

Estimation automatique de l'orientation relative en imagerie terrestre

■ Mahzad KALANTARI - Franck JUNG

Le calcul des paramètres extrinsèques d'une caméra à partir de plusieurs vues avec l'aide des points d'intérêt est la base de toutes les applications photogrammétriques ou de vision par ordinateur. Le défaut majeur de l'approche photogrammétrique, basée sur la condition de colinéarité, est dû à la non-linéarité du problème, et donc des valeurs approchées sont nécessaires. Dans les applications terrestres, ces valeurs initiales sont inconnues, et pour cette raison la communauté scientifique de vision par ordinateur est intéressée depuis longtemps au calcul de ces paramètres extrinsèques en utilisant des équations linéaires, basées sur la condition classique de coplanarité, enlevant le besoin de valeurs approchées. Dans un cas où les paramètres internes de la caméra sont connus, le calcul des paramètres extrinsèques utilise généralement la matrice dite essentielle. Beaucoup d'approches ont été employées pour résoudre cette matrice. Les plus connues sont les algorithmes dits des 8, 7 ou 6 points. Mais la condition de coplanarité a seulement 5 degrés de liberté. Nister a proposé une résolution polynomiale de la matrice essentielle demandant seulement 5 points, mais cette approche semble ne pas bien fonctionner avec des points coplanaires. Dans cet article nous présentons une nouvelle méthode de résolution directe de la rotation et la translation à partir de 5 points. Notre méthode n'utilise pas la matrice essentielle, elle exige une résolution polynomiale, en utilisant les bases de Gröbner, et se comporte correctement avec les points coplanaires. Cette méthode est également satisfaisante pour des images panoramiques, ce qui est un avantage comparatif important.

■ mots-clés

Photogrammétrie,
vision par ordinateur,
bases de Gröbner,
orientation relative

Le calcul des paramètres externes (décrivant l'orientation relative des prises de vues successives) de la caméra à partir de plusieurs vues à l'aide des points d'intérêts est la base de pratiquement toutes les applications aussi bien du domaine de la photogrammétrie, que de celui de la vision par ordinateur. Bien que le problème soit le même pour ces deux communautés, chacune a proposé une approche différente afin de calculer la position et l'attitude des caméras au moment des acquisitions de prises de vues.

L'inconvénient majeur de l'approche photogrammétrique, basée sur la condition de colinéarité, est dû à la non-linéarité du problème, et qui dit non-linéaire dit nécessité de disposer de valeurs approchées autour desquelles

on opère une linéarisation. Dans le cas de la photogrammétrie aérienne, les valeurs initiales sont très simples à estimer. En plus, l'avion ne subit que de faibles variations de rotation, et un ensemble de points d'appui peut toujours être obtenu.

En applications terrestres le calcul des valeurs initiales devient plus compliqué, c'est pour cela que la communauté de vision par ordinateur depuis plusieurs années s'est intéressée au calcul des paramètres extrinsèques de la caméra à l'aide d'équations linéaires. Le modèle employé est alors basé sur la condition de coplanarité, ce qui enlève le besoin d'avoir des valeurs approchées. Au cas où les paramètres internes de la caméra sont connus, le calcul des paramètres extrinsèques se

fait avec l'aide de la matrice dite "essentielle", le but devient alors de définir les composantes de cette matrice. C'est H. Longuet-Higgins [10] qui publia la première fois en 1981 cette notion de matrice essentielle. La contrainte de coplanarité a été longuement exploitée depuis lors par la communauté de vision par ordinateur, car en utilisant la matrice essentielle (E) la condition de coplanarité devient linéaire. La matrice essentielle suppose que la calibration de la caméra soit disponible.

