

# Couplage Visuel Inertiel pour une navigation précise et robuste

Gaëtan CHEVRIN<sup>1</sup>, Sébastien CHANGEY<sup>1</sup>, David MONNIN<sup>1</sup>, Jean-Philippe LAUFFENBURGER<sup>2</sup>

<sup>1</sup>Institut Franco-Allemand de Recherches de Saint-Louis (ISL)  
5 rue du Général Cassagnou, 68300 Saint-Louis, France

<sup>2</sup>Université de Haute-Alsace, IRIMAS UR 7499  
12 rue des Frères Lumière, 68093 Mulhouse Cedex, France  
prenom.nom@isl.eu, jean-philippe.lauffenburger@uha.fr

**Résumé** – Dans cet article, nous présentons une étude contribuant au développement d’un système de navigation visuo-inertiel ayant pour objectifs la précision, l’agnosticité au système mobile auquel il est appliqué et la robustesse aux erreurs de mesure des capteurs. Les méthodes visuelle et inertielle pouvant toutes deux être sujettes à des erreurs que l’autre méthode peut potentiellement compenser, nous présentons ici deux méthodes à couplage lâche pour tirer parti de ce constat ; un ”Error-State Kalman Filter” (ESKF) et un ”Multi Sensor Fusion Extended Kalman Filter” (MSF-EKF). Les deux méthodes permettent des estimations correctes en cas de défaillance de la composante visuelle ou inertielle du système de navigation. Cependant, la méthode MSF-EKF s’est avérée plus efficace et plus polyvalente que la méthode ESKF.

**Abstract** – In this paper, we present a study contributing to the development of a visual-inertial navigation system with the objectives of accuracy, agnosticity to the mobile system to which it is applied and robustness to sensor measurement errors. Since both the visual and inertial method are potentially subject to weaknesses that the other method may possibly compensate, we present here two loose coupling methods to take advantage of this statement; an Error-State Kalman Filter (ESKF) and a Multi Sensor Fusion Extended Kalman Filter (MSF-EKF). Both methods provide correct estimations in case of failure of the visual or inertial component of the navigation system. However, the MSF-EKF method has proven to be more efficient and versatile compared to the ESKF method.

## 1 Introduction

La navigation est une étape nécessaire à toute conduite autonome de véhicule. L’estimation de position est essentielle pour certaines applications comme la cartographie et l’évitement d’obstacles. La majorité des systèmes de navigation reposent sur les technologies Global Navigation Satellites Systems (GNSS) afin d’obtenir une estimation correcte de position [1]. Cependant, cette solution est dépendante de la qualité du signal GNSS qui peut être altérée par un large éventail d’interférences (brouillage, perte du signal, réception d’un signal de mauvaise qualité, etc.). Une telle altération de l’estimation de position doit donc être évitée, en particulier pour des systèmes de défense. Bien que des techniques pour lutter contre le brouillage soient développées [7], une solution à ce problème est de ne plus dépendre d’aucune technologie GNSS pour la navigation. C’est le cas des systèmes ISL-STAMINA : utilisant une caméra monoculaire, ils estiment leur mouvement par la vision seule à l’aide de l’algorithme Parallax Beam (PB) [15] qui offre des résultats comparables à l’initialisation de l’algorithme ORB-SLAM [14] et une robustesse renforcée face aux scènes quasi-planes, communes dans le cadre des applications d’ISL-STAMINA.

Néanmoins, cet algorithme atteint ses limites lorsque la caméra n’est pas en mesure d’obtenir une image exploitable (occultation du champ de vision, etc.). Dans ces cas extrêmes de rupture

du flux d’images, il est nécessaire de toujours être en mesure d’assurer la fonction de navigation et de garantir un niveau de robustesse suffisant vis-à-vis d’erreurs capteurs.

Les centrales inertielle, ou Inertial Measurement Units (IMU) sont un composant clé de la navigation autonome [4] et semblent être une solution adaptée au problème sus-mentionné. Pour un coût modeste, elles permettent d’accéder à plusieurs informations proprioceptives comme l’accélération et la vitesse de rotation.

L’objectif de nos travaux est d’étudier le couplage entre PB et une centrale inertielle afin de conserver la précision tout en garantissant la robustesse de navigation vis-à-vis de toute rupture du flux d’images ou défaut de la caméra. Notre objectif est également de réaliser un couplage agnostique du système sur lequel il est embarqué afin de le rendre aussi polyvalent que possible.

