Reconstruction 3D générique et temps réel

Generic and Real-Time Structure from Motion

E. Mouragnon[1,2], M. Lhuillier[1], M. Dhome[1], F. Dekeyser[2], P. Sayd[2]

[1] LASMEA-UMR 6602, Université Blaise Pascal/CNRS, 63177 Aubière
[2] CEA, LIST, Laboratoire Systèmes de Vision Embarqués, Boite Courrier 94, Gif-sur-Yvette, F-91191 France ;

Résumé

Nous introduisons une méthode de reconstruction 3D générique et incrémentale. Par générique, on entend que la méthode proposée ne dépend pas d'un modèle de caméra spécifique. Pendant la reconstruction 3D incrémentale, les paramètres des points 3D et les poses des caméras sont raffinés simultanément par un ajustement de faisceaux local qui minimise une erreur angulaire entre des rayons. Cette méthode a trois avantages : elle est générique, rapide et précise. On l'expérimente sur des séquences réelles avec plusieurs types de caméras calibrées : une paire stéréo, une caméra perspective et une caméra catadioptrique.

Mots Clef

Méthode générique, ajustement de faisceaux, matrice essentielle généralisée, reconstruction 3D automatique, SLAM.

Abstract

We introduce a generic and incremental Structure from Motion method. By generic, we mean that the proposed method is independent of any specific camera model. During the incremental 3D reconstruction, parameters of 3D points and camera poses are refined simultaneously by a local bundle adjustment that minimizes an angular error between rays. This method has three main advantages : it is generic, fast and accurate. The proposed method is evaluated by experiments on real sequences with many calibrated cameras : stereo rig, perspective and catadioptric cameras.

Keywords

Generic Method, Bundle Adjustment, Generalized Essential Matrix, Automatic Structure from Motion, SLAM.

1 Introduction

L'estimation automatique du mouvement d'une caméra et de points 3D d'une scène à partir d'une séquence d'images ("structure from motion" ou SfM) a déjà été beaucoup étudiée. Plusieurs modèles de caméras ont été utilisés : perspectifs, fish-eyes, catadioptriques, systèmes de multi-caméras (par exemple : une paire stéréo rigide) etc. Beaucoup d'algorithmes **spécifiques** (c'est à dire spécifiques à un modèle de caméra donné) ont été développés avec succès et sont maintenant couramment utilisés pour les modèles perspectifs et stéréo [18, 16, 9]. Les cas des caméras omnidirectionnelles centrales (catadioptriques et fish-eye) ou non centrales (multi-caméras) qui offrent un grand champ de vue ont aussi été examinés [2, 12, 17]. C'est un défi intéressant que de développer des outils **génériques** pour le SfM qui sont exploitables pour tout type de caméras. D'autres auteurs ont commencé à examiner cela en introduisant des modèles de caméra génériques [22, 8]. Dans un modèle de caméra générique, chaque pixel définit un rayon dans le système de coordonnée des caméras qui peut intersecter ou non un point unique appelé centre de projection. Dans les travaux récents sur le SfM générique, le mouvement de caméra peut être estimé à l'aide de la généralisation de la matrice essentielle classique [17, 19] donnée par l'équation de Pless [17] ou avec des méthodes minimales d'estimation de pose relative [21].

Une méthode est nécessaire pour le raffinement des points 3D et des poses de la caméra. La meilleure solution pour la précision est l'ajustement de faisceaux (AF) [23] appliqué sur tout les paramètres (AF global) de la séquence d'image entière. Cependant, il est clair qu'une telle méthode ne peut être temps réel. En général, les méthodes rapides de SfM ou de SLAM basées vision [3, 15] ("simultaneous localization and mapping") sont moins précises que les méthodes hors lignes où la solution optimale est calculée à l'aide d'un AF global [23]. Dans cet article, nous présentons une méthode qui a une précision similaire à celle de l'AF global dans un contexte générique et temps réel. Ceci est possible car nous avons développé une méthode incrémentale qui ne raffine pas la structure 3D entière, mais seulement les paramètres estimés en fin de séquence courante. Dans le cas générique où plusieurs types de caméras sont possibles (perspective, stéréo, catadioptrique ...), l'AF est différent de l'AF classique utilisé pour les caméras perspectives. Notre méthode générique est basé sur les rayons s'échappant de la caméra et minimise une erreur d'angle entre des rayons. Naturel-

lement, le premier avantage est la possibilité d'échanger facilement un modèle de caméra par un autre. Le second avantage est que la méthode reste efficace même si la fonction de projection n'est pas explicite (comme dans le cas catadioptrique non-central).

Comparaisons avec les travaux précédents : Pour résumer, les plus proches travaux sont

- génériques mais pas temps réel [8, 19, 10] (et seulement central dans le cas de [8])
- temps réels mais pas génériques [3, 15] (et sans ajustement de faisceau, même local)
- génériques grâce à l'équation de Pless [17, 19] (généralisation de la contrainte épipolaire), mais aucun détail sur la résolution de cette équation n'est donné dans les situations courantes
- incrémentale avec ajustement de faisceaux local mais pas génériques [13, 4] (et ne sont pas validés dans un système sur des séquences d'images réelles dans le cas de [4]).

Contributions : La première contribution de notre travail est une méthode de SfM générique et temps réel basée sur une méthode d'ajustement de faisceaux local qui minimise une erreur angulaire. La seconde contribution est une méthode détaillée pour résoudre l'équation de Pless (contrairement à ce qui est suggéré dans [17, 19] cela n'est pas un "simple" problème linéaire). Nous comparons aussi nos résultats avec la vérité terrain (un GPS différentiel) et avec des résultats obtenus avec la méthode la plus précise connue (mais non générique et non temps réelle) : un ajustement de faisceaux spécifique et global.

La suite de l'article est organisée de la façon suivante : la partie 2 résume notre approche et donne le modèle de caméra générique. La méthode d'initialisation et l'ajustement de faisceau modifié sont présentés dans les parties 3 et 4. Finalement, les expériences sont données dans la partie 5. Cet article est la traduction en français de [14].

2 Résumé de l'approche

2.1 Modèle de caméra

Pour tout pixel $\mathbf{p}$ d'une image générique, la fonction de calibration f (connue) de la caméra définit un rayon optique $r = f(\mathbf{p})$. Ce rayon est une droite orientée $r = (\mathbf{s}, \mathbf{d})$ avec $\mathbf{s}$ le point de départ ou origine et $\mathbf{d}$ la direction du rayon dans le repère de la caméra (avec $||\mathbf{d}|| = 1$). Pour une caméra centrale, $\mathbf{s}$ est un point unique (le centre) pour tout pixel $\mathbf{p}$. Dans le cas général, $\mathbf{s}$ peut être n'importe quel point donné par f.

L'ensemble des points de départs $\mathbf{s}$ est appelé la surface des rayons [1]. On note que plusieurs surfaces sont possibles car l'origine d'un rayon peut être choisie n'importe où sur ce rayon. Un choix judicieux de la surface des rayons (par exemple pour une précision maximale de la géométrie estimée) dépasse le cadre de cet article.

