



HAL
open science

Estimation à l'échelle du mouvement d'un réseau multi-caméras non synchronisées

Rawia Mhiri, Pascal Vasseur, Stéphane Mousset, Rémi Boutteau, Abdelaziz
Bensrhair

► **To cite this version:**

Rawia Mhiri, Pascal Vasseur, Stéphane Mousset, Rémi Boutteau, Abdelaziz Bensrhair. Estimation à l'échelle du mouvement d'un réseau multi-caméras non synchronisées. *Reconnaissance de Formes et Intelligence Artificielle (RFIA) 2014*, Jun 2014, Rouen, France. hal-00989061

HAL Id: hal-00989061

<https://hal.science/hal-00989061v1>

Submitted on 9 May 2014

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 à l'échelle du mouvement d'un réseau multi-caméras non synchronisées

R. MHIRI¹ P. VASSEUR¹ S. MOUSSET¹ R. BOUTTEAU² A. BENSRAHAIR¹

¹ LITIS - Université de Rouen - INSA de Rouen

² IRSEEM - ESIGELEC

Résumé

Nous proposons une nouvelle méthode d'estimation à l'échelle du mouvement d'un véhicule en utilisant un système multi-caméras non synchronisées que nous appelons la méthode des triangles. L'algorithme proposé utilise les informations issues du calibrage hors ligne des caméras. Nous supposons que la trajectoire entre deux poses de la même caméra en deux instants différents est approximée par un segment linéaire. Les poses relatives des caméras sont estimées par l'algorithme des 5-points [5]. Ensuite, les facteurs d'échelle métrique absolus sont calculés par une résolution linéaire d'un système d'équations qui tient compte de la connaissance du calibrage extrinsèque et de la linéarité du déplacement. Nous vérifions la validité de notre méthode sur des images simulées et des images réelles de la base de données KITTI.

Mots Clef

Estimation du mouvement, réseau multi-caméra non synchronisées, facteur d'échelle, méthode des triangles.

Abstract

We propose a new method to estimate the motion of a vehicle with the scale using non-synchronized multi-camera system which we call the method of triangles. The proposed algorithm uses the information from the offline calibration of the cameras. We assume that the trajectory between two poses of the same camera in different time steps is approximated by a linear segment. The relative cameras poses are estimated by the 5-points algorithm [5]. Then, the scale factors are calculated by a linear resolution of a system of equations that take into account the knowledge of the extrinsic calibration and the linearity of the displacement. We verify the validity of our method on simulated images and real images from KITTI dataset.

Keywords

motion estimation, asynchronous multi camera network, scale factor, triangle-based method.

1 Introduction

Les systèmes d'assistance à la conduite et les véhicules autonomes ont besoin d'avoir une parfaite connaissance de leur environnement proche afin d'être efficace et d'assurer la sécurité. Les capteurs basés vision permettent de concevoir des systèmes avec plusieurs configurations : mono-caméra [2][7], stéréovision [4][16] et réseaux de caméras multiples [1][3][13]. Ces systèmes sont utilisés pour accomplir plusieurs tâches, notamment la détection d'obstacles et de marquage routier, la localisation et la cartographie. Toutefois, une étape commune à ces applications porte sur l'estimation de mouvement, appelée aussi odométrie visuelle [5]. L'estimation des poses d'un système représente l'étape cruciale pour l'odométrie visuelle. Il s'agit de déterminer la rotation, la translation et le facteur d'échelle entre deux positions consécutives. L'estimation du facteur d'échelle est la partie la plus sensible qui doit être traitée soigneusement.

Lorsque l'échelle métrique absolue est désirée, l'odométrie visuelle fait recours à une connaissance sur la structure 3D de la scène. Dans le cas de la configuration mono caméra, une étape d'initialisation suivie du maintien du facteur d'échelle durant le mouvement est nécessaire, par exemple par une paramétrisation contrainte d'ajustement de faisceaux dans [8]. Pour une configuration stéréo ou multiple, l'estimation du mouvement et de la structure (SFM) se base sur la géométrie épipolaire pour déduire l'échelle métrique [17]. Cependant, l'estimation de mouvement à l'échelle n'est possible que si les caméras sont synchronisées et par conséquent la scène est prise sous différents angles de vue au même instant.