La fusion multi-capteurs et en particulier les couplages des systèmes de navigation visuels et inertiels (VINS) [16] sont un sujet de recherche largement exploré avec un grand nombre de méthodes disponibles. Les méthodes de cartographie (mapping), qui reposent essentiellement sur l’association de features avec des positions et des mesures inertielle, qui n’exploitent donc pas les contributions du PB ne sont pas considérées et ne seront donc pas discutées ici.

Dans le cas général, les techniques de couplage peuvent être distinguées selon deux catégories : le couplage lâche et le cou-

plage serré [5]. Un couplage lâche entre une IMU et un autre capteur (une caméra monoculaire dans notre cas) implique généralement d'utiliser les données issues des accéléromètres et des gyroscopes afin de réaliser l'étape de propagation d'un filtre de Kalman pour faire une prédiction qui sera ensuite mise à jour à l'aide de l'estimation de navigation obtenue par le deuxième capteur comme pour les deux méthodes suivantes. Une architecture intéressante d'algorithme VINS est présentée dans [16]. Elle emploie un EKF (filtre de Kalman étendu) ou un ESKF (error-state Kalman filter) afin de coupler une IMU pour la prédiction et une caméra monoculaire avec un capteur de distance infrarouge pour la mise à jour du filtre. Ce dispositif est utilisé avec une grande fréquence d'échantillonnage (jusqu'à 200 Hz pour tous les capteurs) et obtient des résultats précis avec une RMSE de (0.47 0.67 0.035)(m), l'axe z étant l'axe optique de la caméra.

Le Multi Sensor Fusion Extended Kalman Filter (MSF-EKF) [12] peut être perçu comme une généralisation des méthodes de couplage lâche précédentes. Tout en conservant l'idée de base d'utiliser l'IMU afin de réaliser sa prédiction comme le ferait un ESKF, ce filtre permet l'addition modulaire de plusieurs capteurs d'estimation de position pour la mesure et la mise à jour du filtre. Des travaux présentés dans [6] utilisent cette technique afin de réaliser un système VINS pour UAV avec une erreur finale absolue de moins de 5 cm. L'existence d'un MSF-UKF (Unscented Kalman Filter) [17] est intéressante car elle contourne les problèmes numérique du calcul de la Jacobienne propre à un EKF afin de produire de bonnes performances de navigation, mais n'offre pas un gain de performance significatif pour des modèles à faible non linéarité comme c'est le cas ici.

Nous ne considérerons pas les méthodes de couplages serrés ici car nous souhaitons profiter du plein potentiel du Parallax Beam dont la contribution principale n'intervient qu'après l'étape de détection et de description des coins [15] là où le couplage serré utilise le plus souvent uniquement ce type de données issues des systèmes basés vision [5].

Comme le filtre de Kalman est un outil commun et efficace afin de réaliser un couplage lâche [2], la majorité de nos travaux se sont tournés vers eux.

Le présent article présente tout d'abord les deux méthodes d'estimations que nous souhaitons coupler, la navigation inertielle et Parallax Beam, avant de présenter deux méthodes de couplages, une à travers un ESKF et une autre utilisant un MSF-EKF. Les performances des ces algorithmes seront évaluées sur le jeu de données KITTI.

## 2 Méthodes d'estimations du mouvement

Parallax Beam et la navigation inertielle ont pour objectifs de calculer les angles d'Euler  $r_n \in \mathbb{R}^3$  et la position  $p_n \in \mathbb{R}^3$  à chaque instant  $n$  d'une plateforme mobile par rapport à sa position précédente  $(r_{n-1}, p_{n-1})$ . Les estimations du Parallax Beam sont notées  $r_n^{pb}$  et  $p_n^{pb}$  tandis que les estimations iner-

tielles sont notées  $r_n^{imu}$  et  $p_n^{imu}$ .

### 2.1 Estimation du mouvement par la vision : Parallax Beam

Parallax Beam réalise l'estimation du mouvement à partir d'images successives. En prenant en entrée deux images successives, il évalue le mouvement relatif entre les deux images. Pour ce faire, il détecte les points clés de l'image à l'aide d'un détecteur de coins de Harris modifié [9] qui assure une distribution homogène sur l'image des éléments-clés détectés. Ces éléments sont ensuite décrits par le descripteur BRIEF. Une homographie entre les deux images est alors estimée en utilisant une procédure de RANSAC modifiée [13]. A condition que le plan du sol soit visible par la caméra, le résultat de cette procédure est une matrice d'homographie notée  $M_n^{pb}$ . Si la scène n'est pas plane, une homographie ne suffit pas à contenir toute les informations fournis par les points-clés de l'image. Dans ce cas, l'algorithme utilise un modèle de PB expliqué en détail dans [15] pour estimer la matrice fondamentale informant sur le mouvement de la caméra.

