvent le plan le plus présent dans la scène. Lorsque ce n'est pas le cas, la décomposition obtenue peut donner un mouvement erroné entre deux prises de vues. On a donc cherché à utiliser la décomposition d'une autre entité obtenue à partir de correspondances dans les images, à savoir la matrice essentielle. L'apport en justesse et en fidélité sur la qualité de la décomposition pour obtenir la transformation rigide entre deux prises de vue a été évaluée.

En simulation, il a été montré que la décomposition de la matrice essentielle est à privilégier lorsque la scène n'est

pas complètement planaire, auquel cas une troisième vue serait alors nécessaire. En revanche, la décomposition de la matrice essentielle est influencée de façon significative par la précision des points utilisés pour l'estimation de la matrice. En effet lorsque les points sont donnés en coordonnées entières, la décomposition homographique est plus précise. Concernant les erreurs de calibrage, celles-ci semblent influencer de la même façon la décomposition de la matrice essentielle et homographique. On conseillera d'utiliser un détecteur de coins sub-pixelique afin d'obtenir une meilleure estimation de E et ainsi avoir un mouvement plus précis.

Sur des images réelles, la décomposition de la matrice essentielle est bien supérieure à la décomposition homographique. Cette dernière étant moins précise et estimant parfois de manière erronée le mouvement entre deux prises de vues. La dérive devient rapidement importante.

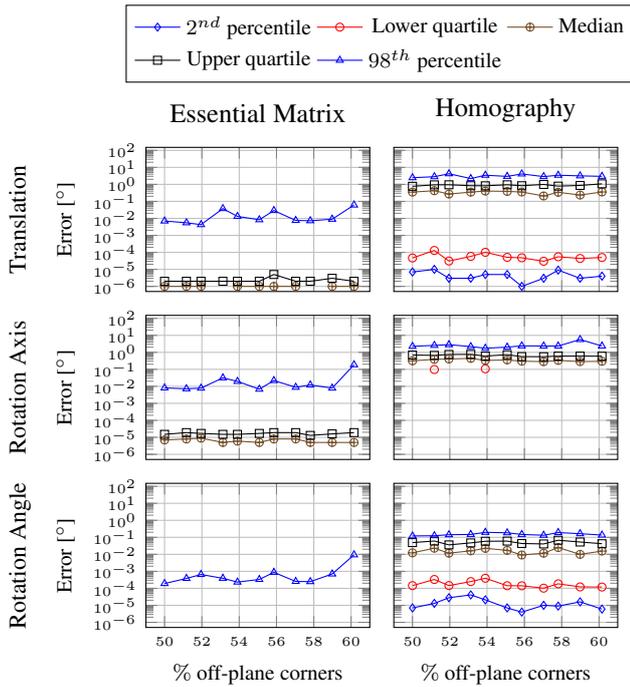


FIGURE 4 – Erreurs obtenues lorsque le calibrage est parfait et les points donnés en coordonnées entières

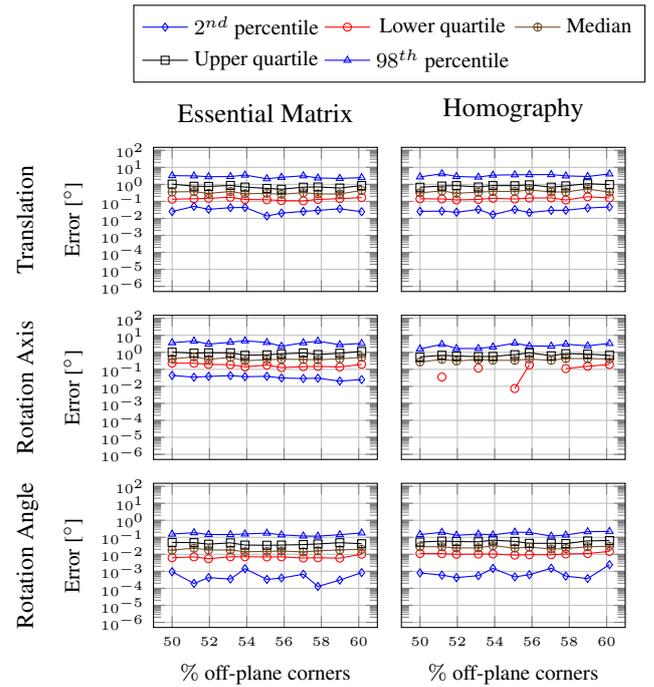


FIGURE 5 – Erreurs obtenues lorsque le calibrage est parfait et les points donnés en coordonnées entières