Pour les véhicules intelligents, les systèmes de vision embarqués sont généralement de faible coût pour être commercialisés à grande échelle, comme dans [4], la synchronisation devient alors un inconvénient majeur. En effet, elle impose l'ajout d'un circuit électronique et un câblage supplémentaire. De plus, un système synchrone utilise en général un matériel coûteux. En outre, l'acquisition d'images à partir d'un système asynchrone peut être facilement im-

plémentée sans se baser sur la vitesse d'acquisition de la caméra la plus lente.

Le SfM est désormais un problème bien maîtrisé dans le cas d'un réseau multi-caméras synchronisées. Jusqu'à présent, les caméras asynchrones ont été rarement étudiées sauf quelques exceptions dans le cadre de l'estimation de mouvement, [10][15].

Pour s'affranchir du caractère asynchrone, nous proposons une nouvelle méthode d'estimation de mouvement qui n'impose aucune contrainte sur la synchronisation des caméras ni sur leurs positions relatives. Les principales contributions de notre méthode, que nous appelons la méthode des triangles, se résument dans ces aspects de non-synchronisation afin d'estimer la trajectoire d'un véhicule à l'échelle sans aucune calibration temporelle. Cette méthode est basée sur l'hypothèse que le mouvement entre deux vues consécutives est rectiligne (linéaire ou lisse). La calibration géométrique des caméras est par ailleurs effectuée hors ligne.

Ce document présentera un aperçu de notre système après un rappel des travaux connexes à l'estimation de mouvement et du facteur d'échelle. Ensuite, notre méthode des triangles sera présentée avant que les résultats expérimentaux soient discutés pour en tirer des conclusions.

1.1 Travaux antérieurs

Estimer le mouvement relatif entre deux caméras calibrées est une étape classique. La solution de base pour cette étape est l'algorithme des 5-points proposé par D. Nister [5]. Cet algorithme cherche les solutions possibles pour les poses relatives entre deux positions de caméras calibrées par la correspondance de 5-points. Dans [6], Li et Hartley simplifient l'algorithme de 5-points en éliminant toutes les variables inconnues d'un seul coup au lieu de les éliminer une par une.

Pour un réseau multi-caméras, dans [1], Lee et al. utilisent un réseau de quatre caméras synchronisées dans l'estimation de mouvement d'un véhicule autonome. Le réseau a été modélisé par une caméra généralisée. Le mouvement relatif de la caméra généralisée est obtenu entre deux trames consécutives des caméras qui donnent un ensemble d'images synchronisées.

Dans [9], les auteurs s'affranchissent de la contrainte de chevauchement des champs de vision pour une odométrie visuelle en utilisant une configuration stéréo synchrone dans des environnements intérieurs. Chaque caméra a été traitée comme un système mono caméra pour estimer son mouvement à l'échelle près. Une solution linéaire a été utilisée par la suite afin de fusionner les facteurs d'échelle provenant des deux estimations en imposant la transformation statique connue entre les deux capteurs pour retrouver enfin l'échelle métrique absolue.

Tous les travaux précédents sont capables d'estimer le mouvement à l'échelle. Néanmoins, ils utilisent des caméras synchrones. Pour relâcher cette contrainte coûteuse, un système composé de dispositifs non synchronisés peut être

utilisé. Il existe très peu de configuration asynchrone possibles qui nécessite un complément d'information : en utilisant des approches 3D-3D mais les facteurs d'échelles relatifs sont différents ou en utilisant des approches 3D-2D qui nécessitent une structure 3D connue a priori. Ces approches ne sont pas très appropriées pour les applications temps réel.

Une configuration multi-caméras non synchronisées a été utilisée avec des techniques basées sur le SLAM dans [10]. Dans un environnement intérieur, un robot effectue la reconstruction 3D à l'aide de deux caméras non synchronisées. L'image synchronisée manquante est estimée par interpolation des points caractéristiques à partir des images acquises. Cette méthode retrouve des caractéristiques 3D en tenant compte des mesures odométriques pour interpoler l'image manquante. Cette approche traite un cas asynchrone mais pas des méthodes d'estimation de mouvement d'un système multi-caméras asynchrones.

Notre approche permet d'estimer le mouvement d'un système de caméras non synchronisées à l'échelle. L'approche est basée sur deux hypothèses : la première hypothèse est que deux poses consécutives d'une même caméra effectuent une trajectoire linéaire. La trajectoire d'un véhicule peut être approximée à une succession de segment de droite. La deuxième hypothèse est que les caméras sont calibrées hors ligne.