Résoudre l'orientation relative revient donc à estimer les 9 éléments de la matrice essentielle. 8 points donc suffisent pour résoudre le système, car la dernière composante de cette matrice peut être prise égale à 1 vu que l'estimation se fait à une échelle près. Cette

■■■ résolution, bien qu'efficace en termes de calcul, ajoute 3 degrés de liberté à l'orientation relative qui à la base ne nécessite que 5 paramètres, comme on le verra ci-après. Un autre inconvénient majeur est qu'elle ne peut fonctionner si tous les points homologues sont coplanaires dans l'espace objet, ce qui peut très bien arriver par exemple sur des façades. Il y a eu donc beaucoup d'efforts au sein de la communauté de vision par ordinateur [7] pour réduire le degré de liberté de la matrice essentielle et réduire au minimum le nombre de points homologues requis. D'ailleurs, les différentes méthodes de résolution de la matrice essentielle ont pour nom ce nombre de points homologues minimum. Il y a donc eu les méthodes des 8, 7, 6 et plus récemment 5 points. Nous ne nous intéressons par la suite qu'aux méthodes de 5 points, qui représentent un minimum absolu.

J. Philip, en 1996 [13] a donné une première résolution avec 5 points. Il s'était basé sur les équations de Kuppra datant de 1913 [9]. en 2004, D. Nister [12] a poursuivi les travaux de Philip et a donné une solution numérique stable. D. Stewenius [17] a repris les mêmes équations que Nister, et donné une résolution algébrique avec l'aide des bases de Gröbner. Toutes les résolutions citées précédemment cherchent à estimer d'une façon ou d'une autre la matrice essentielle. La méthode proposée dans cet article cherche à estimer directement la rotation et la translation sans passer par la résolution de la matrice essentielle, ce qui fait sa grande différence avec les autres méthodes de résolution à 5 points. En effet aujourd'hui, et grâce aux moyens de calculs qui existent, cet article montre qu'avec une résolution polynomiale directe, il est tout à fait possible de donner une solution aux équations de coplanarité. Cette méthode peut aussi traiter des cas avec de faibles translations entre les images successives, elle peut donc être ainsi utilisée pour des images panoramiques. C'est son avantage principal par rapport aux autres méthodes. Par ailleurs, elle parvient aussi à fournir un résultat satisfaisant même si tous les points entrant dans le calcul sont coplanaires.

Présentation de la nouvelle méthode d'estimation de l'orientation relative

■ Géométrie de l'orientation relative

Avant de poursuivre il est important de préciser que la calibration de la caméra, c'est-à-dire l'ensemble des paramètres dits intrinsèques (focale, et point principal d'autocollimation) ainsi que la distorsion, est supposée définie préalablement. L'étape de l'orientation relative consiste à retrouver, comme son nom l'indique, la position entre deux caméras (avec un recouvrement stéréoscopique) au moment de la prise de vues. En général la position de la première caméra est prise comme origine du système terrain S_1 (Figure 1) et donc la position de la deuxième caméra est calculée par rapport à la première. Afin de formaliser cette géométrie, plusieurs méthodes existent.

Une façon de modéliser l'orientation relative est connue sous le nom de condition de coplanarité. Cette contrainte a été beaucoup utilisée par la communauté de vision par ordinateur depuis trois décennies. Comme on le voit mieux sur la figure 1, la condition de coplanarité entre deux images exprime le fait que le vecteur de visée depuis le premier sommet de prise de vue \vec{V}_1 , le vecteur de visée depuis le deuxième sommet de prise de vue (et exprimé dans le référentiel du premier) \vec{V}_2 , ainsi que le vecteur de la translation (entre les deux sommets de prises de vues) \vec{T} se trouvent dans le même plan, qui est appelé le plan épipolaire.

On peut traduire cette condition par un produit mixte nul entre ces 3 vecteurs. En d'autres termes : l'équation (1)

$$\vec{V}_1^t \cdot (\vec{R} \vec{V}_2 \wedge \vec{T}) = 0$$

En exprimant le produit vectoriel de manière algébrique (la translation étant exprimée sous forme d'axeateur), l'équation (1) peut être simplifiée dans sa forme matricielle : l'équation (2)

$$\begin{bmatrix} x_{a1} & y_{a1} & f \end{bmatrix} \begin{bmatrix} 0 & Tz & -Ty \\ -Tz & 0 & Tx \\ Ty & -Tx & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_{a2} \\ y_{a2} \\ f \end{bmatrix} = 0$$

