



HAL
open science

HICALIB : Méthode hybride d'étalonnage et d'estimation de pose RVB-D-IMU

Abanob Soliman, Fabien Bonardi, Désiré Sidibé, Samia Bouchafa

► **To cite this version:**

Abanob Soliman, Fabien Bonardi, Désiré Sidibé, Samia Bouchafa. HICALIB : Méthode hybride d'étalonnage et d'estimation de pose RVB-D-IMU. ORASIS 2023, Laboratoire LIS, UMR 7020, May 2023, Carqueiranne, France. hal-04219525

HAL Id: hal-04219525

<https://hal.science/hal-04219525v1>

Submitted on 27 Sep 2023

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.

HICALIB : Méthode hybride d'étalonnage et d'estimation de pose RVB-D-IMU

A. Soliman¹

F. Bonardi¹

D. Sidibé¹

S. Bouchafa¹

¹ Université Paris-Saclay, Univ Evry, Laboratoire IBISC, 91020, Essonne, France

abanob.soliman@universite-paris-saclay.fr

Résumé

Cet article propose une nouvelle méthode basée sur l'optimisation pour la fusion de données issues de capteurs multimodaux dans l'estimation de la pose 3D (odométrie) en robotique. La méthode consiste à calibrer les paramètres intrinsèques et extrinsèques invariants dans le temps à l'aide d'un algorithme d'initialisation (bootstrapping) optimisé par des données GPS. Une première étape fournit des estimations initiales fiables pour la caméra RVB et l'étape finale optimise les paramètres spatio-temporels tels que la pose de la cible et les biais de centrale inertielle (IMU). L'algorithme proposé est validé par des résultats expérimentaux sur des séquences réelles et simulées et des études d'ablation acquises à partir de véhicules terrestres et aériens afin d'estimer la contribution de chaque capteur à la précision de l'estimation de la pose. Dépôt Github : <https://github.com/AbanobSoliman/HICALIB>

Mots Clef

RVB-D-IMU, étalonnage, optimisation.

Abstract

This paper proposes a novel optimization-based method for fusing data from multimodal sensors in 3D pose estimation (odometry) in robotics. The method involves calibrating time-invariant intrinsic and extrinsic parameters using a GPS-optimized bootstrapping algorithm. A first step provides reliable initial estimates for the RGB camera, and the final step optimizes spatio-temporal parameters such as target pose and inertial measurement unit (IMU) biases. The proposed algorithm is validated through experimental results on both real and simulated sequences and ablation studies on terrestrial and aerial vehicles to estimate the contribution of each sensor to pose estimation accuracy. Github repository: <https://github.com/AbanobSoliman/HICALIB>

Keywords

RGB-D-IMU, calibration, optimization.

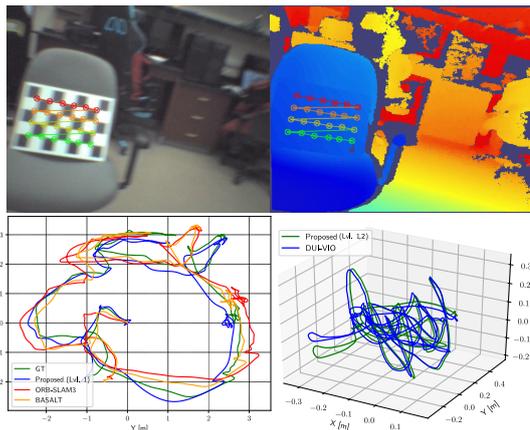


Figure 1: Notre processus de calibration et d'estimation de pose pour la configuration RVB-D-IMU appliqué à la séquence de calibration main-œil VCU-RVI (en haut et en bas à droite) et à la séquence EuRoC V2-01 (en bas à gauche).

1 Introduction et travaux connexes

Une méthode d'odométrie de véhicule autonome efficace repose sur la disponibilité continue des informations de la scène et le véhicule, telles que la structure de la scène et les propriétés physiques du véhicule (position, vitesse ou accélération). Ces propriétés sont mesurées par des capteurs dits extéroceptifs (Caméras/LiDAR/Radar/GPS) et proprioceptifs (IMU / odomètre sur roue). Ajouter une IMU (ou plusieurs IMU comme dans [1]) à une méthode de calibration multi-caméras augmente la complexité du processus d'alignement des poses initiales estimées à une échelle arbitraire avec leurs positions et écarts relatifs réels [2]. Dans les travaux récents [3], les auteurs ont étudié une approche d'optimisation basée sur les graphes qui fusionne les informations GPS et IMU avec celles des caméras stéréo-RVB. Ils montrent une précision d'estimation supérieure, en particulier dans une opération hors-ligne, ce qui est idéal pour les applications de calibration multi-modale (voir Fig. 1).

La multi-modalité est devenue la tendance principale des travaux de calibration les plus récents [4], car une solution d'odométrie multi-modale efficace dépend d'un système

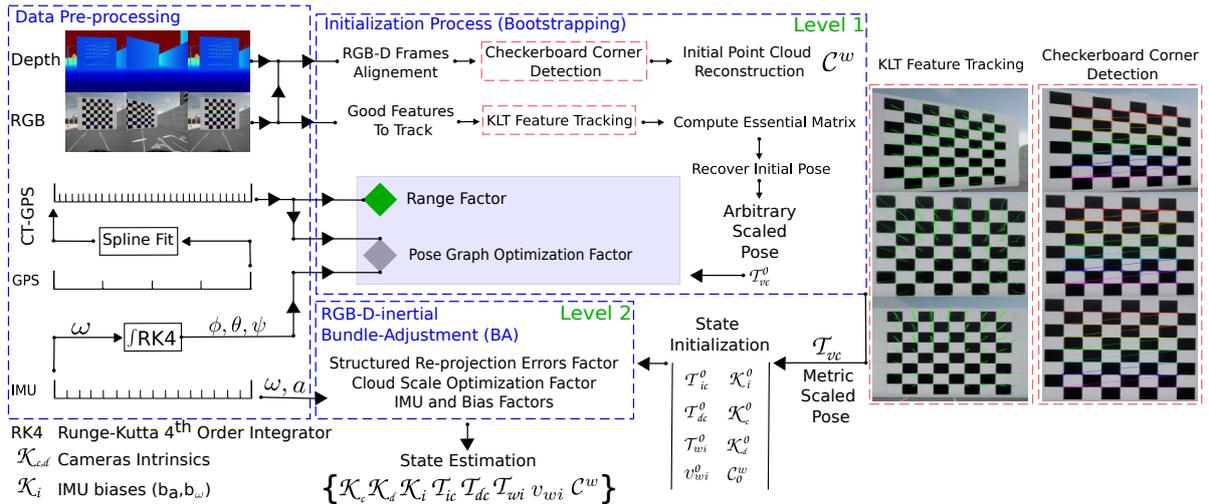


Figure 2: Le pipeline de notre méthode. Le processus avant est une couche de traitement de données initiale après l'acquisition de trames RVB-D alignées. Le processus arrière est la couche de traitement centrale avec deux niveaux d'optimisation.