2 Description du système



FIGURE 1 – Une configuration multi-caméras modélisée pour un prototype de laboratoire (à droite) et embarquée sur un véhicule (à gauche)

La configuration du réseau multi-caméras est conçue de façon que la scène soit vue par deux ou plusieurs caméras. Les caméras voisines partagent un champ de vision commun afin d'avoir une vue complète tout autour d'un véhicule. Le modèle proposé est flexible, le nombre des caméras peut être modifié facilement. La configuration peut être montée sur une plate-forme mobile dépendant de l'application. La figure 1 illustre un exemple d'une configuration multi-caméras.

Le système est composé de N caméras non synchronisées, chaque caméra acquiert une image indépendamment des autres. La figure 2 montre une distribution simple d'un système synchrone (où toutes les caméras acquièrent des

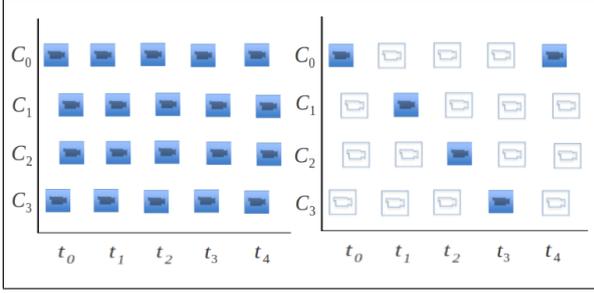


FIGURE 2 – (a) distribution d’images provenant d’un réseau de N caméras synchronisées (b) distribution d’images provenant d’un réseau de N caméras non synchronisées

images en même temps) et d’un système asynchrone (à chaque instant, uniquement une seule caméra peut acquérir une image). La modélisation est choisie en coordination avec le système de la figure 1. La représentation montre que les caméras ne sont pas alignées. Les caméras du système puissent être implémentées sur une voiture dans différents emplacements en tenant compte de l’hypothèse de chevauchement entre les caméras voisines.

2.1 Estimation du mouvement

Nous séparons la méthode en deux parties : l’estimation de poses relatives et l’estimation des facteurs d’échelle absolus. Les poses relatives des caméras sont estimées via SFM. Ensuite, les facteurs d’échelles absolus sont calculés en utilisant le calibrage extrinsèque et l’hypothèse de linéarité du déplacement.

Pour simplifier le problème, nous considérons un système avec seulement deux caméras qui acquièrent des images en trois instants. La première et la deuxième caméra, C_i et C_j , acquièrent respectivement des images I_i et I_j . La notation de temps est placée comme un indice fixé à la notation de la caméra. Par exemple, la position de la caméra C_i à l’instant t_0 est désignée par C_{i0} . Les transformations entre les poses des caméras sont désignées par T . La transformation euclidienne T représente la position d’une caméra dans le système de coordonnées de l’autre caméra. Par exemple, la transformation de la position de la caméra C_{i0} à la position de la caméra à C_{j1} est T_{i0}^{j1} . De la même façon, la matrice de rotation et le vecteur unitaire de la translation de la position de la caméra C_{i0} exprimés dans le système de coordonnées de la caméra C_{j1} sont désignés par R_{i0}^{j1} et t_{i0}^{j1} .

2.2 Estimation de Pose

Deux poses relatives d’une caméra peuvent être décrites par un vecteur unitaire de translation t , une matrice de rotation R et un facteur d’échelle λ . La transformation entre les deux positions peut être exprimée comme suit :

$$T = \begin{bmatrix} R & \lambda t \\ 0 & 1 \end{bmatrix} \quad (1)$$

Nous obtenons la matrice de rotation R et le vecteur de translation t entre deux images à partir de la matrice essentielle définissant la géométrie entre les deux poses.

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

En effet, la matrice essentielle peut être décomposée en valeurs singulières (SVD) pour déterminer la rotation et la translation, l’eq. 2.