2.2 Résumé

La méthode est basée sur la détection et la mise en correspondance de points d'intérêts (Figure 1). Dans chaque nouvelle image, des points de Harris [6] sont détectés et appariés avec des points détectés dans une image précédente en utilisant un score de corrélation ZNCC (Zero Mean Normalized Cross Correlation) dans une région d'intérêt, et cela quelque soit le type de caméra. Les paires de points avec les meilleurs scores sont alors retenues. Pour assurer une estimation stable de la 3D, un sous ensemble d'images appelées images clefs sont sélectionnées. Le critère de sélection est basé sur un nombre de points appariés entre deux images clefs consécutives, qui doit être supérieur à une constante.

L'initialisation de la géométrie est obtenue à l'aide d'une méthode basée sur la résolution de l'équation de Pless (partie 3). Ensuite, l'algorithme est incrémental. Pour chaque nouvelle image de la vidéo, (1) des points d'intérêts sont détectés et appariés avec ceux de la dernière image clef (2) la pose de la caméra de la nouvelle image est estimée de façon robuste (partie 3.4) (3) on regarde si la nouvelle image est sélectionnée comme image clef (4) si c'est le cas, de nouveaux points 3D sont estimés et un ajustement de faisceaux est appliqué localement en fin de sequence (partie 4).

3 Initialisation générique

3.1 Équation de Pless

Étant donnée une liste de points mis en correspondance entre deux images génériques, on souhaite estimer la pose relative $(\mathtt{R}, \mathbf{t})$ des deux caméras génériques correspondantes. Pour chaque couple de point appariés, on a une correspondance entre des rayons optiques $(\mathbf{s}_0, \mathbf{d}_0)$ et $(\mathbf{s}_1, \mathbf{d}_1)$. Un rayon $(\mathbf{s}, \mathbf{d})$ est défini par ses coordonnées de Plücker $(\mathbf{q}, \mathbf{q}')$ telles que $\mathbf{q} = \mathbf{d}$ et $\mathbf{q}' = \mathbf{d} \wedge \mathbf{s}$, qui sont utiles pour notre calcul. Choisissons pour système de coordonnées global le système de coordonnées liées à la caméra 0 et $(\mathtt{R}, \mathbf{t})$ la pose de la caméra 1 dans ce repère. Les deux rayons devant s'intersecter, ils vérifient la contrainte épipolaire généralisée (ou équation de Pless [17])

$$\mathbf{q}_0'^{\top} \mathtt{R} \mathbf{q}_1 - \mathbf{q}_0^{\top} [\mathbf{t}]_{\times} \mathtt{R} \mathbf{q}_1 + \mathbf{q}_0^{\top} \mathtt{R} \mathbf{q}_1' = 0 \qquad (1)$$

avec $[\mathbf{t}]_{\times}$ la matrice antisymétrique du produit vectoriel par le vecteur 3D $\mathbf{t}$ (autrement dit, $[\mathbf{t}]_{\times} \mathbf{x} = \mathbf{t} \wedge \mathbf{x}$).

Nous avons identifié deux cas où cette équation a une infinité de solutions. Le premier cas est évident : il y a une infinité de solution si la caméra est centrale (on a le choix du facteur d'échelle pour la 3D). On note que l'équation 1 est la contrainte épipolaire habituelle définie par la matrice essentielle $\mathtt{E} = [\mathbf{t}]_{\times} \mathtt{R}$ si le centre de la caméra est à l'origine du repère caméra.

Le second cas est moins évident mais il se produit fréquemment en pratique. Dans nos expériences, nous supposons que nous avons seulement des mises en correspondance "simples" : tous les rayons $(\mathbf{s}^i, \mathbf{d}^i)$ d'un point 3D donné passent par un même centre de caméra (dans le repère local

à la caméra générique). Autrement dit, on a $\mathbf{q}'_0 = \mathbf{q}_0 \wedge \mathbf{c}^0$ et $\mathbf{q}'_1 = \mathbf{q}_1 \wedge \mathbf{c}^1$ avec $\mathbf{c}^0 = \mathbf{c}^1$. Pour un système composé de plusieurs caméras centrales (par exemple la paire stéréo rigide), cela signifie que les points 2D sont seulement mis en correspondance avec d'autres points de la même sous image. Ceci est fréquemment le cas en pratique pour deux raisons : des régions d'intérêt peu étendues pour une mise en correspondance fiable, et des intersections de champs de vue vides pour deux caméras composantes du système multi-caméras. Si le mouvement de la caméra générique est une translation pure ($\mathtt{R} = \mathtt{I}_3$), l'équation 1 devient $\mathbf{q}_0^\top [\mathbf{t}]_\times \mathbf{q}_1 = \mathbf{q}_0'^\top \mathbf{q}_1 + \mathbf{q}_0^\top \mathbf{q}'_1 = 0$ avec pour inconnue $\mathbf{t}$. Dans ce contexte, l'échelle de $\mathbf{t}$ ne peut être estimée. On supposera dans cet article que le mouvement de caméra n'est pas une translation pure au début de la séquence d'images.

3.2 Résolution de l'équation de Pless

On re-écrit l'équation 1 sous la forme

$$\mathbf{q}_0'^\top \tilde{\mathtt{R}} \mathbf{q}_1 - \mathbf{q}_0^\top \tilde{\mathtt{E}} \mathbf{q}_1 + \mathbf{q}_0^\top \tilde{\mathtt{R}} \mathbf{q}'_1 = 0 \qquad (2)$$

avec les deux matrices $(\tilde{\mathtt{R}}, \tilde{\mathtt{E}})$ de dimensions 3×3 comme nouvelles inconnues. On peut ranger les coefficients de $(\tilde{\mathtt{R}}, \tilde{\mathtt{E}})$ dans un vecteur $\mathbf{x}$ à 18 dimensions et on voit que chaque valeur du 4-uplet $(\mathbf{q}_0, \mathbf{q}'_0, \mathbf{q}_1, \mathbf{q}'_1)$ donne une équation linéaire $\mathbf{a}^\top \mathbf{x} = 0$. Si on a 17 valeurs de ce 4-uplet pour chaque couple de points appariés, on a 17 équations $\mathbf{a}_k^\top \mathbf{x} = 0$. Cela suffit pour déterminer $\mathbf{x}$ à un facteur d'échelle près [19]. On construit la matrice $\mathtt{A}_{17}$ contenant les 17 mises en correspondance telles que $\|\mathtt{A}_{17}\mathbf{x}\| = 0$ avec $\mathtt{A}_{17}^\top = [\mathbf{a}_1^\top | \mathbf{a}_2^\top | \cdots \mathbf{a}_{17}^\top]$. La résolution dépends de la dimension du noyau de $\mathtt{A}_{17}$, qui dépends lui-même du type de caméra utilisée. Nous déterminons $Ker(\mathtt{A}_{17})$ et sa dimension par une décomposition en valeur singulière de $\mathtt{A}_{17}$. On distingue plusieurs cas dans cet article : (1) caméra centrale à un seul centre optique (2) caméra axiale où les centres sont alignés (ce qui inclus la paire stéréo) et (3) les caméras non axiales.

Il n'est pas surprenant que la dimension du noyau du système linéaire à résoudre soit plus grand que 1. En effet, l'équation linéaire 2 a plus d'inconnues (18 inconnues) que l'équation non-linéaire 1 (6 inconnues). Des dimensions possibles sont données dans la table 1 (sous notre hypothèse de mise en correspondance "simple") et sont justifiées dans les paragraphes qui suivent. Les travaux précédents [17, 19] ignorent ces dimensions, bien qu'une méthode (éventuellement linéaire) devrait en tenir compte.