calibré de manière optimale. Dans ce travail, nous proposons une méthode robuste pour calibrer les paramètres complets du système RVB-D-IMU en tenant compte des performances efficaces en termes de latence, de précision et de robustesse de la configuration. Au cours des dernières années, les algorithmes de calibration RVB-D [5] ont évolué pour intégrer diverses stratégies de correction de profondeur basées sur une étape supplémentaire d'optimisation sur variété. Le travail présenté dans [5] affine l'estimation de la profondeur avec un ajustement de courbe paramétrique, tandis que [6] ajuste le nuage de points sur une sphère. Ajouter une centrale inertielle à une méthode de calibration RVB-D est abordé dans le travail de [7] en utilisant des filtres de Kalman étendus (EKF). Cependant, ces travaux de calibration RVB-D-IMU visent principalement à estimer la pose et à effectuer une calibration extrinsèque IMU/CAM pendant la tâche d'odométrie. Nous proposons une solution fiable de calibration RVB-D-IMU et d'estimation de poses assistée par GPS et nous résumons nos principales contributions en trois points :

- Une nouvelle méthode pour initialiser l'ajustement de faisceaux (*Bundle Adjustment* - BA) visuel-inertiel avec un niveau préalable d'optimisation de graphe de poses basé sur des mesures GPS bruitées à faible débit combinées à des mesures issue d'un gyroscope.
- Un nouveau facteur d'optimisation de l'échelle du nuage de points qui intègre les cartes de profondeur non texturées n'ayant pas de caractéristiques distinctives dans un BA visuel-inertiel, comme toute caméra conventionnelle dans une configuration de vision stéréo par une double réprojection avec estimation de facteurs de distorsion.
- Un algorithme de calibration multi-modale robuste pour la configuration des capteurs RVB-D-IMU avec une méthodologie d'estimation de pose 3D à l'échelle métrique fiable facilement étendue à un algorithme d'odométrie

RVB-D-IMU-GPS multi-modale (voir Fig. 2).

2 Méthodologie

2.1 Odométrie visuelle basée sur le flux

Les coins et leurs caractéristiques correspondantes de la scène sont d'abord extraits avec une taille de bloc de 17 pixels. Pour améliorer la robustesse et la polyvalence de l'odométrie visuelle, nous adoptons la méthode de suivi de caractéristiques basée sur le flux optique Kanade-Lucas-Tomasi (KLT), pour faire correspondre les caractéristiques correspondantes dans une approche de résolution pyramidale à 7 niveaux avec une fenêtre de taille 17×17 pixels. En suivant les caractéristiques les plus robustes et stables sur 10 images consécutives, nous calculons la *matrice essentielle* avec le rejet des valeurs aberrantes des caractéristiques par MAGSAC++ [8]. Bien que RANSAC et MAGSAC++ soient utiles pour estimer les paramètres du modèle à partir de données bruitées, MAGSAC++ offre une amélioration de la précision, de la robustesse et du temps de calcul. La transformation relative entre chaque paire d'images consécutives $\mathcal{T}_{vc} \in SE(3)$ est récupérée ensuite à partir de la *matrice essentielle*, que nous utilisons pour initialiser notre processus d'optimisation de niveau 1 avec le graphe de poses initial en utilisant la transformation à l'échelle arbitraire suivante:

$$\mathcal{T}_{wc} \doteq \mathcal{T}_{wv} \mathcal{T}_{vc}, \quad (1)$$

où $\mathcal{T}_{wv} \in SE(3)$ est la transformation corps rigide entre les repères inertiels de l'IMU/corps (monde) et de la caméra RVB (visuel) w, v respectivement. Lors de l'initialisation, nous supposons qu'il n'y a pas de translation entre les repères de référence IMU-caméra, c'est-à-dire $t_{wv} = [0, 0, 0]^T$. La rotation R_{wv} entre eux est donnée dans la Fig. 3 (b), sachant que le repère de la caméra c est

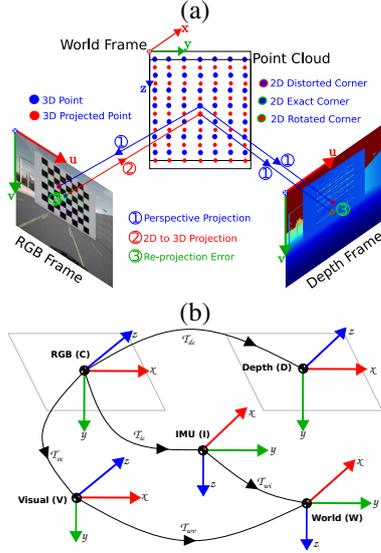


Figure 3: Illustration des facteurs d'erreur de reprojection sur les trames RVB et de profondeur, ainsi que les repères de coordonnées pour tous les capteurs soumis à l'optimisation : (a) erreur de reprojection 3D à 2D et 2D à 3D à 2D pour trianguler le même point 3D de la cible sur les deux trames RVB-D alignées actuellement ; (b) Repère de coordonnées pour tous les capteurs à calibrer par rapport au repère du monde. Pour plus de cohérence : tous les repères suivent la règle de la main droite comme la bibliothèque OpenCV.

son repère inertiel de référence (repère visuel v) coïncident initialement l'un sur l'autre. Jusqu'à cette étape, le mouvement de corps rigide de la caméra RVB \mathcal{T}_{wc} est considéré comme le mouvement de corps rigide au facteur d'échelle de toute la configuration de capteurs multi-modale \mathcal{T}_{wi}^0 .

En parallèle, une détection de coins via une mire en damier est effectuée sur toutes les images de la caméra RVB. Lorsqu'un damier est détecté, une image RVB est considérée comme une image-clé de calibration (*keyframe* - KF). Nous intégrons l'image de profondeur correspondante, synchronisée dans le temps et alignée spatialement (d), pour construire un nuage de points 3D des coins détectés.

2.2 Initialisation robuste de l'optimiseur

Après avoir obtenu les poses de la mire et construit initialement des nuages de points du motif en damier, l'amorçage de l'optimiseur est essentiel pour un processus de calibration fiable. Cette méthode est efficace en termes de complexité, car l'amorçage ne repose que sur des mesures GPS bruitées à faible débit et des données préintégréées issues du gyroscope. Pour résoudre ces problèmes de GPS, nous appliquons une interpolation de B-spline cumulative sur variété [9] pour synthétiser une trajectoire en temps continu (CT) plus lisse en \mathbb{R}^3 à partir des lectures GPS bruitées à faible débit, comme illustré sur la Fig. 4.