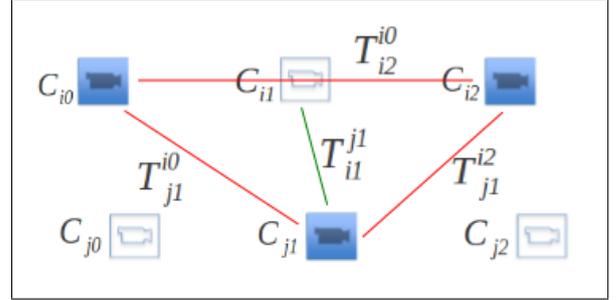


FIGURE 3 – méthode des triangles pour deux caméras non synchronisées : les lignes rouges désignent les transformations obtenues par SFM et les lignes vertes désignent la transformation statique issues de la calibration extrinsèque

Notre méthode des triangles repose sur trois matrices essentielles calculées entre trois images à trois instants : deux images provenant de la même caméra (C_i par exemple) et une image de l’autre caméra (C_j dans ce cas). Le triangle peut être modélisé comme dans la figure. 3 qui présente toutes les transformations possibles entre les trois images formant le “triangle”. Les transformations entre les trois images sont déterminées par l’algorithme des 5-points : T_{i2}^{i0} , T_{j1}^{i0} et T_{j1}^{i2} . La transformation statique T_{i1}^{j1} est obtenue à partir du processus de calibrage extrinsèque hors ligne.

Nous commençons notre algorithme par l’extraction et la mise en correspondance des points d’intérêt détectés dans les trois images du “triangle”. Nous effectuons des appariements entre les trois paires d’images. Dans notre implémentation, nous utilisons le détecteur SURF [11] et le descripteur FREAK [12]. Ces étapes sont appliquées entre les images I_{i0} , I_{i2} et I_{j1} .

La caméra C_i passe par une position intermédiaire où elle n’acquiert pas d’image en raison du fait de la non synchronisation. Il s’agit de la position C_{i1} qui peut être estimée en utilisant la transformation T_{i1}^{j1} . En résumé, nous utilisons quatre transformations : trois calculées via SFM (T_{i2}^{i0} , T_{j1}^{i0} et T_{j1}^{i2}) et une transformation statique (T_{i1}^{j1}) issue de la calibration extrinsèque.

2.3 Estimation des facteurs d’échelle : méthode des triangles

Jusque là, les facteurs d’échelle absolus sont inconnus entre les poses de la caméra. Les hypothèses initialement posées permettent de moduler le système par quatre principales transformations comme il a été montré dans la Fig.3.

Les trois images (I_{i0} , I_{i2} et I_{j1}) forment un "grand triangle" entre les positions de C_{i0} , C_{i2} et C_{j1} . La pose virtuelle de la caméra C_{i1} peut être considérée comme une position intermédiaire dans le grand triangle. Cette position donne deux sous-triangles : le premier sous-triangle est formé par C_{i0} , C_{i1} et C_{j1} et le second est formé par C_{i1} , C_{i2} et C_{j1} . Dans le premier sous-triangle, Fig.4 , la

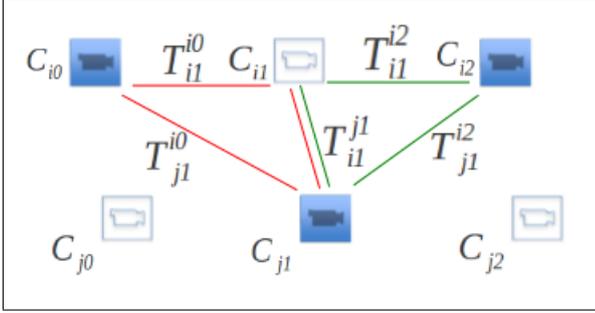


FIGURE 4 – le premier sous-triangle entre les caméras (C_{i0} , C_{i1} et C_{j1}) et le deuxième sous-triangle entre les caméras (C_{i1} , C_{i2} et C_{j1})

transformation de la position C_{i1} vers le repère de la position C_{i0} , T_{i1}^{i0} est égale à la transformation de la position de la caméra C_{j1} vers le repère de la position C_{i0} , T_{j1}^{i0} , multipliée par la transformation de la position de la caméra C_{i1} pour le repère de la position C_{j1} , T_{i1}^{j1} , l'eq. 3.

$$\mathbf{T}_{i1}^{i0} = \mathbf{T}_{j1}^{i0} \mathbf{T}_{i1}^{j1} \quad (3)$$

Les transformations euclidiennes sont exprimées dans le système de coordonnées homogènes, comme indiqué dans l'eq. 1. La transformation statique T_{i1}^{j1} obtenue par le calibrage extrinsèque est une transformation à l'échelle, le facteur d'échelle de cette transformation est connu. Développons l'eq. 3, nous introduisons les facteurs d'échelles inconnues (λ_1 , α) de chaque transformation dans l'eq. 4. λ_1 est le facteur d'échelle associé à la transformation T_{i1}^{i0} et α est le facteur d'échelle associé à la transformation T_{i1}^{j1} .