Caméra	Centrale	Axiale	Non axiale
$dim(Ker(\mathtt{A}_{17}))$	10	4	2

TAB. 1 – $dim(Ker(\mathtt{A}_{17}))$ dépends du type de caméra

Caméra centrale. Pour les caméras centrales (par exemple perspectives), tout les rayons optiques convergent sur le centre optique $\mathbf{c}$. Puisque $\mathbf{q}'_i = \mathbf{q}_i \wedge \mathbf{c} = [-\mathbf{c}]_\times \mathbf{q}_i$, l'équation 2 devient $\mathbf{q}_0{}^\top([\mathbf{c}]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{E}} - \tilde{\mathtt{R}}[\mathbf{c}]_\times)\mathbf{q}_1 = 0$. On note que $(\tilde{\mathtt{R}}, \tilde{\mathtt{E}}) = (\tilde{\mathtt{R}}, [\mathbf{c}]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}]_\times)$ est une solution de l'équation 2 pour toute matrice $\tilde{\mathtt{R}}$. De telles solutions sont "exactes" : l'équation 2 est exactement égale à 0 (à la précision de la machine près) quelque soit $(\mathbf{q}_0, \mathbf{q}_1)$. Notre "vraie" solution est $(\tilde{\mathtt{R}}, \tilde{\mathtt{E}}) = (\mathtt{0}, [\mathbf{t}]_\times \mathtt{R})$ si $\mathbf{c} = 0$, et l'égalité n'est pas exacte à cause du bruit dans les images. Donc, la dimension de $Ker(\mathtt{A}_{17})$ est au moins $9 + 1$. Des expériences confirment que cette dimension est 10 (au bruit près). Dans ce cas, on résout simplement la contrainte épipolaire habituelle $\mathbf{q}_0{}^\top [\mathbf{t}]_\times \mathtt{R} \mathbf{q}_1 = 0$ comme cela est décrit dans [7].

Caméra axiale. Ce cas inclus la paire stéréo habituelle composée de deux caméras perspectives rigidement liées. Soient $\mathbf{c}_a$ et $\mathbf{c}_b$ deux points sur l'axe de la caméra (contenant tous les centres). L'annexe montre que des solutions exactes sont définies par

$$\tilde{\mathtt{E}} = [\mathbf{c}_a]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_a]_\times \text{ et}$$
$$\tilde{\mathtt{R}} \in Vect\{\mathtt{I}_{3\times 3}, [\mathbf{c}_a - \mathbf{c}_b]_\times, (\mathbf{c}_a - \mathbf{c}_b)(\mathbf{c}_a - \mathbf{c}_b)^\top\}$$

avec nos hypothèses d'appariements "simples" (partie 3.1). La "vraie" solution n'est pas exacte à cause du bruit dans les images, et donc la dimension de $Ker(A_{17})$ est au moins $3 + 1$. Des expériences confirment que cette dimension est 4 (au bruit près).

Nous construisons une base de 3 solutions exactes $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ et une solution non exacte $\mathbf{y}$ avec les vecteurs singuliers correspondants aux 4 plus petites valeurs singulières de $\mathtt{A}_{17}$. Les valeurs singulières associées à $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ sont 0 à la précision de la machine près et celle de $\mathbf{y}$ est 0 au bruit image près. Nous recherchons la solution réelle $(\tilde{\mathtt{R}}, \tilde{\mathtt{E}})$ par combinaison linéaire de $\mathbf{y}$, $\mathbf{x}_1$, $\mathbf{x}_2$ et $\mathbf{x}_3$ telle que la matrice résultante $\tilde{\mathtt{R}}$ vérifie $\tilde{\mathtt{R}}^\top \tilde{\mathtt{R}} = \lambda \mathtt{I}_{3\times 3}$ ou $\tilde{\mathtt{E}}$ est une matrice essentielle. Soit $\mathbf{l}$ un vecteur tel que $\mathbf{l}^\top = [\lambda_1\ \lambda_2\ \lambda_3]^\top$. On note alors $\tilde{\mathtt{R}}(\mathbf{l})$ et $\tilde{\mathtt{E}}(\mathbf{l})$ les matrices $\tilde{\mathtt{R}}$ et $\tilde{\mathtt{E}}$ extraites de la solution $\mathbf{y} - [\mathbf{x}_1|\mathbf{x}_2|\mathbf{x}_3]\mathbf{l}$. Avec ces notations, on a $\tilde{\mathtt{R}}(\mathbf{l}) = \mathtt{R}_0 - \sum_{i=1}^3 \lambda_i \mathtt{R}_i$ et $\tilde{\mathtt{E}}(\mathbf{l}) = \mathtt{E}_0 - \sum_{i=1}^3 \lambda_i \mathtt{E}_i$ avec $(\mathtt{R}_i, \mathtt{E}_i)$ extraites de $\mathbf{x}_i$.

Une fois la base $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ calculée, nous calculons les coordonnées de la solution par minimisation non linéaire de la fonction $(\lambda, \mathbf{l}) \rightarrow \|\lambda \mathtt{I}_{3\times 3} - \mathtt{R}(\mathbf{l})^\top . \mathtt{R}(\mathbf{l})\|^2$ pour obtenir $\mathbf{l}$ et donc $\tilde{\mathtt{E}}$. Une décomposition SVD est appliquée sur $\tilde{\mathtt{E}}$ et on obtient 4 solutions [7] pour $([\mathbf{t}]_\times, \mathtt{R})$. La solution avec une contrainte épipolaire minimale $\|\mathtt{A}_{17}\mathbf{x}\|$ est choisie. Enfin, on raffine le facteur d'échelle k en minimisant $k \rightarrow \sum_i (\mathbf{q}'_{0i}{}^\top \mathtt{R} \mathbf{q}_{1i} - \mathbf{q}_{0i}^\top k.[\mathbf{t}]_\times \mathtt{R} \mathbf{q}_{1i} + \mathbf{q}_{0i}^\top \mathtt{R} \mathbf{q}'_{1i})^2$ et on applique $\mathbf{t} \leftarrow k\mathbf{t}$.

Caméra non axiale. Pour les caméras non axiales (par exemple un système de trois caméras perspectives rigidement liées dont les centres ne sont pas alignés), le problème est encore différent. L'annexe montre que des solutions "exactes" sont $(\tilde{\mathtt{R}}, \tilde{\mathtt{E}}) \in Vect\{(\mathtt{I}_{3\times 3}, \mathtt{0}_{3\times 3})\}$ avec notre hypothèse d'appariements "simples" (partie 3.1). La "vraie" solution n'est pas exacte à cause du bruit dans les

images, et donc la dimension de $Ker(A_{17})$ est au moins $1+1$. Des expériences confirment que cette dimension est 2 (au bruit près). Nous n'avons pas encore expérimenté ce cas sur des séquences réelles.