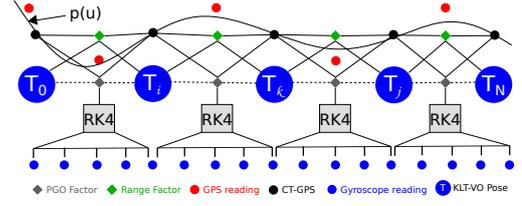


Figure 4: Graphe de facteurs de niveau 1. $p(u)$ est la trajectoire CT-GPS générée à haute fréquence. RK4 est le schéma d'intégration du gyroscope Runge-Kutta d'ordre 4^{ème}. Les lignes pointillées représentent le terme d'erreur ($\hat{T}_i^{-1} \hat{T}_j$) dans l'équation (5) entre deux poses KLT-VO quelconques.

La forme matricielle pour la variété de B-spline cumulative d'ordre $k = n + 1$, où n est le degré de la spline, est modélisée à $t \in [t_i, t_{i+k-1}]$ comme suit:

$$p(u) = p_i + \sum_{j=1}^{k-1} \tilde{B}_j^{(k)} \tilde{u}_j^{(k)} d_j \in \mathbb{R}^3, \quad (2)$$

où $p(u)$ est l'incrément de B-spline en temps continu qui interpole k mesures GPS sur l'unité normalisée de temps $u(t) = (t - t_i)/\Delta t_s - P_n$ où $1/\Delta t_s$ dénote la fréquence de génération de la spline, P_n est le numéro de pose qui contribue au segment de spline actuel $P_n \in [0, \dots, k-1]$. p_i est la mesure de position GPS en temps discret (DT) initiale au moment t_i . $d_j^i = p_{i+j} - p_{i+j-1}$ est le vecteur de différence entre deux données DT-GPS consécutives. $\tilde{B}_j^{(k)}$ est la matrice de mélange de base cumulative et $\tilde{u}_j^{(k)}$ est le vecteur de temps normalisé. Ces éléments sont définis comme suit:

$$\begin{aligned} \tilde{B}_j^{(k)} &= \tilde{b}_{j,n}^{(k)} = \sum_{s=j}^{k-1} b_{s,n}^{(k)}, \\ b_{s,n}^{(k)} &= \frac{C_{k-1}^n}{(k-1)!} \sum_{l=s}^{k-1} (-1)^{l-s} C_k^{l-s} (k-1-l)^{k-1-n}, \\ \tilde{u}_j^{(k)} &= [u^0, \dots, u^{k-1}, u^k]^\top, u \in [0, \dots, 1]. \end{aligned} \quad (3)$$

Notre système d'initialisation assisté par GPS-IMU comprend deux facteurs; le premier facteur, r^p , optimise les 6 degrés de libertés (DDL) de chaque pose, tandis que le second facteur, r^s , optimise les 3 DDL positionnels entre deux poses avec une contrainte de portée.

La fonction objectif de niveau 1 $L^{p,s}$ est modélisée comme suit:

$$L^{p,s} = \arg \min_{\mathcal{T}_{wi}} \left[\sum_{(i,j)}^N \left(\|r^p(i,j)\|_{\Sigma_{i,j}^p}^2 + \|r^s(i,j)\|_{\Sigma_{i,j}^s}^2 \right) \right]. \quad (4)$$

$\Sigma_{i,j}^p, \Sigma_{i,j}^s$ sont les matrices d'information associées à la covariance des lectures GPS, reflétant les bruits des facteurs PGO et Range sur le processus d'estimation du facteur d'échelle global entre deux images RVB-D alignées.

Facteur d'optimisation de graphe de pose (PGO). Le PGO est un facteur 6 DDL qui contrôle l'erreur de pose

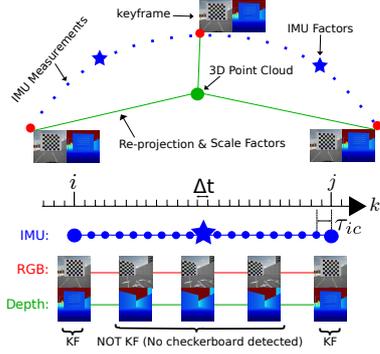


Figure 5: Graphe de facteurs de niveau 2 entre les images clés alignées RVB-D (KF). Ce graphe de facteurs illustre le processus BA non linéaire pour calibrer l'ensemble de capteurs RVB-D-IMU. Δt désigne le pas de temps IMU. τ_{ic} désigne le décalage temporel entre la caméra et l'IMU.

relative entre deux arêtes consécutives i, j et est formulé comme suit:

$$r^p = \left\| \left(\hat{T}_i^{-1} \hat{T}_j \right) \ominus \Delta T_{ij}^{\omega, GPS} \right\|_2, \quad (5)$$

où $\|\cdot\|_2$ est la norme L2, $\hat{T}_i, j \in SE(3)$ est le \mathcal{T}_{wi}^0 estimé à partir du pipeline front-end aux trames i, j . \ominus est la carte logarithmique $SE(3)$ telle que définie dans [10]. $\Delta T_{ij}^{\omega, GPS} \in \mathfrak{se}(3)$, où $\delta p^{GPS}_{ij} = p_j - p_i$ est l'incrément de mesure CT-GPS et $\delta R^{\omega}_{ij} = [\delta\phi, \delta\theta, \delta\psi]$ est l'incrément intégré du gyroscope $\delta R^{\omega}_{ij} = \int_{k=i}^j (\omega_k) dk$ en utilisant la méthode d'intégration Runge-Kutta d'ordre 4^{ème} (RK4) [11] entre les deux images clés i, j .

Facteur de portée. Le facteur de portée limite la dérive visuelle en bout de chaîne et maintient l'échelle métrique globale sous contrôle dans une plage de valeurs définie par le signal GPS et est formulé comme suit:

$$r^s = \left\| \left\| \hat{t}_j - \hat{t}_i \right\|_2 - \left\| p_j^{GPS} - p_i^{GPS} \right\|_2 \right\|_2, \quad (6)$$

où la norme euclidienne interne $\|\cdot\|_2$ est entre les vecteurs de translation $\hat{t}_i, j, p^{GPS}_{i, j} \in \mathbb{R}^3$ de deux poses consécutives du front-end (KLT-VO) et des signaux CT-GPS, respectivement.