$$\begin{bmatrix} R_{i1}^{i0} & \lambda_1 t_{i1}^{i0} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} R_{j1}^{i0} & \alpha t_{j1}^{i0} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_{i1}^{j1} & t_{i1}^{j1} \\ 0 & 1 \end{bmatrix} \quad (4)$$

En développant l'eq. 4, nous obtenons :

$$R_{i1}^{i0} = R_{j1}^{i0} R_{i1}^{j1} \quad (5)$$

$$\lambda_1 t_{i1}^{i0} - \alpha t_{j1}^{i0} = R_{j1}^{i0} t_{i1}^{j1} \quad (6)$$

Par ailleurs, l'équation eq. 6 peut être écrite :

$$\begin{bmatrix} t_{i1}^{i0} & -t_{j1}^{i0} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \alpha \end{bmatrix} = R_{j1}^{i0} t_{i1}^{j1} \quad (7)$$

Dans le deuxième sous triangle, Fig.4, la transformation de la position C_{i1} vers le repère de la position C_{i2} , T_{i1}^{i2} est égale à la transformation de la position de la caméra C_{j1} vers le repère de la position C_{i2} (T_{j1}^{i2}) multipliée par

la transformation statique de la position de la caméra C_{i1} pour le repère de la position C_{j1} , T_{i1}^{j1} , l'eq. 8.

$$T_{i1}^{i2} = T_{j1}^{i2} T_{i1}^{j1} \quad (8)$$

De la même façon que le premier sous-triangle, l'eq. 8 est développée pour obtenir :

$$\begin{bmatrix} R_{i1}^{i2} & \lambda_2 t_{i1}^{i2} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} R_{j1}^{i2} & \beta t_{j1}^{i2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_{i1}^{j1} & t_{i1}^{j1} \\ 0 & 1 \end{bmatrix} \quad (9)$$

λ_2 représente le facteur d'échelle associé à la transformation T_{i1}^{i2} . β représente le facteur d'échelle associé à la transformation T_{j1}^{i2} . Séparons dans l'eq. 8 les termes en rotation et en translation, nous obtenons :

$$R_{i1}^{i2} = R_{j1}^{i2} R_{i1}^{j1} \quad (10)$$

$$\lambda_2 t_{i1}^{i2} - \beta t_{j1}^{i2} = R_{j1}^{i2} t_{i1}^{j1} \quad (11)$$

Dans le "grand" triangle entre les positions C_{i0} , C_{i2} et C_{j1} , Fig. 3, les transformations euclidiennes entre les poses peuvent être exprimées comme dans l'eq. 12. La transformation de la position C_{i2} exprimée dans le repère de la caméra C_{i0} (T_{i2}^{i0}) est égale à la transformation de C_{j1} exprimée dans le repère de la caméra C_{i0} (T_{j1}^{i0}) multipliée par la transformation de C_{i2} exprimée dans le repère de la caméra C_{j1} (T_{i2}^{j1}).

$$T_{i2}^{i0} = T_{j1}^{i0} T_{i2}^{j1} \quad (12)$$

Comme pour les sous-triangles, l'eq. 12 devient :

$$\begin{bmatrix} R_{i2}^{i0} & (\lambda_1 + \lambda_2) t_{i2}^{i0} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} R_{j1}^{i0} & \alpha t_{j1}^{i0} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_{i2}^{j1} & \beta t_{i2}^{j1} \\ 0 & 1 \end{bmatrix} \quad (13)$$

Après le développement de l'eq. 13, nous séparons les termes en rotation et en translation. La translation conduit à l'eq. 14.

$$\begin{bmatrix} t_{i2}^{i0} & t_{i2}^{i0} & -t_{j1}^{i0} & -R_{j1}^{i0} t_{i2}^{j1} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \alpha \\ \beta \end{bmatrix} = 0 \quad (14)$$

En résumé, les équations en terme de translation pour les trois triangles étudiés sont exprimées par le système suivant :

$$\begin{cases} \lambda_1 t_{i1}^{i0} - \alpha t_{j1}^{i0} = R_{j1}^{i0} t_{i1}^{j1} \\ \lambda_2 t_{i1}^{i2} - \beta t_{j1}^{i2} = R_{j1}^{i2} t_{i1}^{j1} \\ \lambda_1 t_{i2}^{i0} + \lambda_2 t_{i2}^{i0} - \beta R_{j1}^{i0} t_{i2}^{j1} - \alpha t_{j1}^{i0} = 0 \end{cases} \quad (15)$$