3.3 Initialisation robuste pour trois vues

La première étape de la méthode est la reconstruction 3D du premier triplé d'images clefs $\{0,1,2\}$ avec une méthode de type RANSAC. On effectue plusieurs tirages aléatoires de 17-uplets de couples de points appariés. Pour chaque tirage, la pose relative entre les vues 0 et 2 est calculée avec la méthode qui précède et les points appariés sont reconstruits par triangulation. La pose de la caméra 1 est estimée par des mises en correspondances 3D/2D par raffinements itératifs en minimisant une erreur angulaire définie dans la partie 4.2 (plus de détails dans la partie 3.4). La même erreur est minimisée pour trianguler les points. Finalement, la solution (parmi celles obtenues avec tous les tirages) qui donne le maximum de points consistants avec la géométrie des points de vues 0, 1, et 2 est retenue. Le j-ème point 3D est consistant avec la i-ème vue si l'erreur angulaire $||\epsilon_j^i||$ est inférieure à un seuil ϵ ($\epsilon = 0.01$ radians dans nos expériences).

3.4 Estimation robuste de pose

Le calcul de pose générique est utile pour les deux étapes de notre méthode (initialisation et processus incrémental). On suppose que la i-ème pose $P^i = (\mathtt{R}^i, \mathbf{t}^i)$ de la caméra est proche de la i-1-ème pose $P^{i-1} = (\mathtt{R}^{i-1}, \mathbf{t}^{i-1})$. P^i est estimé par optimisation non linéaire initialisée à P^{i-1} avec 5 mises en correspondance 3D/2D, en conjonction avec la méthode robuste RANSAC (une mise en correspondence 3D/2D est un point suivi sur 3 images et reconstruit sur 2 d'entre elles). Pour chaque tirage aléatoire, la pose est estimée en minimisant une erreur angulaire (partie 4.2) et nous comptons ensuite le nombre de points 3D consistants avec cette pose. La pose avec le maximum de points consistants est retenue, et un raffinement de cette pose est effectué avec tous les points qui ont été déclarés consistants à l'étape précédente.

4 Ajustement de faisceaux local et générique

4.1 Définitions

L'ajustement de faisceaux global (AF global) est le raffinement simultanés de tous les points 3D et toutes les poses de la caméra par minimisation d'une fonction de coût. Le nombre de paramètres est 3 pour chaque points 3D et 6 pour chaque pose (3 pour la translation et 3 pour la rotation). Soit $\mathbf{P}_j = [x_j, y_j, z_j, t_j]^\top$ les coordonnées homogènes du j-ème point dans le repère de la scène. Soient $\mathtt{R}^i$ et $\mathbf{t}^i$ l'orientation (une matrice rotation) et l'origine du repère de la i-ème caméra dans le repère de la scène.
Si $(\mathbf{s}_j^i, \mathbf{d}_j^i)$ est le rayon optique correspondant à l'observation de $\mathbf{P}_j$ par la i-ème caméra, la direction de la droite définie par $\mathbf{s}_j^i$ et $\mathbf{P}_j$ est $\mathbf{D}_j^i = \mathtt{R}^{i\top}[\mathtt{I}_3 \mid -\mathbf{t}^i]\mathbf{P}_j - t_j\mathbf{s}_j^i$ dans le repère de la i-ème caméra. Dans le cas idéal, les directions $\mathbf{d}_j^i$ et $\mathbf{D}_j^i$ sont parallèles (ce qui est équivalent à une erreur de reprojection de zéro pixel).

4.2 Choix de l'erreur générique

L'approche classique [23, 13] consiste à minimiser une somme de carrés $||\epsilon_j^i||^2$ avec ϵ_j^i une erreur spécifique dépendante du modèle de la caméra : l'erreur de reprojection en pixels. Dans notre cas, il faut minimiser une erreur générique. Nous définissons ϵ_j^i comme l'angle entre les directions $\mathbf{d}_j^i$ et $\mathbf{D}_j^i$ définies précédemment.
Des expériences de synthèse ont montré que la convergence de l'AF est mauvaise avec $\epsilon_j^i = \arccos(\mathbf{d}_j^i . \frac{\mathbf{D}_j^i}{||\mathbf{D}_j^i||})$. Une explication theorique est donnee dans [11] : on peut montrer que cet ϵ_j^i n'est pas de classe C^1 lorsque $\mathbf{d}_j^i$ et $\mathbf{D}_j^i$ sont égaux une fois normalisés (i.e. au point de la solution exacte). Une fonction de classe C^2 au voisinage de ce point est souhaitable afin d'obtenir une convergence rapide (quadratique) de l'algorithme de Levenberg-Marquardt impliqué dans l'AF.
La convergence est satisfaisante avec ϵ_j^i défini comme on le décrit maintenant. Nous choisissons $\epsilon_j^i = \pi(\mathtt{R}_\mathtt{j}^\mathtt{i}\mathbf{D}_j^i)$ avec $\mathtt{R}_\mathtt{j}^\mathtt{i}$ une matrice rotation telle que $\mathtt{R}_\mathtt{j}^\mathtt{i}\mathbf{d}_j^i = [0\ 0\ 1]^\top$ et π la fonction $\mathcal{R}^3 \rightarrow \mathcal{R}^2$ telle que $\pi([x\ y\ z]^\top) = [\frac{x}{z}\ \frac{y}{z}]^\top$. On note que ϵ_j^i est un vecteur 2D dont la norme euclidienne $||\epsilon_j^i||$ est égale à la tangente de l'angle entre $\mathbf{d}_j^i$ et $\mathbf{D}_j^i$. La tangente est une bonne approximation de l'angle si celui ci est petit, et c'est le cas dans ce contexte. On note que cette fonction ϵ_j^i est de classe C^2.

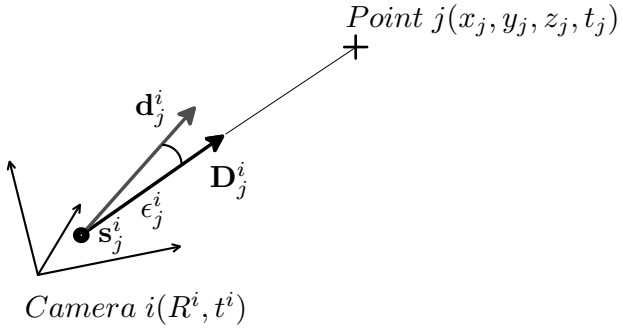

FIG. 2 – Angle entre le rayon correspondant à l'observation $(\mathbf{s}_j^i, \mathbf{d}_j^i)$ et le rayon passant par $\mathbf{s}_j^i$ et le points 3D $\mathbf{P}_j$. L'ajustement de faisceaux générique minimise une somme de carrés de tels angles.