2.3 RVB-D-IMU Local Bundle Adjustment

Pour estimer les paramètres de calibration du RVB-D-IMU, nous fusionnons les coins de damier suivis et les nuages de points avec le facteur de mesures pré-intégrées de l'IMU proposé dans [12]. La Fig. 5 montre notre approche de fenêtre glissante. Le BA local est effectué sur tous les coins 2D collectés \mathcal{B} au sein de leur nuage de points 3D correspondant \mathcal{C} entre deux images clés alignées de la caméra RVB c et de la caméra de profondeur d i, j , et les lectures de l'IMU I entre les deux. Notre fonction objective de minimisation de l'ajustement local de paquets $L^{c,d,I}$ est définie par:

$$L^{c,d,I} = \arg \min_{\mathcal{X}} \left[\sum_{(i,j)}^N \rho_{\mathcal{H}}(\|r^I(i,j)\|_{\Sigma_i^I}^2) + \sum_{C_i}^N \sum_{\mathcal{B}_i}^M \left(\rho_{\mathcal{H}}(\|r^c(\mathcal{B}_i|C_i)\|_{\Sigma_i^c}^2) + \rho_C(\|r^d(\mathcal{B}_i|C_i)\|_{\Sigma_i^d}^2) \right) \right], \quad (7)$$

avec \mathcal{X} , les états d'optimisation du BA local complet, qui sont définis comme suit:

$$\begin{aligned} \mathcal{X} = & \{ \mathcal{K}_c, \mathcal{K}_d, \mathcal{K}_i, \mathcal{T}_{ic}, \mathcal{T}_{dc}, \mathcal{T}_{wi}, v_{wi}, C^w \}, \\ \mathcal{K}_c, \mathcal{K}_d = & [f_x, f_y, c_x, c_y, k_1, k_2, p_1, p_2, k_3, \lambda] \in \mathbb{R}^{10}, \\ \mathcal{K}_i^k = & [\tau_{ic}, b^\omega, b^a] \in \mathbb{R}^7, \forall k \in [0, N], \\ \mathcal{T}_{ic}, \mathcal{T}_{dc}, \mathcal{T}_{wi} = & [R_{ic} | t_{ic}, R_{dc} | t_{dc}, R_{wi} | t_{wi}] \in SE(3), \\ C_k^w = & [X^w, Y^w, Z^w] \in \mathbb{R}^3, \forall k \in [0, N], \end{aligned} \quad (8)$$

où $\mathcal{K}_c, \mathcal{K}_d$ sont les paramètres intrinsèques contenant les longueurs focales des caméras f_x, f_y , les centres focaux c_x, c_y , les coefficients de distorsion radiale-tangentielle $k_1, 2, 3, p_1, 2$ et le facteur d'échelle de nuage λ . $\mathcal{T}_{ic}, \mathcal{T}_{dc}$ sont les transformations rigides extrinsèques entre les capteurs. Alors que les paramètres spatio-temporels comprennent la structure de la scène C^w , la pose métrique du corps \mathcal{T}_{wi} , la vitesse v_{wi} par rapport aux coordonnées du monde, $\tau_{ic}, [sec]$ est le décalage temporel IMU-caméra [13] et $b^\omega \in \mathbb{R}^3, b^a \in \mathbb{R}^3$ sont les biais du gyroscope et de l'accéléromètre, respectivement. N, M sont le nombre de trames de calibration et d'observations de coin, respectivement. r^I, r^c, r^d sont les facteurs de l'IMU, de la reprojection du coin et de l'échelle du nuage, respectivement. $\Sigma^I, j, \Sigma^c, \Sigma^d$ sont les matrices d'information associées aux lectures de l'IMU I , aux coins détectés \mathcal{B} et au bruit de covariances de l'échelle du nuage reconstruit \mathcal{C} . ρ est la fonction de perte définie par la norme de Huber [14] $\rho_{\mathcal{H}}$ pour r^I, r^c et la norme de Cauchy [15] ρ_C pour r^d .

Facteur d'erreurs de reprojection structurées. Nous appliquons le modèle de caméra à trou d'épingle RVB avec des coefficients de distorsion radiale-tangentielle avec la matrice des paramètres intrinsèques \mathcal{K}_c . Comme illustré dans la Fig. 3 (a), nous considérons un nuage de points 3D construit C_k^w en utilisant la trame k^{th} alignée de la caméra de profondeur avec la trame clé RVB courante k . Nous avons $H \times W$ observations de caractéristiques pour chaque échiquier, représentant les coins détectés de la trame clé $\mathcal{B}_k^c[u, v]$.

Il y a un facteur pour chaque coin détecté sur l'image clé actuelle k qui minimise l'erreur entre la position de ce coin $\mathcal{B}_k^c[u, v]$ et la réprojection du point 3D correspondant du nuage $C_k^w(u, v)$ sur l'image clé $k^{ième}$ après distorsion $\hat{\mathcal{B}}_k^c[\hat{u}, \hat{v}]$. Ce facteur est défini par:

$$r^c = \left\| \mathcal{B}_k^c[u, v] - \hat{\mathcal{B}}_k^c[\hat{u}, \hat{v}] \right\|_2. \quad (9)$$

En appliquant le modèle de distorsion radial-tangentielle de la caméra sténopé [16] pour calculer la position du pixel déformé du point 3D re-projeté sur l'image courante $\hat{\mathcal{B}}_k^c[\hat{u}, \hat{v}]$, nous obtenons:

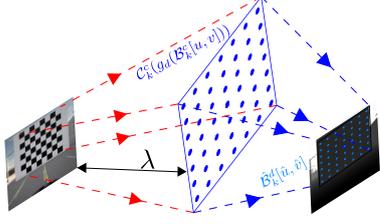


Figure 6: Illustration de la projection 2D-3D-2D des points caractéristiques du damier $H \times W = 7 \times 7$ de l'image RVB vers le nuage de points puis vers l'image de profondeur. λ est le facteur de correction pour les intrinsèques de la caméra RVB permettant d'estimer le facteur d'échelle du nuage de manière optimale.

$$\begin{aligned}
C_k^c(u, v) &= \mathcal{T}_{ic}^{-1} \mathcal{T}_{wi}^{-1} C_k^w(u, v) = [X_k^c, Y_k^c, Z_k^c], \\
\bar{u} &= X_k^c / Z_k^c + c_x / f_x, \quad \bar{v} = Y_k^c / Z_k^c + c_y / f_y, \\
r^2 &= \bar{u}^2 + \bar{v}^2, \\
\hat{u} &= f_x (\bar{u} (1 + k_1 r^2 + k_2 r^4 + k_3 r^6 \\
&\quad + 2p_1 \bar{v}) + p_2 (r^2 + 2\bar{u}^2)), \\
\hat{v} &= f_y (\bar{v} (1 + k_1 r^2 + k_2 r^4 + k_3 r^6 \\
&\quad + 2p_2 \bar{u}) + p_1 (r^2 + 2\bar{v}^2)).
\end{aligned} \tag{10}$$

Facteur d'optimisation de l'échelle du nuage. Ce facteur est modélisé pour fusionner les caractéristiques de coin des images RVB avec les cartes de profondeur non texturées afin de bénéficier des avantages des deux capteurs en minimisant l'erreur entre la re-projection déformée du point 3D du nuage $C_k^w(u, v)$ sur la $k^{\text{ème}}$ image de profondeur $\hat{\mathcal{B}}_k^d[\hat{u}, \hat{v}]$ et l'observation actuelle de la caractéristique de coin $g_d(\mathcal{B}_k^c[u, v])$ par rapport à celle-ci.