Références

- [1] T. Botterill, S. Mills, and R. Green. Fast RANSAC hypothesis generation for essential matrix estimation. In *Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on*, pages 561–566. IEEE, 2011.
- [2] C. Engels, H. Stewénius, and D. Nistér. Bundle adjustment rules. *Photogrammetric computer vision*, 2 :124–131, 2006.
- [3] K. Fathian and N. R. Gans. A new approach for solving the Five-Point Relative Pose Problem for vision-based estimation and control. In *American Control Conference (ACC), 2014*, pages 103–109. IEEE, 2014.
- [4] O. Faugeras and S. Maybank. Motion from point matches : multiplicity of solutions. Research Report RR-1157, INRIA, 1990.
- [5] M. A. Fischler and R. C. Bolles. Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6) :381–395, 1981.
- [6] 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.
- [7] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN : 0521540518, second edition, 2004.
- [8] J. Hedborg and M. Felsberg. Fast iterative five point relative pose estimation. In *Robot Vision (WORV), 2013 IEEE Workshop on*, pages 60–67. IEEE, 2013.
- [9] L. Kneip, D. Scaramuzza, and R. Siegwart. A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2969–2976. IEEE, 2011.
- [10] H. Li and R. Hartley. Five-Point Motion Estimation Made Easy. In *18th International Conference on Pattern Recognition (ICPR'06)*, volume 1, pages 630–633, 2006.
- [11] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry. *An invitation to 3-d vision : from images to geometric models*, volume 26. Springer Science & Business Media, 2012.
- [12] M. Maimone, Y. Cheng, and L. Matthies. Two years of visual odometry on the mars exploration rovers. *Journal of Field Robotics*, 24(3) :169–186, 2007.
- [13] 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.
- [14] D. Monnin, A. L. Schneider, and E. Bieber. Detecting suspicious objects along frequently used itineraries. In *Security+ Defence*, pages 78340F–78340F. International Society for Optics and Photonics, 2010.
- [15] S. Negahdaripour. Closed-form relationship between the two interpretations of a moving plane. *J. Opt. Soc. Am. A*, 7(2) :279–285, Feb. 1990.
- [16] 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.

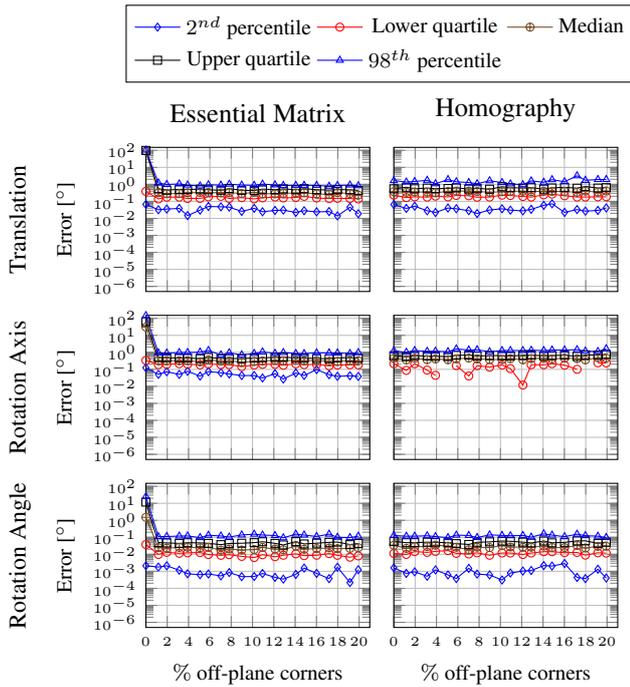


FIGURE 6 – Erreurs obtenues lorsque le calibrage est perturbé par $\mathcal{N}(0, 100)$ et les points donnés en coordonnées flottantes