4.3 Ajustement de faisceaux local

Lors de la reconstruction 3D incrémentale, lorsqu'une nouvelle image clef I^i est sélectionnée, de nouveaux points appariés sont triangulés. Ensuite, une étape d'optimisation est effectuée : un ajustement de faisceaux. Il s'agit d'une minimisation par la méthode de Levenberg-Marquardt d'une

fonction de coût $f^i(\mathcal{C}^i, \mathcal{P}^i)$, avec $\mathcal{C}^i$ et $\mathcal{P}^i$ des paramètres de la caméra générique (paramètres extrinsèques) et les paramètres de points 3D choisis pour cette i-ème étape. Comme il bien connu qu'un ajustement de faisceaux global (global à la séquence entière) est coûteux en temps, notre idée [13] est de réduire le nombre de paramètres calculés et d'éviter les redondances dans les calculs. Dans notre AF modifié, on n'optimise pas tout les paramètres des caméras, mais seulement les paramètres des n dernières poses (d'images clefs). Les coordonnées de tout les points 3D vus dans les n dernières images clefs sont raffinés (ce qui inclus aussi les nouveaux points 3D). Pour assurer la consistance du processus incrémental (c'est à dire assurer que les nouveaux paramètres sont compatibles avec ceux estimés avant), nous tenons compte des reprojections de points dans les N dernières images clefs avec $N > n$, typiquement $n = 3$ and $N = 10$. Ainsi, $\mathcal{C}^i$ est la liste des poses $\{C^{i-n+1} \ \dots \ C^i\}$ de la caméra et $\mathcal{P}^i$ contient tous les points 3D qui ont été détectés dans les images correspondantes à $\mathcal{C}^i$. La fonction de coût f^i est la somme des erreurs angulaires au carré pour toutes les observations obtenues dans les dernières images clefs $C^{i-N+1} \ \dots \ C^i$ de tout les points 3D dans $\mathcal{P}^i$:

$$f^i(\mathcal{C}^i, \mathcal{P}^i) = \sum_{C^k \in \{C^{i-N+1} \ \dots \ C^i\}} \ \sum_{p_j \in \mathcal{P}^i} ||\epsilon_j^k||^2.$$

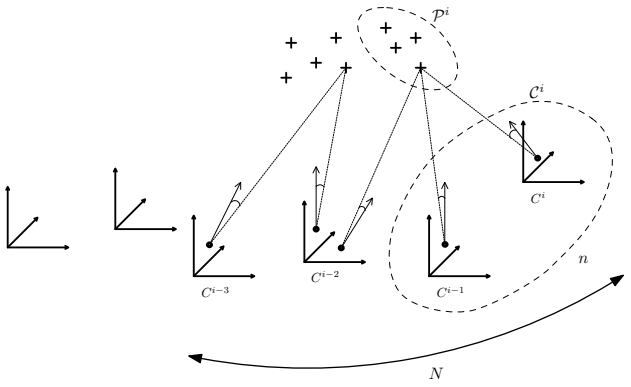

FIG. 3 – Ajustement de faisceaux local et angulaire appliqué lorsque la i-ème image clef (de pose C^i) est sélectionnée. Seuls les points $\mathcal{P}^i$ et poses $\mathcal{C}^i$ entourées sont optimisées. La fonction de coût minimisée tient compte des projections des points 3D dans les N dernières images clefs.

4.4 Remarques

Chaque optimisation par ajustement de faisceaux local comporte en fait deux étapes d'optimisation par la méthode de Levenberg-Marquardt, séparées par une mise à jour des points 3D et 2D satisfaisant la géométrie (le j-ème point 3D et la i-ème pose clef sont consistants si l'erreur angulaire $||\epsilon_j^i||$ est inférieure à un seuil $\epsilon = 0.01$ radians). Une étape est stoppée si l'erreur ne décroît pas assez vite ou bien si un nombre maximal d'itérations est atteint (5 dans notre cas). En pratique, le nombre d'itérations est vraiment faible ; cela est du au fait que, excepté pour la dernière pose, celles d'avant ont déjà été raffinées lors de l'ajustement de faisceaux précédant.

On remarque aussi que la condition $N > n$ énoncée dans le paragraphe précédant ne suffit pas dans le cas où la caméra est centrale. En effet, prendre $N = n + 1$ ne permet pas de fixer le facteur d'échelle relatif entre les n dernières poses clefs de la caméra (les poses raffinées) et le début de la séquence (les poses fixées). La condition $N \geq n + 2$ est donc nécessaire dans ce cas, et elle suffit en pratique dans nos expérimentations.

5 Expérimentations

La méthode de reconstruction 3D générique et incrémentale a été expérimentée sur des séquences réelles avec trois caméras différentes : une caméra perspective, une caméra catadioptrique, et une paire stéréo. Des exemples d'images sont montrées dans la figure 1 et les caractéristiques des séquences sont données dans la table 2. Les temps de calculs sont dans la table 3. Dans les expériences qui suivent, la trajectoire obtenue avec notre méthode générique est comparée à la vérité terrain (fournie par un GPS différentiel) ou bien par un AF global et spécifique. Une transformation rigide (rotation, translation, et facteur d'échelle) est appliquée sur la trajectoire pour recoler au mieux avec la trajectoire de référence [5]. Ensuite, l'erreur 3D moyenne (ou 2D moyenne dans le plan horizontal) est calculée entre la trajectoire de référence et la trajectoire estimée par notre méthode.

5.1 Comparaison avec la vérité terrain

Les résultats suivants sont obtenus avec une caméra perspective embarquée sur un véhicule expérimental équipé d'un GPS différentiel (de précision environ 2.5 cm). La trajectoire du véhicule est un "S" de 88 m de long (Séquence 1). Le mouvement obtenu avec notre algorithme est comparé à celui donné par le GPS, et la figure 4 montre les deux trajectoires recalées. Comme la trajectoire obtenue par GPS est donnée dans un repère métrique, on peut mesurer l'erreur de positionnement en mètres : l'erreur moyenne 3D est 1.57 m et l'erreur moyenne 2D dans le plan horizontal est 1.16 m. On a noté que la précision de la pose est difficile à obtenir pour cet exemple. Il y a deux raisons : la majorité des points 3D reconstruits sont distants, et il y a deux demi-tours serrés. Le temps de calcul est 2 min. 32 s pour la séquence entière, et 6.55 images sont traitées par seconde (en moyenne).

5.2 Comparaison avec un ajustement de faisceaux global et spécifique

Dans les exemples qui suivent, la vérité terrain est inconnue. On compare donc nos résultats avec ceux de la meilleure méthode connue : un AF global et spécifique (tous les paramètres de la séquence complète sont raffinés en minimisant la somme des carrés des erreurs de repro-

jection en pixels). Les caractéristiques et résultats sur les séquences sont donnés dans la table 2.

La séquence 2 est prise dans un environnement intérieur avec une caméra perspective tenue à la main. Un résultat très précis est obtenu : l'erreur 3D moyenne est inférieure à 6.5 cm pour une trajectoire d'environ 15 m. L'erreur relative est 0.45%.

La séquence 3 est prise dans un environnement extérieur avec une caméra catadioptrique centrale tenue à la main (le miroir "0-360" avec la caméra Sony HDR-HC1E en utilisant le format DV, visibles dans la figure 5). La partie utile de l'image rectifiée est contenue dans un disque dont le diamètre est de 458 pixels. Le résultat est également précis : l'erreur 3D moyenne est inférieure à 9.2 cm pour une trajectoire (d'environ) 40 m. L'erreur relative est 0.23%.

La séquence 4 est prise avec une paire stéréo (la baseline est de 40 cm) dans un couloir (Figure 5). L'image est composée de deux sous-images de dimensions 640×480. La trajectoire (longue de 20 m) est comparée avec les résultats obtenus avec un AF global appliqué sur la caméra de gauche/droite. L'erreur 3D moyenne est 2.7/8.4 cm comparée à la caméra de gauche/droite, et l'erreur relative est de $0.13\%/0.42\%$.