L'efficacité de ce facteur provient de l'hypothèse que la suppression de la distorsion de l'image de profondeur permettra, en retour, de dédistordre les coordonnées planaires du nuage de points $C_k^d[X_k^d, Y_k^d]$. Dans la Fig. 6, nous appliquons l'échelle du nuage λ (connue sous le nom de profondeur inverse) pour optimiser les longueurs focales de la caméra RVB avec la 3^{ème} coordonnée du nuage $C_k^d[Z_k^d]$ qui est optimisée au sein du modèle de calibration conjoint, en connaissant l'échelle métrique de la pose. Ce facteur est défini par:

$$r^d = \|g_d(\mathcal{B}_k^c[u, v]) - \hat{\mathcal{B}}_k^d[\hat{u}, \hat{v}]\|_2, \tag{11}$$

où $\hat{\mathcal{B}}_k^d[\hat{u}, \hat{v}]$ suit le même modèle dans l'équation (10) en remplaçant $C_k^c(u, v)$ par $C_k^d(u, v) = \mathcal{T}dc C_k^c(u, v)$. $g_d(\cdot)$ est une fonction de double re-projection avec distorsion, qui **premièrement** projette l'observation $\mathcal{B}_k^c[u, v]$ sur le nuage de points 3D $C_k^c(\mathcal{B}_k^c[u, v])$ comme illustré par la flèche rouge numérotée (2) dans la Fig. 3 (a) en utilisant la transformation de corps rigide $\mathcal{T}wc = \mathcal{T}wi\mathcal{T}ic$ des coordonnées c vers w avec la formule suivante:

$$C_k^c(\mathcal{B}_k^c[u, v]) = R_{wc}(\lambda \mathcal{X}_c^{-1} \mathcal{B}_k^c[u, v]) + t_{wc}. \tag{12}$$

Ensuite, **deuxièmement**, fait tourner $C_k^c(\mathcal{B}_k^c[u, v])$ en $C_k^d(\mathcal{B}_k^c[u, v])$ en utilisant $\mathcal{T}dc$, et **finalemt**, re-projette et déforme ce point doublement tourné sur la trame de profondeur $C_k^d(\mathcal{B}_k^c[u, v])$ en utilisant le même modèle dans l'équation (10).

Facteurs IMU. Les facteurs de pré-intégration IMU entre deux images clés consécutives i, j sont définis dans [12] par:

$$\begin{aligned}
r^I &= [\Delta R_{i,j}, \Delta v_{i,j}, \Delta p_{i,j}, \Delta b_{i,j}^{0,a}] \in \mathbb{R}^{15}, \\
r_{\Delta R_{i,j}}^I &= \log((\Delta \tilde{R}_{i,j})^\top R_i^\top R_j), \\
r_{\Delta v_{i,j}}^I &= R_i^\top (v_j - v_i - g \Delta t_{i,j}) - \Delta \tilde{v}_{i,j}, \\
r_{\Delta p_{i,j}}^I &= R_i^\top (t_j - t_i - v_i \Delta t_{i,j} - \frac{1}{2} g \Delta t_{i,j}^2) - \Delta \tilde{p}_{i,j}, \\
r_{\Delta b_{i,j}}^I &= \|b_j^0 - b_i^0\|_2 + \|b_j^a - b_i^a\|_2.
\end{aligned} \tag{13}$$

où $\Delta \tilde{R}_{i,j}, \Delta \tilde{v}_{i,j}, \Delta \tilde{p}_{i,j}$ sont les incréments pré-intégrés de rotation, de vitesse et de translation. Toutes ces dérivations d'incrément de pré-intégration sur la variété, ainsi que la propagation de covariance $\Sigma^I_{i,j}$, sont données dans le matériel supplémentaire de [12], et pour une meilleure lisibilité, nous écrivons $R_{i,j}, t_{i,j}, v_{i,j}$ au lieu de $[R_{wi}, t_{wi}, v_{wi}]$.

3 Évaluation

Nous évaluons les performances de notre méthode sur deux applications : la calibration RVB-D-IMU et l'estimation de pose assistée par GPS. En utilisant les API d'acquisition de données basées sur CARLA du benchmark IBIS-Cape [17], nous collectons trois séquences de calibration simulées avec une vaste gamme de tailles. De plus, la validation d'algorithme sur des séquences simulées facilite le changement de réglages vers différentes configurations de capteurs pour une validation robuste de tous les cas limites et fournit une référence pour la plupart des paramètres système. En outre, pour une évaluation dans le monde réel, nous évaluons notre méthode de calibration sur la séquence de calibration main-œil avec damier RVB-D-IMU du benchmark VCU-RVI [18]. Enfin, nous menons des études d'ablation sur les séquences IBIS-Cape (Véhicule) et EuRoC [19] (MAV) pour évaluer la contribution de chaque capteur dans une configuration RVB-D-IMU-GPS à l'exactitude de l'estimation de pose pour une navigation à long terme fiable.

3.1 Application I : Calibration RVB-D-IMU

Pour les séquences VCU-RVI et CARLA, les valeurs initiales des matrices intrinsèques des caméras sont fixées à $W/2$ pour c_x, f_x , $H/2$ pour c_y, f_y et zéros pour les distorsions radiales-tangentielles. La valeur initiale de λ est fixée à 0.1643, qui est la densité de pixels des caméras CARLA. Pour l'initialisation des paramètres extrinsèques \mathcal{T}_{ic}^0 et \mathcal{T}_{dc}^0 , nous fixons la partie translation à zéros et la matrice de rotation est définie comme indiqué sur la Fig. 3 (b). Comme la séquence portative VCU-RVI peut fournir une excitation IMU suffisante mais sans données GPS disponibles, l'amorçage du système de calibration est effectué par la méthode traditionnelle basée sur l'IMU [2].

Nous validons notre nouveau facteur d'optimisation globale de nuage sur deux critères : le nuage de points estimé après optimisation et l'estimation de la distorsion du cadre de profondeur en tant qu'indicateur de correction de la profondeur. La Fig. 7 montre que le nuage optimisé converge vers une distribution normale dont la moyenne est la position exacte dans le monde de simulation à 60 m, qui est à l'emplacement du damier comme indiqué sur la Fig. 8. Le tableau 1 montre les valeurs considérablement élevées pour les coefficients de distorsion du cadre de profondeur, indiquant l'effet de notre facteur sur l'indistorsion planaire du nuage.

En utilisant Kalibr [1] comme référence pour les intrinsèques de la caméra RVB pour les séquences CARLA et VCU-RVI, nous évaluons la qualité de l'estimation de notre optimiseur dans le Tab. 1. Puisque l'échelle de la carte λ est un paramètre d'optimisation de la caméra RVB basé sur la contrainte de liaison géométrique RVB-D introduite dans l'équation (12), les estimations de la longueur focale nécessitent une correction d'échelle en utilisant : $f_{x,y}^{corr} = f_{x,y}^{est} * \lambda$. Pour la séquence hand-eye VCU-RVI, nous remarquons que le facteur d'échelle du nuage se rapproche de la valeur 1, ce qui indique que le nuage de points initial est construit avec un capteur de profondeur de haute qualité.