### 2.2 Estimation du mouvement inertielle

Dans nos travaux, nous choisissons d'ignorer les effets de la rotation terrestre et l'accélération de Coriolis. La Terre est considérée localement plate et le vecteur gravité  $g \in \mathbb{R}^3$  est une constante connue.

Une centrale inertielle fournit des estimations bruitées et biaisées du vecteur de vitesse angulaire  $\omega_n^{imu}$  et de l'accélération spécifique  $a_n^{imu}$  [11].

$$\omega_n^{imu} = \omega_n + b_n^\omega + w_n^\omega \quad (1)$$

$$a_n^{imu} = a_n + b_n^a + w_n^a \quad (2)$$

$b_n^\omega$  et  $b_n^a$  sont des biais tandis que  $w_n^\omega$  et  $w_n^a$  sont des bruits blancs gaussiens. Les biais suivent une marche aléatoire bruitées par un bruit blanc gaussien.

À partir de ces données inertielles, il est possible de reconstruire la pose de la plateforme équipée par la méthode de navigation à l'estime [3]. Cette méthode dans sa forme la plus simple offre des performances faibles à cause des dérives. C'est généralement par ce biais que l'on réalise l'étape de prédiction des filtres de Kalman pour le couplage VINS.

Comme on peut le voir dans la table 1, Parallax Beam est plus précis que la navigation à l'estime, ce qui peut être expliqué par la dérive inertielle. Cependant, on constate que la précision du PB se dégrade si on provoque une rupture du flux d'image (échantillonnage à 5 Hz au lieu de 10 Hz). Un couplage lâche devrait permettre de garantir la précision du PB en dépit de quelques ruptures du flux d'images tout en minimisant l'impact de la dérive inertielle.

TABLE 1 – Erreurs (RPE) en rotation et translation moyennes de nos méthodes d’estimations obtenues sur les séquences 05, 09 et 10 de KITTI [8]

KITTI sequence	05	09	10
<b>Parallax Beam</b>			
Translation [%]	2.14	2.7	1.54
Rotation [°/m]	0.0083	0.0074	0.0096
<b>Parallax Beam dégradé</b>			
Translation [%]	3.8	15.8	8.7
Rotation [°/m]	0.0207	0.0764	0.0337
<b>Navigation à l’estime</b>			
Translation [%]	275	37.5	27.5
Rotation [°/m]	0.0056	0.0040	0.0041

### 3 Error-State Kalman Filter

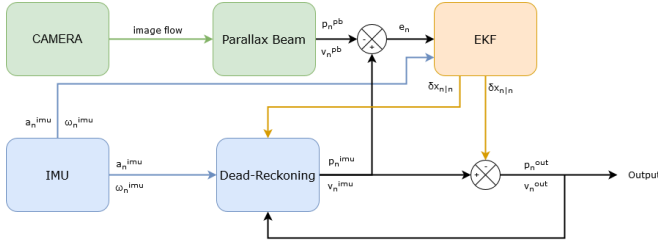
Une première étude de couplage lâche consiste à réaliser un Error-State Kalman filter (ESKF). Pour ce faire, on définit notre état comme une erreur [10] :

$$\delta x_n = \delta x_{n|n} = [\delta r_n; b_n^\omega; \delta p_n; \delta v_n; b_n^a] \quad (3)$$

Avec  $v_n$  la vitesse de la plate-forme.

Le filtre ESKF réalise une première prédiction à l’aide de la centrale inertielle puis se met à jour grâce à la mesure de la différence entre Parallax Beam et la navigation à l’estime simple comme l’illustre la figure 1. Nous avons testé cette architecture

FIGURE 1 – Schéma-block de l’ESKF proposé



sur le jeu de données KITTI [8] selon deux scénarios : un couplage classique sans rupture du flux d’image (vision échantillonnée à 10 Hz) et un second couplage avec la vision sous échantillonnée à 5 Hz.

Ces travaux (Table 2) ont montré la possibilité d’un couplage lâche entre une IMU et l’algorithme PB. Le filtre ESKF, bien que moins précis que le PB, est capable d’assurer une navigation correcte en l’absence d’information venant du PB.