6 Conclusion

Nous avons proposé et expérimenté une méthode générique pour l'estimation robuste et temps réel de la géométrie d'une séquence d'images. Elle comporte deux étapes : une initialisation générique, et une reconstruction 3D incrémentale de la scène et du mouvement de la caméra. La précision est obtenue grâce à un ajustement de faisceaux local qui minimise une erreur angulaire. Les expériences ont montré qu'il est facile de changer une caméra avec une autre. De plus, des résultats prometteurs ont été obtenus avec trois différentes caméras (perspective, catadioptrique, stéréo) sur des séquences d'images réelles. On souhaite maintenant étendre notre méthode à des systèmes multi-caméras plus complexes, mais aussi (d'un point de vue purement technique), améliorer l'implémentation pour l'accélérer. Enfin, il faudrait étudier l'influence du choix des surfaces de rayon sur la précision de la reconstruction.

Annexe : solutions exactes à l'équation de Pless (version linéaire)

Dans cette partie, on exhibe des solutions à l'équation de Pless (version linéaire) qui ne sont pas la solution que l'on recherche.

La i-ème paire de rayons est définie par les coordonnées de Plücker $(\mathbf{q}_{0i}, \mathbf{q}'_{0i})$ et $(\mathbf{q}_{1i}, \mathbf{q}'_{1i})$ telles que $\mathbf{q}'_{0i} = \mathbf{q}_{0i} \wedge \mathbf{c}_{0i} = -[\mathbf{c}_{0i}]_\times \mathbf{q}_{0i}$ et $\mathbf{q}'_{1i} = \mathbf{q}_{1i} \wedge \mathbf{c}_{1i} = -[\mathbf{c}_{1i}]_\times \mathbf{q}_{1i}$. Si les rayons de la i-ème paire s'intersectent, alors ils satisfont l'équation de Pless et donc sa version linéaire :

$$\begin{aligned} 0 &= {\mathbf{q}'_{0i}}^\top \tilde{\mathtt{R}} \mathbf{q}_{1i} - \mathbf{q}_{0i}^\top \tilde{\mathtt{E}} \mathbf{q}_{1i} + \mathbf{q}_{0i}^\top \tilde{\mathtt{R}} \mathbf{q}'_{1i} \\ &= {\mathbf{q}_{0i}}^\top ([\mathbf{c}_{0i}]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{E}} - \tilde{\mathtt{R}}[\mathbf{c}_{1i}]_\times) \mathbf{q}_{1i} \end{aligned} \quad (3)$$

avec les deux matrices $(\tilde{\mathtt{R}}, \tilde{\mathtt{E}})$ de taille 3×3 comme inconnues. Dans cette partie, on souhaite trouver $(\tilde{\mathtt{R}}, \tilde{\mathtt{E}})$ tel que $\forall i, \tilde{\mathtt{E}} = [\mathbf{c}_{0i}]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_{1i}]_\times$. On recherche alors $\tilde{\mathtt{R}} \in \Re^{3\times3}$ tel que $\tilde{\mathtt{E}}$ soit indépendant de toutes les paires de centres possibles $(\mathbf{c}_{0i}, \mathbf{c}_{1i})$. La conséquence est que l'équation 3 est exactement égale à 0 (à la précision de la machine près) pour tout $(\mathbf{q}_{0i}, \mathbf{q}_{1i})$. On considère plusieurs cas.

Cas 1 : appariements simples $(\forall i, \mathbf{c}_{0i} = \mathbf{c}_{1i})$

Comme cela a été dit dans la partie 3.1, ce cas particulier est important en pratique.

Caméra Centrale Soit $\mathbf{c}$ le centre. Ce cas est simple : on a $\mathbf{c}_{0i} = \mathbf{c}_{1i} = \mathbf{c}$ et $\tilde{\mathtt{E}} = [\mathbf{c}]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}]_\times$. Tout $\tilde{\mathtt{R}} \in \Re^{3\times3}$ est possible.

Paire stéréo Soient $\mathbf{c}_a$ et $\mathbf{c}_b$ les deux centres. On a $(\mathbf{c}_{0i}, \mathbf{c}_{1i}) \in \{(\mathbf{c}_a, \mathbf{c}_a), (\mathbf{c}_b, \mathbf{c}_b)\}$ et $\tilde{\mathtt{E}} = [\mathbf{c}_a]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_a]_\times = [\mathbf{c}_b]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_b]_\times$. La contrainte sur $\tilde{\mathtt{R}}$ est donc

$$[\mathbf{c}_a - \mathbf{c}_b]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_a - \mathbf{c}_b]_\times = 0.$$

Tout $\tilde{\mathtt{R}}$ dans l'espace vectoriel des polynômes en $[\mathbf{c}_a - \mathbf{c}_b]_\times$ est possible. De plus, il est simple de prouver que cette contrainte ne permet pas d'autres $\tilde{\mathtt{R}}$ en changeant le système de coordonnées de sorte que $\mathbf{c}_a - \mathbf{c}_b = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}^\top$. On en déduit que

$$\begin{aligned} &\tilde{\mathtt{E}} = [\mathbf{c}_a]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_a]_\times \text{ et} \\ &\tilde{\mathtt{R}} \in Vect\{\mathtt{I}_{3\times3}, [\mathbf{c}_a - \mathbf{c}_b]_\times, (\mathbf{c}_a - \mathbf{c}_b)(\mathbf{c}_a - \mathbf{c}_b)^\top\} \end{aligned}$$

Caméra axiale Tous les centres de la caméra sont alignés : il existe $\mathbf{c}_a$ et $\mathbf{c}_b$ tels que $\mathbf{c}_{0i} = \mathbf{c}_{1i} = (1 - \lambda_i)\mathbf{c}_a + \lambda_i \mathbf{c}_b$ avec $\lambda_i \in \Re$. Donc,

$$\begin{aligned} &\tilde{\mathtt{E}}(\lambda) = [(1-\lambda)\mathbf{c}_a + \lambda \mathbf{c}_b]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[(1-\lambda)\mathbf{c}_a + \lambda \mathbf{c}_b]_\times \\ &= [\mathbf{c}_a]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_a]_\times + \lambda([\mathbf{c}_b - \mathbf{c}_a]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_b - \mathbf{c}_a]_\times) \end{aligned}$$

ne doit pas dépendre de λ. On voit que $[\mathbf{c}_b - \mathbf{c}_a]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_b - \mathbf{c}_a]_\times = 0$. Ce cas est le même que celui qui précède.

Caméra non axiale Il existe alors trois centres $\mathbf{c}_a, \mathbf{c}_b, \mathbf{c}_c$ non alignés. Maintenant, la contrainte sur $\tilde{\mathtt{R}}$ est

$$\begin{aligned} 0 &= [\mathbf{c}_a - \mathbf{c}_b]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_a - \mathbf{c}_b]_\times \\ &= [\mathbf{c}_b - \mathbf{c}_c]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_b - \mathbf{c}_c]_\times \\ &= [\mathbf{c}_c - \mathbf{c}_a]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_c - \mathbf{c}_a]_\times. \end{aligned}$$

Cette contrainte est trois fois celle de la paire stéréo. Changeons le système de coordonnées de sorte que $\{\mathbf{c}_a - \mathbf{c}_b, \mathbf{c}_b - \mathbf{c}_c, \mathbf{c}_c - \mathbf{c}_a\}$ soit la base canonique et écrivons les solutions des trois cas stéréo . On voit alors que seul $\tilde{\mathtt{R}} \in Vect\{\mathtt{I}_{3\times3}\}$ est possible.