Dans le Tab. 2, nous montrons la performance optimale de notre optimiseur pour estimer les paramètres extrinsèques inter-capteurs par rapport aux valeurs GT. Comparé à la référence, notre optimiseur estime efficacement la rotation et la translation inter-capteurs dans le cas des capteurs RVB-D. Pour les paramètres extrinsèques de la caméra IMU et contrairement aux rotations, la translation rigide de la caméra IMU dépend principalement des valeurs initiales définies dans l'optimiseur. Afin d'estimer les valeurs optimales pour la partie translation, plusieurs expériences doivent être réalisées avec des zéros comme conditions initiales avec de grands ensembles de données. En fonction de la qualité des valeurs de calibration IMU fixes, toutes les expériences convergeront vers des valeurs relatives, comme le montre le Tab. 2.

3.2 Application II: Estimation de la position assistée par GPS

Deux études d'ablation sont réalisées pour évaluer la contribution du capteur GPS à la précision de l'estimation de la pose lorsque les informations de profondeur sont disponibles ou non disponibles. Les métriques d'évaluation VIO standard [23] sont utilisées pour l'évaluation : l'erreur de trajectoire absolue moyenne à racine carrée (RMS ATE_p [m]) et l'erreur de pose relative (RPE_p [m]).

Étude d'ablation sur un véhicule terrestre simulé. Dans la première étude d'ablation, nous évaluons la performance de notre estimation de pose intégrant la profondeur avec l'aide de GPS comparativement aux derniers systèmes VIO de pointe qui n'utilisent pas de mesures GPS dans

Table 1: Estimation des paramètres intrinsèques de la configuration des capteurs RVB-D-IMU. Comme le simulateur CARLA ne fournit pas de valeurs intrinsèques exactes, les valeurs GT pour les intrinsèques de la caméra RVB sont obtenues avec Kalibr [1]. KF : nombre de trames clés. TL : Longueur de la trajectoire de la séquence. D : Durée de la séquence. * désigne une valeur calculée à partir des spécifications de la caméra RVB-D Structure Core (SC) avec un champ de vision en profondeur (FOV) de 70°. ** désigne une valeur issue de la fiche technique de l'IMU Bosch BMI085.

Paramètre	Simulateur CARLA (IBIScape [17])				VCU-RVI [18]	
	S1	S2	S3	GT	main-oeil	GT
Spécifications						
RVB		20 Hz - 1024 × 1024 px			30 Hz - 640 × 480 px	
Depth		20 Hz - 1024 × 1024 px			30 Hz - 640 × 480 px	
IMU		6-axis acc/gyro @200Hz			6-axis acc/gyro @100Hz	
#KF	353	994	2126	-	1118	-
TL[m]	122.06	345.42	737.88	-	11.16	-
D[sec]	17.640	49.730	106.29	-	46.59	-
Caméra RVB						
λ, f_x	164.01	122.71	148.42	151.51	375.67	459.36
λ, f_y	163.30	122.22	149.39	151.89	398.44	459.76
c_x	498.89	506.21	507.59	510.01	315.48	332.69
c_y	514.01	515.49	518.61	510.71	289.64	258.99
k_1	-5.10e-3	-6.20e-3	-6.15e-3	2.42e-5	-1.62e-2	-2.98e-1
k_2	-1.95e-3	-1.96e-3	-2.07e-3	2.89e-6	-3.62e-3	9.22e-2
p_1	-1.25e-3	-1.96e-3	-8.31e-4	1.71e-4	-2.31e-3	-1.19e-4
p_2	-3.20e-3	-2.27e-3	-3.53e-3	-3.22e-5	-1.09e-2	-7.46e-5
k_3	-8.16e-4	-8.70e-4	-8.64e-4	0.0	-7.84e-4	-
λ	0.3581	0.2819	0.3432	-	0.9831	-
Caméra de profondeur						
f_x	511.42	511.51	511.51	512.0	456.82	457.01*
f_y	511.91	511.83	511.82	512.0	456.06	457.01*
c_x	512.20	512.22	512.30	512.0	333.29	320.0*
c_y	511.81	512.01	512.02	512.0	259.17	240.0*
k_1	-3.53e-2	-3.37e-2	-3.54e-2	-	-5.74e-2	-
k_2	-5.60e-3	-6.20e-3	-6.25e-3	-	-9.07e-3	-
p_1	-3.41e-2	-3.22e-2	-3.29e-2	-	-4.13e-2	-
p_2	-3.93e-2	-3.50e-2	-3.82e-2	-	-6.09e-2	-
k_3	-1.10e-3	-1.45e-3	-1.38e-3	-	-2.98e-4	-
Capteur IMU						
τ_{ic}	4.986e-3	4.989e-3	4.998e-3	5e-3	4.473e-3	-
b_x^{ω}	-7.549e-3	-2.242e-2	-4.907e-3	-2.383e-3	1.512e-4	9.69e-5**
b_y^{ω}	-3.283e-2	3.813e-2	-2.054e-2	-3.364e-3	9.337e-5	9.69e-5**
b_z^{ω}	8.151e-2	2.659e-2	-2.540e-2	1.555e-3	-2.967e-4	9.69e-5**
b_x^t	0.109	-0.062	0.147	-0.951	-5.704e-4	-
b_y^t	-0.707	-1.069	-0.091	-0.691	6.757e-4	-
b_z^t	-1.926	-2.295	-2.364	0.183	-9.304e-4	-

Table 2: Estimation des paramètres extrinsèques pour les séquences de calibration IBIScape (S1, S2, S3) et VCU-RVI (main-oeil).

Paramètre	t_x [m]	t_y [m]	t_z [m]	q_x	q_y	q_z	q_w
RVB-D (T_{ic})	S1	4.95e-3	0.017	0.037	-0.037	-0.022	0.030
	S2	5.47e-3	0.020	0.065	-0.041	0.005	0.019
	S3	9.10e-3	0.018	0.065	-0.036	-0.010	0.025
	GT	0.0	0.020	0.060	0.0	0.0	0.0
	hand-eye	-0.103	0.003	0.018	0.041	0.081	0.009
GT	-0.100	0.0	0.0	0.0	0.0	0.0	
RVB-IMU (T_{ic})	S1	-0.806	0.154	-0.308	0.493	0.507	0.499
	S2	-0.854	-0.057	0.006	0.503	0.495	0.501
	S3	-0.808	-0.028	-0.102	0.503	0.501	0.499
	GT	-0.800	0.0	0.0	0.500	0.500	0.500
	hand-eye	0.077	0.020	-0.041	0.699	-0.713	-0.009
GT	-0.008	0.015	-0.011	0.708	-0.706	0.001	

Table 3: Étude d'ablation sur la contribution du capteur GPS à la précision du système lorsque les informations de profondeur sont disponibles.