Néanmoins, cette solution peut être améliorée sur plusieurs points comme sa fréquence d’échantillonnage, sa précision et sa robustesse à des ruptures ponctuelles du flux images.

### 4 MSF-EKF

Afin d’obtenir de meilleures performances, nous avons mis en place un filtre Multi-Sensor Fusion Extended Kalman Filter (MSF-EKF) en nous inspirant des travaux de [12] [19]. Ce filtre réalise une prédiction à partir des données inertielles et utilise

TABLE 2 – Erreurs (RPE) en rotation et translation moyennes de l’ESKF avec et sans sous-échantillonnage obtenues sur les séquences 05, 09 et 10 de KITTI

KITTI sequence	05	09	10
<b>ESKF (10 Hz)</b>			
Translation [%]	3.76	7.9	3.03
Rotation [°/m]	0.0056	0.0040	0.0041
<b>ESKF (5 Hz)</b>			
Translation [%]	7.6	51	21.2
Rotation [°/m]	0.0056	0.0040	0.0041

pour sa mise à jour les mesures de position du Parallax Beam. L’innovation du filtre MSF-EKF réside dans sa capacité à utiliser plusieurs mesures de capteurs différents afin de réaliser ses mises à jour, même avec des mesures non-synchrones. Contrairement au ESKF, nous utilisons les données brutes de KITTI afin de disposer d’une IMU à 100 Hz pour nos prédictions, le PB restant à 10 Hz. Ainsi nous garantissons une plus grande fréquence d’échantillonnage du filtre et de l’estimation de position. De plus, nous avons décidé d’augmenter le vecteur d’état avec l’objectif d’améliorer les performances de notre navigation.

D’une manière similaire à l’ESKF, ce filtre réalise l’estimation d’un état-erreur, ceci afin d’améliorer la stabilité de notre filtre face aux erreurs numériques [18].

$$\tilde{x}_n = [\Delta p_n, \Delta v_n, \delta \theta_n, \Delta b_n^\omega, \Delta b_n^a, \Delta \lambda_n, \delta \theta_n^{wv}, \delta \theta_n^{ci}, \Delta p_n^{ci}] \quad (4)$$

Avec  $\Delta \lambda_n$  l’erreur du facteur d’échelle du Parallax Beam,  $\delta \theta_n$  l’erreur du quaternion d’attitude,  $\delta \theta_n^{wv}$  l’erreur du quaternion signifiant la rotation entre le repère de référence (ou repère monde) et le repère caméra,  $\delta \theta_n^{ci}$  l’erreur de rotation entre les repères caméra et inertiel et  $\Delta p_n^{ci}$  l’erreur de position entre les repères caméra et inertiel.

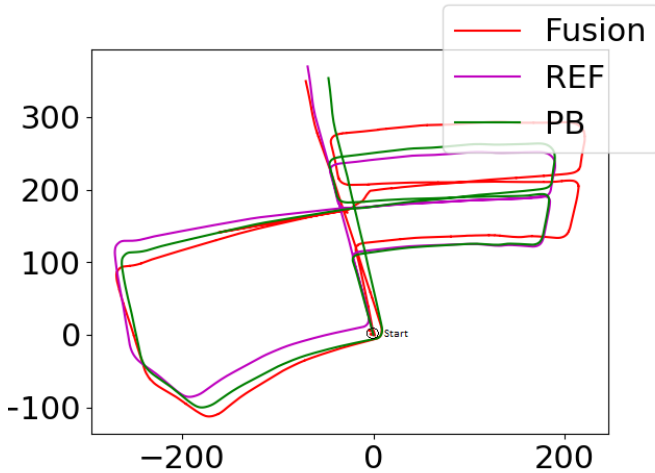
Ce filtre offre des premiers résultats encourageants avec une erreur terminale (erreur finale divisée par la longueur totale de la trajectoire) en position de 25% (contre 32% pour PB avec la même évaluation) et d’attitude de 0.12°/m sur la séquence 05 de KITTI d’une distance de 2 km (figure 2). Les performances moyennes de la méthode sur KITTI restent à déterminer et la présence d’outliers requiert la mise en place d’outils adaptés pour améliorer encore ces premiers résultats.