Pour résoudre ces équations et obtenir les facteurs d'échelle, nous pouvons écrire le système d'équations, eq. 15, comme suit :

$$\begin{bmatrix} t_{i1}^{i0} & 0 & -t_{j1}^{i0} & 0 \\ 0 & t_{i1}^{i2} & 0 & -t_{j1}^{i2} \\ t_{i2}^{i0} & t_{i2}^{i0} & -t_{j1}^{i0} & -R_{j1}^{i0} t_{i2}^{j1} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} R_{j1}^{i0} t_{i1}^{j1} \\ R_{j1}^{i2} t_{i1}^{j1} \\ 0 \end{bmatrix} \quad (16)$$

Les facteurs d'échelle absolus peuvent être retrouvés par la résolution linéaire de l'éq. 16 par la méthode des moindres carrés. L'éq. 16 peut être écrite comme un système linéaire de la forme :

$$A.X = B \quad (17)$$

avec X le vecteur des facteurs d'échelle λ_1 , λ_2 , α et β . La transformation de C_{i0} dans le repère de C_{i1} et la transformation de C_{i2} vers C_{i1} peuvent être obtenues grâce à l'hypothèse de linéarité entre les trois positions C_{i0} , C_{i1} et C_{i2} de la caméra C_i . Les translations entre les trois positions C_{i0} , C_{i1} et C_{i2} peuvent être exprimées en termes de vecteurs unitaires : t_{i1}^{i0} , t_{i2}^{i1} et t_{i2}^{i0} . L'équation suivante représente la relation entre ces vecteurs.

$$(\lambda_1 + \lambda_2)t_{i2}^{i0} = \lambda_1 t_{i1}^{i0} + \lambda_2 t_{i2}^{i1} \Leftrightarrow t_{i2}^{i0} = t_{i1}^{i0} = t_{i2}^{i1} \quad (18)$$

Toutes les transformations sont calculées soit par le SFM ou à partir de la calibration et l'hypothèse de linéarité. Les facteurs d'échelle peuvent être obtenus par une simple résolution de l'éq. 17.

Ainsi, nous pouvons obtenir les poses relatives et les facteurs d'échelle absolus pour tous les triangles d'une séquence.

3 Expérimentation

3.1 Simulation

Nous appliquons l'algorithme précédemment décrit sur des données simulées. Un nuage de 100 points 3D est aléatoirement généré constituant une scène observée par deux caméras. Les points 3D sont projetés dans les deux caméras en utilisant le modèle de caméras perspectives. Nous simulons trois poses pour chacune des caméras. Les caméras virtuelles sont liées par une transformation statique afin de satisfaire nos hypothèses. Ainsi, le système virtuel se déplace linéairement. Trois images (I_{i0} , I_{j1} et I_{i2}) sont obtenues sans bruit puis avec un bruit additif de $\sigma=0.2$. Les points 3D sont devant les caméras. La figure 5 illustre la scène simulée.

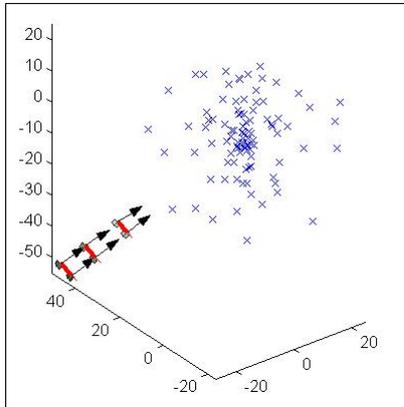


FIGURE 5 – Scène simulée : un nuage de point 3D et deux caméras suivant une trajectoire rectiligne

Nous considérons que tous les points 3D sont devant les caméras et donc vus par les deux caméras (100 points 2D). Les points d'intérêt sont extraits, décrits et appariés entre les trois images formant le triangle. Les matrices essentielles sont calculées puis décomposées en rotations et en translations. En utilisant la transformation T_{i1}^{j1} , les facteurs d'échelle sont estimés par la résolution du système d'équations obtenu par la méthode des moindres carrés.