Méthode	IBIScape [17] (RPE _p ($\mu \pm \sigma$) [m])			Moyenne
	S1	S2	S3	
DUI-VIO [20]	0.115±0.113	0.115±0.114	0.120±0.119	0.117±0.115
BASALT [21]	0.084±0.084	0.052±0.051	0.026±0.026	0.054±0.054
ORB-SLAM3 [22]	0.028±0.013	0.073±0.034	0.031±0.028	0.044±0.025
Proposée (Lv1.1+2)	0.016±0.019	0.025±0.030	0.018±0.025	0.020±0.025

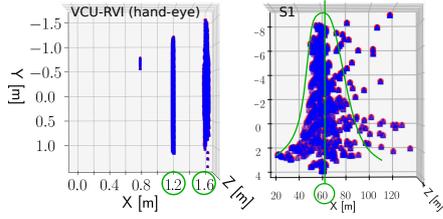


Figure 7: La reconstruction de nuage de points 3D en vue de dessus de la cible ; (gauche) nuages de points initialement construits de VCU-RVI, (droite) nuage de points optimisé de CARLA. Les points bleus avec un contour rouge représentent l'emplacement 3D des coins de l'échiquier. La courbe verte représente la convergence de la distribution normale du nuage de points après optimisation. Les cercles verts représentent une valeur moyenne de profondeur du nuage de points.

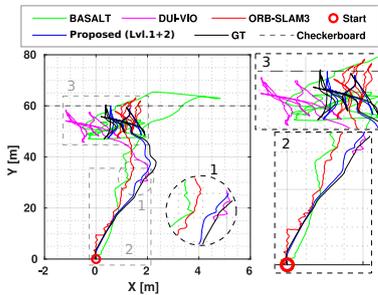


Figure 8: Évaluation de l'estimation de pose de notre méthode par rapport à ORB-SLAM3, BASALT et DUI-VIO sur la séquence S1. Différentes échelles d'axes pour montrer des détails fins.

leurs estimations. Nous comparons la précision de notre estimation de pose RVB-D-IMU aidée par GPS à celle des systèmes ORB-SLAM3 (RVB-D) [22], BASALT (2×RVB-IMU) [21] et DUI-VIO (RVB-D-IMU) [20] en utilisant les séquences VCU-RVI et CARLA.

Lors de l'évaluation du système DUI-VIO [20], nous avons remarqué une erreur d'initialisation avec la séquence S1 jusqu'à ce que le système s'initialise avec succès à la fin du ralentisseur à environ 30 m comme le montre la Fig. 8 (#1). Ce problème d'initialisation n'est pas observé avec la séquence d'étalonnage à la main VCU-RVI en raison de ses mouvements combinés complexes (voir Fig. 1 (droite)). Les séquences (S2, S3) sont simulées avec un mouvement combiné élevé pour assurer une couverture optimale de la grille pour toutes les trames de la caméra RVB-D. Le mouvement complexe génère une excitation suffisante de l'IMU pour initialiser BASALT et DUI-VIO.

Dans notre analyse de la Tab. 3, les résultats quantitatifs montrent une performance supérieure de notre méthode par rapport aux autres approches. En effet, l'erreur d'estimation de pose est réduite de 54,55%, 62,96% et

Table 4: Étude d'ablation sur la contribution du capteur GPS à la précision du système lorsque les informations de profondeur ne sont pas disponibles. * dénote le suivi de caractéristiques sur 5 images consécutives au lieu de 10 en raison du mouvement rapide du MAV. + dénote la seule référence basée sur l'apprentissage dans le tableau et la seule méthode incorporant des nuages de points LiDAR. V, I, G : Vision, IMU et GPS.

Méthode	EuRoC [19] (RMS ATE _p [m])						Moy.	
	V1-01	V1-02	V1-03	V2-01	V2-02	V2-03		
Mono-VI	OKVIS [24]	0.090	0.200	0.240	0.130	0.160	0.290	0.185
	ROVIO [25]	0.100	0.100	0.140	0.120	0.140	0.140	0.123
	VINS-Mono [2]	0.047	0.066	0.180	0.056	0.090	0.244	0.114
	OpenVINS [26]	0.056	0.072	0.069	0.098	0.061	0.286	0.107
	CodeVIO+ [27]	0.054	0.071	0.068	0.097	0.061	0.275	0.104
Stereo-VI	VINS-Fusion [28]	0.076	0.069	0.114	0.066	0.091	0.096	0.085
	BASALT [21]	0.040	0.020	0.030	0.030	0.020	0.050	0.032
	Kimera [29]	0.050	0.110	0.120	0.070	0.100	0.190	0.107
	ORB-SLAM3 [22]	0.038	0.014	0.024	0.032	0.014	0.024	0.024
Mono-VII/G	CT (V+I+G) [30]	0.024	0.014	0.011	0.012	0.010	0.010	0.014
	CT (V+G) [30]	0.011	0.013	0.012	0.009	0.008	0.012	0.011
	CT (I+G) [30]	0.062	0.102	0.117	0.112	0.164	0.363	0.153
	DT (V+I+G) [30]	0.016	0.024	0.018	0.009	0.018	0.033	0.020
	DT (V+G) [30]	0.010	0.025	0.024	0.010	0.012	0.029	0.018
	DT (I+G) [30]	0.139	0.137	0.138	0.138	0.138	0.139	0.138
Proposée (Lvl.1)	0.008	0.017*	0.023*	0.008	0.022	0.025*	0.017	

82,91% par rapport à ORB-SLAM3, BASALT et DUI-VIO, respectivement. Cela est dû à notre méthode de bootstrapping rapide aidée par GPS qui diminue l'accumulation d'erreurs de pose relative avec le temps.

Étude d'ablation sur un véhicule aérien réel. Nous avons mené une autre étude d'ablation pour valider davantage les performances de notre méthode d'estimation de pose dans une application réelle. Les expériences de cette étude ont été menées sur le jeu de données EuRoC MAV [19] intégrant des capteurs RVB-IMU et ont été comparées au système SLAM basé sur les signaux GPS en temps continu et en temps discret (CT/DT) proposé dans [30]. Étant donné qu'une comparaison avec la technique concurrente [30], combinant les signaux GPS calculés à partir des mesures du système Vicon, met mieux en évidence les résultats de cette recherche d'ablation, nous avons choisi les six séquences de la salle Vicon identiques au benchmark EuRoC qu'ils ont utilisées dans leur évaluation.