### 5 Conclusion

Les couplages lâches réalisés, ESKF et MSF-EKF montrent des résultats encourageant dans la poursuite d’une fusion lâche de système visuels-inertiels en offrant une précision satisfaisante tout en garantissant la fonction de navigation à des fréquences supérieures à celles du PB (100 Hz contre 10 HZ pour le PB sur KITTI). De plus amples travaux sont requis afin de définir plus précisément les performances du MSF-EKF. Un système d’acquisition de données visuelles et inertielles asyn-

chrones a été réalisé à cet effet afin de tester les conditions limites des deux capteurs mobilisés pour le couplage.

FIGURE 2 – Résultat d'estimation sur la trajectoire 05 de KITTI. On peut observer la référence (magenta), le Parallax Beam (vert) et notre algorithme (rouge).



## Références

- [1] R. Attia, J. Daniel, J.-P. Lauffenburger, R. Orjuela, and M. Basset. Reference generation and control strategy for automated vehicle guidance. In *2012 IEEE Intelligent Vehicles Symposium*, pages 389–394. IEEE, 2012.
- [2] M. Brossard, A. Barrau, and S. Bonnabel. Rins-w : Robust inertial navigation system on wheels. *arXiv preprint arXiv :1903.02210*, 2019.
- [3] M. Brossard, A. Barrau, and S. Bonnabel. Ai-imu dead-reckoning. *IEEE Transactions on Intelligent Vehicles*, 2020.
- [4] T. Brunner, J.-P. Lauffenburger, S. Changey, and M. Basset. Quaternion-based imu and stochastic error modeling for intelligent vehicles. In *2015 IEEE Intelligent Vehicles Symposium (IV)*, pages 877–882. IEEE, 2015.
- [5] Z. Chen, Y. Qu, X. Ling, Y. Li, H. Jiao, and Y. Liu. Study on gps/ins loose and tight coupling. In *2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, volume 1, pages 34–37. IEEE, 2015.
- [6] M. Faessler, F. Fontana, C. Forster, E. Mueggler, M. Pizzoli, and D. Scaramuzza. Autonomous, vision-based flight and live dense 3d mapping with a quadrotor micro aerial vehicle. *Journal of Field Robotics*, 33(4) :431–450, 2016.
- [7] G. X. Gao, M. Sgammini, M. Lu, and N. Kubo. Protecting gnss receivers from jamming and interference. *Proceedings of the IEEE*, 104(6) :1327–1338, 2016.
- [8] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics : The kitti dataset. *The International Journal of Robotics Research*, 32(11) :1231–1237, 2013.
- [9] 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, 2012.
- [10] A. R. Jiménez, F. Seco, J. C. Prieto, and J. Guevara. Indoor pedestrian navigation using an ins/ekf framework for yaw drift reduction and a foot-mounted imu. In *2010 7th Workshop on Positioning, Navigation and Communication*, pages 135–143. IEEE, 2010.
- [11] M. Kok, J. D. Hol, and T. B. Schön. Using inertial sensors for position and orientation estimation. *arXiv preprint arXiv :1704.06053*, 2017.
- [12] S. Lynen, M. W. Achtelik, S. Weiss, M. Chli, and R. Siegwart. A robust and modular multi-sensor fusion approach applied to mav navigation. In *2013 IEEE/RSJ international conference on intelligent robots and systems*, pages 3923–3929. IEEE, 2013.
- [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] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos. Orbslam : a versatile and accurate monocular slam system. *IEEE transactions on robotics*, 31(5) :1147–1163, 2015.
- [15] M. Rebert, D. Monnin, S. Bazeille, and C. Cudel. Parallax beam : a vision-based motion estimation method robust to nearly planar scenes. *Journal of Electronic Imaging*, 28(2) :023030, 2019.
- [16] A. Santamaria-Navarro, G. Loianno, J. Sola, V. Kumar, and J. Andrade-Cetto. Autonomous navigation of micro aerial vehicles using high-rate and low-cost sensors. *Autonomous robots*, 42(6) :1263–1280, 2018.
- [17] S. Shen, Y. Mulgaonkar, N. Michael, and V. Kumar. Multi-sensor fusion for robust autonomous flight in indoor and outdoor environments with a rotorcraft mav. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4974–4981. IEEE, 2014.
- [18] N. Trawny and S. I. Roumeliotis. Indirect kalman filter for 3d attitude estimation. *University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep.*, 2 :2005, 2005.
- [19] S. Weiss and R. Siegwart. Real-time metric state estimation for modular vision-inertial systems. In *2011 IEEE international conference on robotics and automation*, pages 4531–4537. IEEE, 2011.