



HAL
open science

Estimation initiale de la pose avec un échantillonnage aléatoire mais pondéré de paires de triplets redondants d'images

Charles Villard, Ewelina Rupnik, Marc Pierrot-Deseilligny

► **To cite this version:**

Charles Villard, Ewelina Rupnik, Marc Pierrot-Deseilligny. Estimation initiale de la pose avec un échantillonnage aléatoire mais pondéré de paires de triplets redondants d'images. ORASIS 2023, Laboratoire LIS, UMR 7020, May 2023, Carqueiranne, France. hal-04219524

HAL Id: hal-04219524

<https://hal.science/hal-04219524>

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.

Estimation initiale de la pose avec un échantillonnage aléatoire mais pondéré de paires de triplets redondants d'images

Charles Villard¹

Ewelina Rupnik²

Marc Pierrot-Deseilligny³

¹ LRE, EPITA, 94270 Le Kremlin-Bicêtre

^{1 2 3} LASTIG, Univ Gustave Eiffel, ENSG, IGN, 94160 Saint-Mande

¹ charles.villard@epita.fr

² ewelina.rupnik@ign.fr

³ marc.pierrot-deseilligny@ensg.eu

Résumé

Dans le processus de Structure from Motion, l'initialisation est une étape cruciale qui peut empêcher une reconstruction fiable de la scène. Cette étude se focalise sur l'utilisation de graphes de triplets issus de l'appariement des images pour calculer l'orientation des vues à l'aide d'arbres de couverture aléatoires. Un score est calculé en effectuant de nombreux tirages et en pondérant le graphe de triplets. En fin de compte, l'arbre à couverture minimale sur le graphe de triplet permet de construire une orientation initiale convenable sur des cas simple, des pistes sont discutées pour les cas les plus complexes.

Mots Clef

Photogrammétrie, Pose Initiale, Orientation, Structure from Motion, Triplets.

1 Introduction

Il existe différentes stratégies pour traiter de grands jeux de données grâce à la Structure from Motion. La méthode globale permet de résoudre la scène en considérant toutes les vues simultanément en fonction de leurs orientations relatives [2, 6, 7].

Pour l'étape de Bundle Adjustment, nous devons choisir une première orientation à l'initialisation. Plusieurs travaux ont montré que le problème peut être résolu efficacement en commençant par une rotation, puis en effectuant une translation [1]. Ce problème est important pour permettre à l'étape de Bundle Adjustment de converger, et ainsi produire un résultat pour la scène à reconstruire.

Notre approche consiste à utiliser le graphe de vue et l'hypergraphe de triplets produit par MicMac [5] pour déterminer autant la rotation que translation des vues dans l'objectif d'obtenir une initialisation de la scène.

2 Méthodologie

Notre algorithme utilise en entrée un graphe de triplets d'images de la scène. Ce graphe est produit à partir du

graphe de paires d'images, qui est lui-même produit suite à l'appariement des images. Cette étape provient de MicMac [5] et le résultat produit est utilisé sans modification.

L'hypergraphe de triplets $\mathcal{H}_{\text{triplets}} = (\mathcal{V}, \mathcal{E})$ où tous les sommets $\mathcal{V}_i \in \mathcal{V}$ représentent une image, et chaque arête $\mathcal{E}_{i,j,k}$ connecte un triplet d'images $(\mathcal{V}_i, \mathcal{V}_j, \mathcal{V}_k)$. Les arêtes contiennent les informations de déplacement relatives $[\{R_{j,i}, tr_{j,i}\}, \{R_{j,k}, tr_{j,k}\}]$ dans le repère local du triplet avec origine \mathcal{V}_j telle qu'elle a été décrite dans [4]. On rajoute sur les arêtes $\mathcal{E}_{i,j,k}$ les points caractéristiques des images $(\mathcal{V}_i, \mathcal{V}_j, \mathcal{V}_k)$ calculés par MicMac.

L'hypergraphe est dense, ce qui signifie qu'il y a un grand nombre de triplets disponibles. L'objectif est donc d'estimer la qualité des arêtes de l'hypergraphe $\mathcal{H}_{\text{triplets}}$, ce qui permettra de construire un arbre couvrant minimum du graphe, et ainsi d'orienter toutes les vues.

Afin de parvenir à ce résultat, un prétraitement permettra d'obtenir des données locales sur les triplets qui seront ensuite pris en compte pour pondérer la recherche. Une série d'arbres couvrants de $\mathcal{H}_{\text{triplets}}$ vont être générés aléatoirement afin de proposer une orientation pour chaque vue. À l'aide de cette orientation, les triplets vont être évalués pour savoir s'ils confirment ou non la proposition de l'arbre aléatoire. Les itérations sont indépendantes entre chaque arbre couvrant aléatoire, il est donc possible de paralléliser cette étape.

Finalement, un arbre couvrant minimum de $\mathcal{H}_{\text{triplets}}$ permet de proposer une pose initiale pour toutes les images de la scène avec les meilleurs triplets.