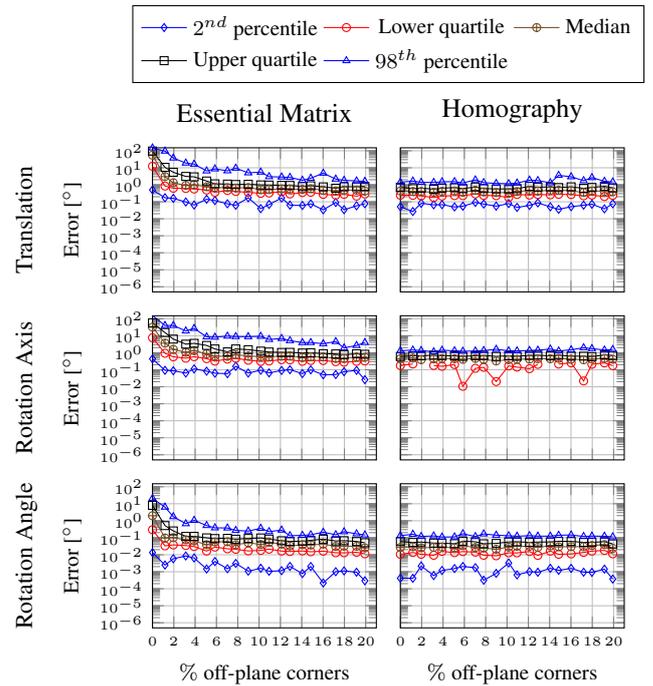


FIGURE 7 – Erreurs obtenues lorsque le calibrage est perturbé par $\mathcal{N}(0, 100)$ et les points donnés en coordonnées entières

- [17] D. Nistér, O. Naroditsky, and J. Bergen. Visual odometry. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–652. IEEE, 2004.
- [18] D. Scaramuzza and F. Fraundorfer. Visual odometry [tutorial]. *IEEE Robotics & Automation Magazine*, 18(4):80–92, 2011.
- [19] K. Simek. The Perspective Camera - An Interactive Tour, Aug. 2012.
- [20] H. Stewenius, C. Engels, and D. Nister. An Efficient Minimal Solution for Infinitesimal Camera Motion. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.
- [21] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle Adjustment — A Modern Synthesis. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms : Theory and Practice : International Workshop on Vision Algorithms Corfu, Greece, September 21–22, 1999 Proceedings*, pages 298–372. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000. DOI : 10.1007/3-540-44480-7_21.
- [22] R. Tsai, T. Huang, and W.-L. Zhu. Estimating three-dimensional motion parameters of a rigid planar patch, II : singular value decomposition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 30(4):525–534, 1982.
- [23] T. D. M. U. Vincent Lui (Monash University). An Iterative 5-pt Algorithm for Fast and Robust Essential Matrix Estimation. In *Proceedings of the British Machine Vision Conference*. BMVA Press, 2013.

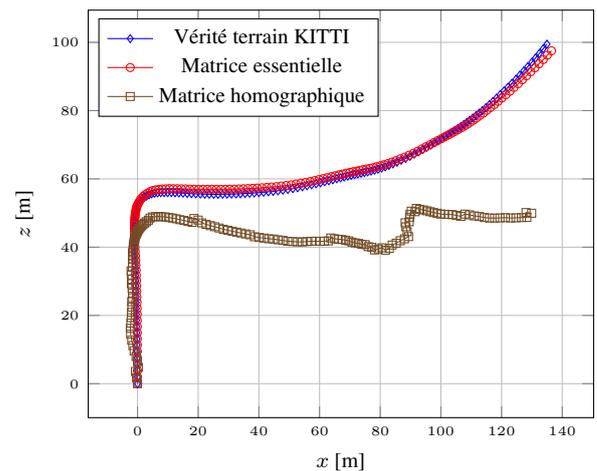


FIGURE 8 – Trajectoires obtenues sur la séquence 03 de *KITTI* jusqu’à l’image 150