Cas 2 : appariements quelconques

Maintenant, les appariements sont simples ou pas.

Caméra Centrale Ce cas ne peut se produire ici.

Paire stéréo Dans ce cas, on a $(\mathbf{c}_{0i}, \mathbf{c}_{1i}) \in \{(\mathbf{c}_a, \mathbf{c}_a), (\mathbf{c}_b, \mathbf{c}_b), (\mathbf{c}_a, \mathbf{c}_b), (\mathbf{c}_b, \mathbf{c}_a)\}$ et $\tilde{\mathtt{E}} = [\mathbf{c}_a]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_a]_\times = [\mathbf{c}_b]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_b]_\times = [\mathbf{c}_a]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_b]_\times = [\mathbf{c}_b]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_a]_\times$. La contrainte sur $\tilde{\mathtt{R}}$ est donc

$$0 = [\mathbf{c}_a - \mathbf{c}_b]_\times \tilde{\mathtt{R}} = \tilde{\mathtt{R}}[\mathbf{c}_a - \mathbf{c}_b]_\times.$$

Cette contrainte est plus forte que dans le cas de la caméra stéréo avec des appariements simples. On voit alors que seul $\tilde{\mathtt{R}} \in Vect\{(\mathbf{c}_a - \mathbf{c}_b)(\mathbf{c}_a - \mathbf{c}_b)^\top\}$ est possible.

Caméra axiale Il existe $\mathbf{c}_a$ et $\mathbf{c}_b$ tels que $\mathbf{c}_{0i} = (1 - \lambda_{0i})\mathbf{c}_a + \lambda_{0i}\mathbf{c}_b$ et $\mathbf{c}_{1i} = (1 - \lambda_{1i})\mathbf{c}_a + \lambda_{1i}\mathbf{c}_b$ avec $\lambda_{0i} \in \Re$ et $\lambda_{1i} \in \Re$. Donc,

$$\begin{aligned}\tilde{\mathtt{E}}(\lambda_0, \lambda_1) &= [(1 - \lambda_0)\mathbf{c}_a + \lambda_0\mathbf{c}_b]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[(1 - \lambda_1)\mathbf{c}_a + \lambda_1\mathbf{c}_b]_\times \\ &= [\mathbf{c}_a]_\times \tilde{\mathtt{R}} - \tilde{\mathtt{R}}[\mathbf{c}_a]_\times + \lambda_0[\mathbf{c}_b - \mathbf{c}_a]_\times \tilde{\mathtt{R}} - \lambda_1 \tilde{\mathtt{R}}[\mathbf{c}_b - \mathbf{c}_a]_\times\end{aligned}$$

ne doit pas dépendre de λ_0 et λ_1 : on a $0 = [\mathbf{c}_b - \mathbf{c}_a]_\times \tilde{\mathtt{R}} = \tilde{\mathtt{R}}[\mathbf{c}_b - \mathbf{c}_a]_\times$. Ce cas est le même que le précédent.

Caméra non axiale Il existe trois centres non alignés $\mathbf{c}_a, \mathbf{c}_b, \mathbf{c}_c$ et la contrainte sur $\tilde{\mathtt{R}}$ est trois fois celle du cas stéréo :

$$\begin{aligned}0 &= [\mathbf{c}_a - \mathbf{c}_b]_\times \tilde{\mathtt{R}} = \tilde{\mathtt{R}}[\mathbf{c}_a - \mathbf{c}_b]_\times = [\mathbf{c}_b - \mathbf{c}_c]_\times \tilde{\mathtt{R}} \\ &= \tilde{\mathtt{R}}[\mathbf{c}_b - \mathbf{c}_c]_\times = [\mathbf{c}_c - \mathbf{c}_a]_\times \tilde{\mathtt{R}} = \tilde{\mathtt{R}}[\mathbf{c}_c - \mathbf{c}_a]_\times.\end{aligned}$$

Changeons le système de coordonnées de sorte que $\{\mathbf{c}_a - \mathbf{c}_b, \mathbf{c}_b - \mathbf{c}_c, \mathbf{c}_c - \mathbf{c}_a\}$ soit la base canonique et écrivons les solutions des trois cas stéréo induits. On obtient $\tilde{\mathtt{R}} = 0$. Aucune solution exacte n'est obtenue avec cette méthode dans ce cas.

Références

[1] M. Grossberg and S.K. Nayar. A general imaging model and a method for finding its parameters. *ICCV'01*.

[2] P. Chang and M. Hebert. Omni-directional structure from motion. *IEEE Workshop on Omnidirectional Vision*, 2000.

[3] A. Davison. Real-Time Simultaneous Localization and Mapping with a Single Camera. *ICCV'03*.

[4] C. Engels, H. Stewenius, D. Nister. Bundle Adjustment Rules. *Photogrammetric Computer Vision*, 2006.

[5] O.D. Faugeras and M. Hebert. The representation, recognition, and locating of 3d objects. *IJRR*, 5 :27-52, 1986.

[6] C. Harris and M.Stephens. A combined corner and edge detector. *4th Alvey Vision Conference*, 1988.

[7] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, 2000.

[8] J. Kannala and S.S Brandt. A generic camera model and calibration method for conventional, wide angle, and fish-eye lens. *PAMI*, 28 :1335-1340, 2006.

[9] M. Lhuillier and L. Quan. A Quasi-Dense Approach to Surface Reconstruction from Uncalibrated Images. *PAMI*, 27(3) :418-433, 2005.

[10] M. Lhuillier. Effective and Generic Structure from Motion using Angular Error. *ICPR'06*.

[11] M. Lhuillier. Automatic Scene Structure and Camera Motion using a Catadioptric Camera System. *CVIU'08*, doi : 10.1016/j.cviu2007.05.004 (à paraitre).

[12] B. Micusik and T. Pajdla. Autocalibration and 3D reconstruction with non-central catadioptric camera. *CVPR'04*.

[13] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser and P. Sayd. Real Time localization and 3D reconstruction. *CVPR'06*.

[14] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser and P. Sayd. Generic and Real-Time Structure from Motion. *BMVC'07*.

[15] D. Nister. An efficient solution to the five-point relative pose problem. *PAMI*, 26(6) :756-777, 2004.

[16] D. Nister, O. Naroditsky and J. Bergen. Visual Odometry. *CVPR'04*.

[17] R. Pless. Using many camera as one. *CVPR'03*.

[18] M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool. Automated reconstruction of 3D scenes from sequences of images. *ISPRS Journal of Photogrammetry and Remose Sensing*, 55(4) :251-267, 2000.

[19] S. Ramalingam, S. Lodha and P. Sturm. A generic structure from motion framework. *CVIU*, 103(3) :218-228, 2006.

[20] E. Royer, M. Lhuillier, M. Dhome and T. Chateau. Localization in Urban Environment : monocular vision compared to a differential GPS sensor. *CVPR'05*.

[21] H. Stewenius, D. Nister, M. Oskarsson and K. Astrom. Solutions to minimal generalized relative pose problems. *OMNIVIS Workshop*, 2005.