2.1 Prétraitement

Si nous traitons tous les triplets uniformément pour créer les arbres couvrants aléatoires, certains chemins choisis par les arbres ne seront pas forcément intéressants. Pour limiter ce cas, sans pour autant l'empêcher, un prétraitement est fait pour, à priori, identifier la qualité des liaisons des triplets.

Ce prétraitement consiste, pour toutes les liaisons de triplets adjacents qui ont deux vues en commun, de calculer

la distance de reprojection des points caractéristiques entre les deux triplets en fonction de leurs images communes. Pour optimiser les calculs, les points caractéristiques virtuels utilisés sont créés suivant la méthode décrite dans [4]. Ce pré-score nous permet d’avoir une erreur locale entre les deux triplets, et donc l’affecte qu’il aura sur la précision de l’orientation si l’arbre couvrant le traverse.

2.2 Arbre couvrant aléatoire

Un arbre couvrant aléatoire est généré suivant l’algorithme de Prim [3] qui cherche à construire l’arbre couvrant minimum.

Le cout des arêtes est établi aléatoirement et pondéré à partir du score calculé pendant le prétraitement (Voir 2.1). Le score aléatoire permet d’explorer l’hypergraphe en entier, mais la pondération nous évite de prendre des chemins dont à priori les orientations sont mauvaises.

En suivant le chemin de l’arbre aléatoire, l’enchaînement des bascules entre les triplets permet à toutes les vues d’êtres orientées telles que visible dans la figure 2.

2.3 Score

En utilisant l’orientation des images, nous calculons le score pour chaque triplet. Le calcul du score consiste à faire la reprojection des points caractéristique pour obtenir l’erreur sur les trois images du triplet suivant l’orientation de l’arbre couvrant aléatoire. Là aussi, 5 points triples virtuels sont utilisés issue de [4]. L’erreur de reprojection moyenne sur tous les points est utilisée pour calculer le score du triplet.

Cette approche a pour conséquence de pénaliser les triplets qui se trouvent profonds dans l’arbre et qui vont ainsi cumuler les erreurs de tous les triplets précédents lors de leur orientation. Pour limiter l’impact de ce cas-là, une normalisation est effectuée suivant la formule : $Score = \frac{Residue}{Residue + R_0}$ avec R_0 une constante choisie par l’utilisateur commune à tous les calculs.

2.4 Arbre couvrant minimal

La génération des poses initiale est produite à l’aide de l’arbre couvrant minimal du graphe $\mathcal{H}_{triplets}$ pondéré suivant la moyenne des $Score$ calculé sur chaque arbre couvrant aléatoire (Voir 2.3).

L’arbre couvrant minimal est généré suivant l’algorithme de Prim [3]. Le triplet avec le $Score$ le plus faible est choisi en *racine*. Ensuite, l’algorithme original est exécuté pour trouver l’arbre composé des $Scores$ les plus faible couvrant toutes les vues de la scène.

3 Résultats

Les essais ont été faits sur un jeu de donnée issue d’une prise de vue en drone suivant une trajectoire classique avec recouvrement (Voir figure 1). Ce jeu de donnée contient 2317 triplets entre 73 images dont la position GPS précise est connue pour chaque vue permettant d’avoir une vérité terrain (Voir figure 1).

Distance	Moyenne	Écart-type
Position (en mètre)	0.4307	0.1410
Rotation (en radian)	0.0037	0.00041

TABLE 1 – Distance produite par notre méthode sur 50 exécutions avec chacune 500 arbres aléatoires et $R_0 = 10$.

Les triplets générés à partir de ce jeu de donnée sont de bonne qualité, ainsi, on arrive à distinguer la trajectoire visuellement sur les arbres aléatoires générés (Voir figure 2). Après le calcul de l’arbre couvrant minimale, la trajectoire de référence est visible (Voir figure 3).

Pour quantitativement estimer la précision de notre orientation, nous utilisons comme vérité terrain la reconstruction issue du processus de Structure from Motion de MicMac [5] dont les poses ont été initialisées à l’aide des données GPS. La sortie de notre méthode n’ayant pas le même système de coordonnées que celui de la référence, nous devons basculer le résultat dans le même système de coordonnée. Notre approche utilise des nombres aléatoire et produit un résultat différent à chaque exécution. Pour estimer sa robustesse, nous la lançons une multitude de fois et suivons son évolution par rapport à la référence (Voir Table 1).

La faible distance des positions et rotation indique que notre méthode arrive à correctement reproduire la scène, et le faible écart-type la robustesse de notre méthode.

4 Travaux futurs

Les résultats indiquent que notre méthode est efficace sur des graphes de petite taille. En testant sur des jeux de données plus volumineux, nous avons constaté que la largeur du graphe rend les arbres couvrants aléatoires moins efficaces pour estimer la qualité des triplets.

Plusieurs sous-branches de l’arbre sont correctement orientées, mais pas l’orientation globale. Cela montre les limites de notre approche qui consiste à utiliser une bascule pour orienter les triplets adjacents et la nécessité d’utiliser un ajustement de faisceaux.

De futurs travaux se fonderont sur l’utilisation de l’arbre couvrant minimum pour diviser l’hypergraphe dans une approche hiérarchique de Structure from Motion.