TABLE 1 – Résultats de la simulation : les facteurs d'échelle obtenus par la méthode des triangles TM et à partir de la vérité terrain GT

	λ_1	λ_2	α	β
GT	0.7291	1.4810	0.4379	2.3887
TM sans bruit	0.7291	1.4810	0.4379	2.3887
TM avec bruit	0.8313	1.8344	0.3986	2.7781

Dans le cas non bruité, les facteurs d'échelle estimés dévoilent une très haute précision par rapport aux facteurs d'échelle métrique obtenus de la vérité terrain. Dans le cas bruité, les résultats présentent une erreur due au bruit additif, tableau 1. La simulation n'a pour objectif que de valider nos hypothèses pour un cas simple (valider l'aspect théorique de la méthode proposée) avec des données parfaites pour un seul triangle. Un déplacement d'un véhicule selon une trajectoire avec des courbures a été étudié dans le cas réel.

3.2 Données Réelles

Afin de valider la méthode des triangles, nous vérifions la validité de nos hypothèses pour une séquence du monde réel. Nous utilisons la banque d'images réelles de la base KITTI [13],[14]. Nous comparons les résultats de notre méthode pour une séquence issue de deux caméras parfaitement synchronisées par rapport aux positions de la vérité de terrain issus des mesures d'un GPS/INS. La banque d'image KITTI est une base de donnée publique qui a été acquise à partir d'un véhicule instrumenté. Nous utilisons une séquence stéréo en prenant une seule image à chaque instant afin de simuler la situation asynchrone. En effet, notre méthode utilise une image sur deux pour avoir l'aspect asynchrone. Les poses relatives et les facteurs d'échelle sont calculés pour chaque ensemble de trois images formant un grand triangle.

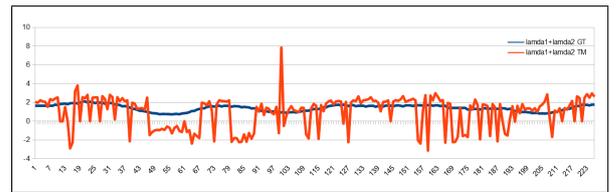


FIGURE 6 – La distribution brute de $\lambda_1 + \lambda_2$ (TM et GT) pour 223 triangles

TABLE 2 – Les valeurs moyennes de λ_1 et λ_2 (TM et GT) pour 223 poses

	TM	GT	TM/GT
λ_1	0.805	0.706	1.14022
λ_2	0.864	0.705	1.22553

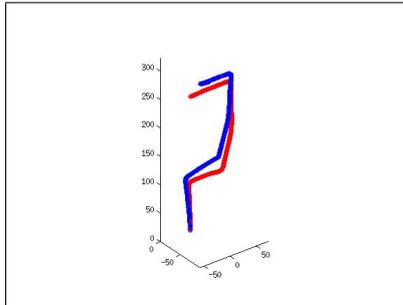


FIGURE 7 – La trajectoire d’un véhicule (549 positions) estimée par le méthode des triangles (la trajectoire en rouge) et la vérité terrain issu du GPS(la trajectoire en bleu)

Les résultats obtenus sont présentés dans le tableau 2 et dans la figure Fig. 6. La figure Fig. 6 montre la distribution des valeurs de $\lambda_1 + \lambda_2$ estimées par la méthode des triangle TM et obtenus de la vérité terrain GT pour une séquence de 223 triangles. Les pics proviennent des nombres insuffisants de points d’intérêt. Cette courbe présente les données brutes sans aucun filtrage d’où l’apparition d’erreurs inhérent aux techniques utilisées. Les erreurs apparaissent sur une image (un triangle) mais n’a pas une conséquence particulière sur la courbe.

La figure Fig. 7 illustre les trajectoires réelles et estimées. La trajectoire estimée et la trajectoire réelle sont très proches. La trajectoire bleue représente les positions du GPS (la vérité terrain) et la rouge représente la trajectoire estimée par la méthode des triangles. Il est clairement démontré que notre méthode est validée et que les hypothèses initialement posées le sont aussi.

La trajectoire est calculée pour une séquence de 549 images. La comparaison entre les trajectoires montre une dérive surtout dans les virages. Nous expliquons cette dérive par les erreurs d’estimation au niveau des rotations et des translations. En effet, les rotations et les translations sont estimées en se basant sur les appariements 2D entre les images. Ces erreurs peuvent être dues à la distribution des appariements dans les images qui n’est peut-être pas toujours homogène, d’où les erreurs d’estimations. En outre, nous avons trouvé deux erreurs majeures dans deux situations qui sont dues à un faible nombre de caractéristiques appariées, il s’agit des occultations dans les images. Comme nous estimons la pose d’une position courante dans le repère d’une pose précédemment estimée, les erreurs d’estimations s’accumulent.