[22] P. Sturm and S. Ramalingam. A generic concept for camera calibration. *ECCV'04*.

[23] B. Triggs, P. McLauchlan, R. Hartley and A. Zisserman. Bundle Adjustment–A modern synthesis. *Vision Algorithms : Theory and Practice*, 2000.

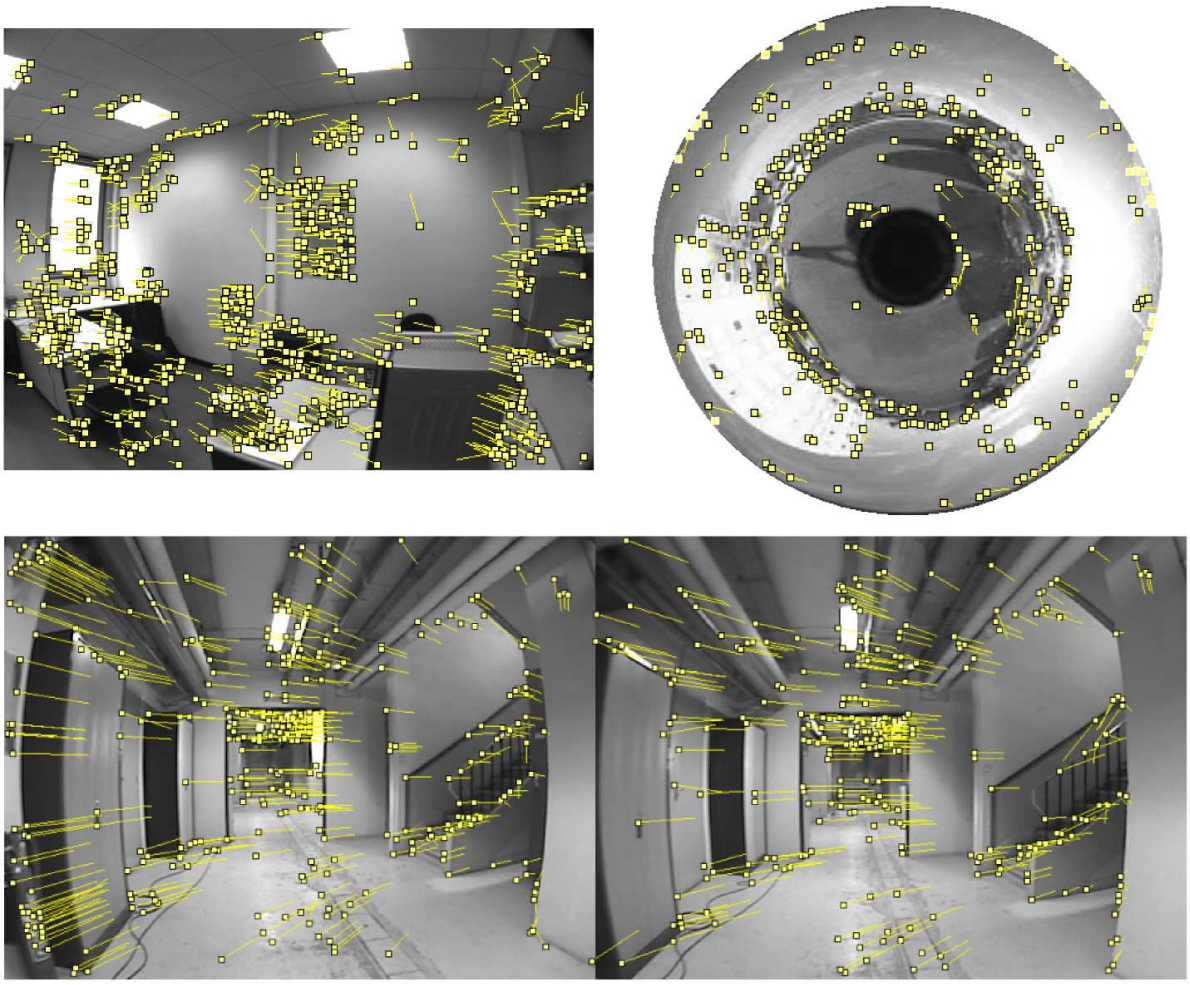

FIG. 1 – Suivi de points pour une image de caméra générique dans trois cas : caméra perspective (en haut a gauche), caméra catadioptrique (en haut a droite) et paire stéréo (en bas). La méthode de mise en correspondance est la même quelque soit le type de caméra utilisée.

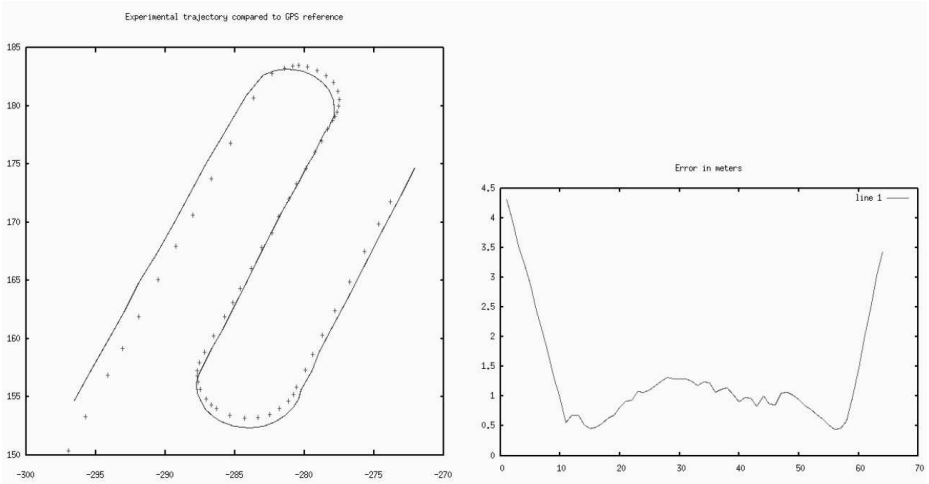

FIG. 4 – A gauche : décalage de la trajectoire estimée (méthode générique locale) avec la trajectoire vérité terrain (GPS différentiel). A droite : l'erreur 3D le long de la trajectoire. Les lignes continues sont pour le GPS et les points représentent les positions estimées des images clefs.

Séquence	Caméra	#images	#images clefs	#3D Pts	#2D Pts	Traj. long.
Séquence 1	perspective	996	66	4808	17038	88 m
Séquence 2	perspective	511	48	3162	11966	15 m
Séquence 3	catadioptrique	1493	132	4752	18198	40 m
Séquence 4	paire stéréo	303	28	3642	14189	20 m

TAB. 2 – Caractéristiques des séquences vidéo.

Caméra	Image	Détection+Matching	images	images clefs	fréquence
Perspective	512×384	0.10	0.14	0.37	**6.3 fps**
Catadioptrique	464×464	0.12	0.15	0.37	**5.9 fps**
Paire stéreo	1280×480	0.18	0.25	0.91	**3.3 fps**

TAB. 3 – Temps de calculs en secondes pour les trois caméras.

FIG. 5 – Trajectoires et vues de dessus des reconstructions 3D (trajectoires en bleu et points 3D en noir). En haut : caméra catadioptrique (séquence 2). En bas : paire stéréo (séquence 3).