Références

- [1] Daniel Barath, Dmytro Mishkin, Iván Eichhardt, Ilia Shipachev, and Jiří Matas. Efficient Initial Pose-graph Generation for Global SfM. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14541–14550.
- [2] David Crandall, Andrew Owens, Noah Snavely, and Dan Huttenlocher. Discrete-continuous optimization for large-scale structure from motion. In *CVPR 2011*, pages 3001–3008.
- [3] R. C. Prim. Shortest Connection Networks And Some Generalizations. 36(6) :1389–1401.

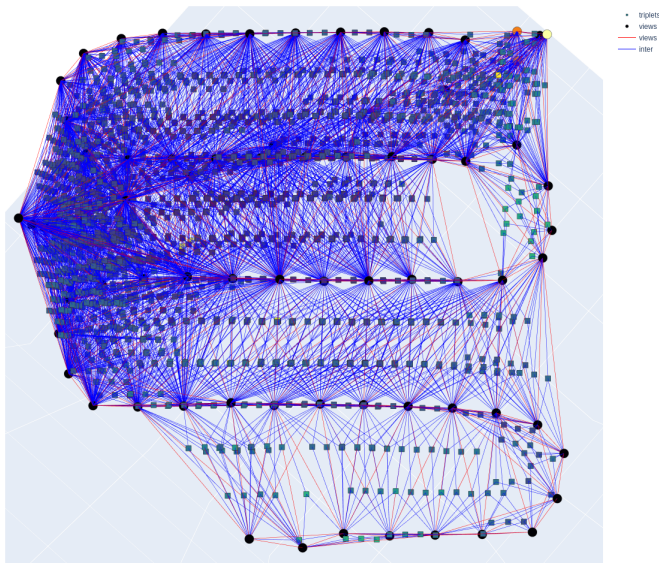


FIGURE 1 – Graphe de triplets $\mathcal{H}_{\text{triplets}}$.

- : image, ■ centre du triplet
- : liaisons du triplet avec ses images
- : liaisons entre les vues.

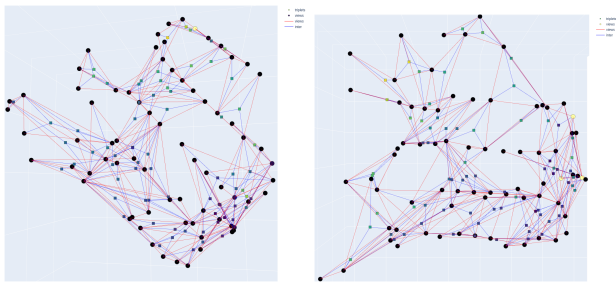


FIGURE 2 – Exemple d'arbres couvrants aléatoire pondéré

- [4] E. Rupnik and M. Pierrot Deseilligny. Towards Structureless Bundle Adjustment With Two- And Three-View Structure Approximation. V-2-2020 :71–78.
- [5] Ewelina Rupnik, Mehdi Daakir, and Marc Pierrot Deseilligny. MicMac – a free, open-source solution for photogrammetry. 2(1) :14.
- [6] Chris Sweeney, Torsten Sattler, Tobias Hollerer, Matthew Turk, and Marc Pollefeys. Optimizing the Viewing Graph for Structure-from-Motion. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 801–809. IEEE.
- [7] Kyle Wilson and Noah Snavely. Robust Global Translations with 1DSfM. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, volume 8691 of *Lecture Notes in Computer Science*, pages 61–75. Springer International Publishing.

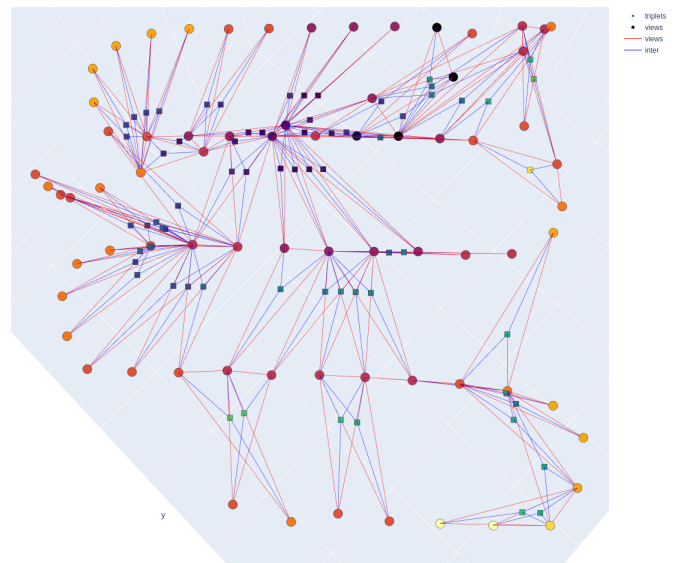


FIGURE 3 – Arbre couvrant minimal de $\mathcal{H}_{\text{triplets}}$.

- : image, ■ centre du triplet
- : liaisons du triplet avec ses images
- : liaisons entre les vues.

Plus l'image est claire, plus elle est profonde dans l'arbre.