Jusqu’à présent, nous n’avons comparé nos estimations

qu’avec les mesures du GPS (la vérité terrain). Des comparaisons de nos résultats avec d’autres approches synchrones seront effectuées dans nos futures travaux.

4 CONCLUSION

Nous avons présenté une nouvelle méthode d’estimation de mouvement, appelée méthode des triangles, pour un réseau multi-caméras non synchronisées. Nous supposons que la trajectoire entre deux images consécutives provenant d’une même caméra est approximée à un segment linéaire et que les caméras sont calibrées. Notre méthode peut être implémentée sur des plate-formes mobiles. La non synchronisation est utile pour de nombreuses applications telles les véhicules autonomes et les applications robotiques. L’approche présente une grande précision au niveau de l’estimation du mouvement pour un déplacement rectiligne et reste acceptable pour les virages. Bien que nous utilisions une méthode robuste, certaines erreurs peuvent se propager en raison des éventuelles erreurs dans l’estimation des poses. Ces erreurs peuvent être améliorées par des méthodes d’optimisation comme les méthodes d’ajustement de faisceaux.

Annexe

Ce travail est sponsorisé par le projet ANR Blanc DrAaCar N ANR-11-IS03-0003 et la Région de la Haute Normandie.

Références

- [1] G. Lee, F. Fraundorfer, and M. Pollefeys. Motion estimation for a self-driving car with a generalized camera. pages 2746 - 2753 In CVPR, 2013.
- [2] P., D. Scaramuzza, and R. Siegwart. Monocular slam based navigation for autonomous micro helicopters in gps denied environments. Journal of Field Robotics, Volume 28 Issue 6, pages 854-874, November 2011.
- [3] Meilland, M. Cartographie RGB-D dense pour la localisation visuelle temps-réel et la navigation autonome. These, Ecole Nationale Supérieure des Mines de Paris, March 2012.
- [4] P. T. Furgale, P.I Newman, R. Triebel et al. Toward Automated Driving in Cities using Close-to-Market Sensors, an overview of the V-Charge project. Intelligent Vehicles Symposium, page 809-816. 2013.
- [5] D. Nistér. An efficient solution to the five-point relative pose problem. In Pattern Analysis and Machine Intelligence, volume 26, pages 756-777, June 2004.
- [6] H. Li and R. Hartley. Five-point motion estimation made easy. ICPR '06 Proceedings of the 18th International Conference on Pattern Recognition - Volume 01, Pages 630-633, 2006.
- [7] D. Scaramuzza, F. Fraundorfer, and R. Siegwart. Real-time monocular visual odometry for on-road vehicles with 1-point RANSAC. Pages 488-494, Proceedings of the 2009 IEEE international conference on Robotics and Automation ICRA'09.

- [8] F. Fraundorfer, D. Scaramuzza, and M. Pollefeys. A constricted bundle adjustment parameterization for relative scale estimation in visual odometry. pages 1899-1904 In ICRA, 2010.
- [9] T. Kazik, L. Kneip, J. Nikolic, M. Pollefeys, and R. Siegwart. Real-time 6d stereo visual odometry with non-overlapping fields of view. In *Computer Vision and Pattern Recognition*, pages 1529-1536, June 2012.
- [10] M. Svedman, L. Goncalves, N. Karlsson, M. E. Munnich, and P. Pirjanian. Structure from stereo vision using unsynchronized cameras for simultaneous localization and mapping. In *IROS*, pages 3069-3074. IEEE, 2005.
- [11] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). In *Computer Vision and Image Understanding*, volume 110, pages 346-359, June 2008..
- [12] A. Alahi, R. Ortiz, and P. Vandergheynst. Freak : Fast retina keypoint. In *CVPR'12 Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 510-517. 2012.
- [13] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics : The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013.
- [14] 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)*, Pages 3354-3361, 2012.
- [15] C. Theobalt, H. Seidel, K. I. Kim, N. Hasler, C. Stoll, A. Elhayek. Spatio-temporal motion tracking with unsynchronized cameras. In *CVPR*, pages 1870-1877. IEEE, 2012.
- [16] M. Bertozzi, L. Bombini, A. Broggi, et al. The VisLab Intercontinental Autonomous Challenge : 13,000 km, 3 months, no driver. *Procs. 17th World Congress on ITS*, Busan, South Korea, October 2010.
- [17] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN :0521540518, second edition, 2004.