L'analyse quantitative dans le tableau 4 montre que nos estimations de niveau 1, sans information de profondeur, peuvent efficacement estimer une trajectoire à échelle métrique qui peut amorcer le niveau 2 et surpasser d'autres systèmes VIO bien établis en termes de précision. Nous remarquons également une amélioration de la précision de l'estimation avec l'ajout d'une modalité de capteur (IMU/GPS), à condition qu'au moins un capteur visuel soit présent dans le système. Une autre conclusion est qu'un GPS peut être suffisant avec le capteur optique pour obtenir une estimation de trajectoire fiable dans un schéma de fusion fortement couplé. Pour un schéma de fusion lâchement couplé (proposé Lvl.1), l'ajout d'un gyroscope augmente la confiance de l'optimiseur à converger vers des valeurs

raisonnables.

4 Conclusion

Cette étude propose une méthode de référence pour l'étalonnage robuste des paramètres intrinsèques et extrinsèques RVB-D-IMU. Nous présentons tout d'abord une méthode de bootstrap optimisée RVB-GPS-Gyro qui estime la pose de la cible à l'échelle métrique de manière fiable pour le processus d'étalonnage. Nous avons ensuite défini un facteur d'échelle de nuage pour une carte de profondeur RVB-D alignée spatialement et non texturée, qui estime son échelle en incorporant l'incertitude initiale du nuage reconstruit. Dans les travaux futurs, il serait intéressant d'incorporer des situations où les limitations des capteurs GPS ne peuvent pas être simulées, telles que les effets multi-trajets sur l'optimiseur. Enfin, il sera essentiel de généraliser davantage le problème d'optimisation de BA pour étendre la capacité d'étalonnage de l'algorithme à plusieurs IMU avec plusieurs capteurs de vision (RVB et profondeur).

References

- [1] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart, "Extending kalibr: Calibrating the extrinsics of multiple IMUs and of individual axes," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4304–4311, IEEE, 2016.
- [2] T. Qin, P. Li, and S. Shen, "VINS-Mono: a robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.
- [3] A. Das, J. Elfring, and G. Dubbelman, "Real-time vehicle positioning and mapping using graph optimization," *Sensors*, vol. 21, no. 8, p. 2815, 2021.
- [4] Y. Zhang, W. Liang, S. Zhang, X. Yuan, X. Xia, J. Tan, and Z. Pang, "High-precision Calibration of Camera and IMU on Manipulator for Bio-inspired Robotic System," *Journal of Bionic Engineering*, vol. 19, no. 2, pp. 299–313, 2022.
- [5] Y. Zhou, D. Chen, J. Wu, M. Huang, and Y. Weng, "Calibration of RGB-D camera using depth correction model," *Journal of Physics: Conference Series*, vol. 2203, no. 1, p. 012032, 2022.
- [6] H. Liu, D. Qu, F. Xu, F. Zou, J. Song, and K. Jia, "Approach for accurate calibration of RGB-D cameras using spheres," *Opt. Express*, vol. 28, pp. 19058–19073, Jun 2020.
- [7] C. Chu and S. Yang, "Keyframe-based RGB-D visual-inertial odometry and camera extrinsic calibration using Extended Kalman Filter," *IEEE Sensors Journal*, vol. 20, no. 11, pp. 6130–6138, 2020.
- [8] D. Barath, J. Noskova, M. Ivaschekin, and J. Matas, "MAGSAC++, a fast, reliable and accurate robust estimator," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [9] A. Soliman, H. Hadj-Abdelkader, F. Bonardi, S. Bouchafa, and D. Sidibé, "Mav localization in large-scale environments: A decoupled optimization/filtering approach," *Sensors*, vol. 23, no. 1, 2023.
- [10] Y. Wang and G. S. Chirikjian, "Nonparametric second-order theory of error propagation on motion groups," *The International journal of robotics research*, vol. 27, no. 11-12, pp. 1258–1273, 2008.
- [11] L. Zheng and X. Zhang, "Chapter 8 - numerical methods," in *Modeling and Analysis of Modern Fluid Problems* (L. Zheng and X. Zhang, eds.), Mathematics in Science and Engineering, pp. 361–455, Academic Press, 2017.
- [12] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual-inertial odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2016.
- [13] R. Voges and B. Wagner, "Timestamp offset calibration for an IMU-Camera system under interval uncertainty," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 377–384, 2018.
- [14] P. J. Huber, "Robust estimation of a location parameter," in *Breakthroughs in statistics*, pp. 492–518, Springer, 1992.
- [15] M. J. Black and P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.
- [16] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [17] A. Soliman, F. Bonardi, D. Sidibé, and S. Bouchafa, "IBISCape: A simulated benchmark for multi-modal SLAM systems evaluation in large-scale dynamic environments," *Journal of Intelligent & Robotic Systems*, vol. 106, p. 53, Oct 2022.
- [18] H. Zhang, L. Jin, and C. Ye, "The VCU-RVI Benchmark: Evaluating Visual Inertial Odometry for Indoor Navigation Applications with an RGB-D Camera," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6209–6214, 2020.
- [19] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The EuRoC micro aerial vehicle datasets," *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, 2016.
- [20] H. Zhang and C. Ye, "DUI-VIO: Depth uncertainty incorporated visual inertial odometry based on an RGB-D camera," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5002–5008, IEEE, 2020.
- [21] V. Usenko, N. Demmel, D. Schubert, J. Stückler, and D. Cremers, "Visual-inertial mapping with non-linear factor recovery," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 422–429, 2019.
- [22] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. Montiel, and J. D. Tardós, "OrbSLAM3: An accurate open-source library for visual, visual-inertial, and multimap slam," *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874–1890, 2021.
- [23] W. Chen, G. Shang, A. Ji, C. Zhou, X. Wang, C. Xu, Z. Li, and K. Hu, "An Overview on Visual SLAM: From Tradition to Semantic," *Remote Sensing*, vol. 14, no. 13, 2022.
- [24] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual-inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.
- [25] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust visual inertial odometry using a direct EKF-based approach," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 298–304, 2015.
- [26] P. Geneva, K. Eickenhoff, W. Lee, Y. Yang, and G. Huang, "OpenVINS: a research platform for visual-inertial estimation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4666–4672, 2020.
- [27] X. Zuo, N. Merrill, W. Li, Y. Liu, M. Pollefeys, and G. P. Huang, "CodeVIO: visual-inertial odometry with learned optimizable dense depth," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 14382–14388, 2021.
- [28] T. Qin, J. Pan, S. Cao, and S. Shen, "A general optimization-based framework for local odometry estimation with multiple sensors," *arXiv preprint arXiv:1901.03638*, 2019.
- [29] A. Rosinol, M. Abate, Y. Chang, and L. Carlone, "Kimera: an open-source library for real-time metric-semantic localization and mapping," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1689–1696, IEEE, 2020.
- [30] G. Cioffi, T. Cieslewski, and D. Scaramuzza, "Continuous-time vs. discrete-time vision-based SLAM: A comparative study," *IEEE Robotics Autom. Lett.*, vol. 7, no. 2, pp. 2399–2406, 2022.