Il existe plusieurs manières de représenter la rotation, et pendant long-

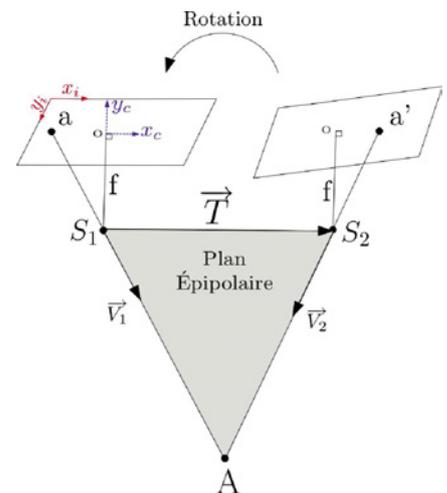


Figure 1 : La matrice rotation de la seconde caméra par rapport à la première est appelée R , et le vecteur de translation T est la base qui relie les centres optiques des caméras (S_1 et S_2). O est le point principal d'autocollimation (ppa). Les images du point terrain A sur les 2 images sont a et a' .

temps en photogrammétrie analogique et analytique, les angles d'Euler (ω, φ, κ) ont été employés. Avec la photogrammétrie numérique il est plus judicieux d'employer une représentation vecteur et angle [8], ou encore les quaternions.

La manipulation numérique de la matrice rotation est beaucoup plus simple avec les quaternions. La représentation d'une matrice de rotation à l'aide du quaternion $Q = a + b i + c j + d k$ (où a, b, c, d sont des nombres réels) se fait de la manière suivante : l'équation (3)

$$\begin{bmatrix} 1 - 2c^2 - 2d^2 & 2bc - 2da & 2bd + 2ca \\ 2bc + 2da & 1 - 2b^2 - 2d^2 & 2cd - 2ba \\ 2bd - 2ca & 2cd + 2ba & 1 - 2b^2 - 2c^2 \end{bmatrix}$$

avec la condition que la norme de Q , c'est-à-dire $(a^2 + b^2 + c^2 + d^2)^{1/2}$, doit être égale à 1.

En remplaçant la matrice de rotation dans l'équation (2), on aboutit à un polynôme de degré 3. Il est clair que dans les inconnues on trouve les éléments du quaternion (a, b, c, d) et la translation (T_x, T_y, T_z) ce qui en tout fait 7 inconnues. Chaque couple de points donne un polynôme. Comme cela a été vu, une contrainte de norme unitaire sur le quaternion doit impérativement être rajoutée. En orientation relative, l'échelle ne peut être estimée ce qui

implique que le vecteur de translation n'a que 2 degrés de liberté. Il est tout à fait possible de contraindre un des éléments du vecteur de la translation, par exemple rajouter comme contrainte que T_x soit égal à 1. Mais ce choix ne s'avère pas judicieux surtout en photogrammétrie terrestre où il peut y avoir des angles forts de rotation sur les axes Y et Z, ce qui par ailleurs n'était pas le cas en photogrammétrie aérienne. Dans cette optique, la seule contrainte qui semble adaptée à rajouter aux systèmes d'équations est la contrainte de vecteur unitaire pour la translation. Avec cette contrainte il est démontré qu'il est tout à fait possible de gérer la stéréo avant-arrière (translation en Z) ainsi que les faibles translations, cas des images panoramiques par exemple.

En résumé, nous avons fabriqué cinq équations d'observations, et rajouté deux contraintes. Les racines de ce système sont les paramètres d'orientation relative, à savoir les quatre paramètres du quaternion (a, b, c, d) et les trois paramètres de la translation (T_x, T_y, T_z).

■ Résolution des systèmes polynomiaux à l'aide des bases de Gröbner

La résolution d'un système polynomial consiste à trouver les zéros d'un système non linéaire polynomial du type $f(x) = 0$ avec $f = (f_1, f_2, \dots, f_n)$, où f_i est un polynôme à l variables $x = (x_1, x_2, \dots, x_l)$ [1][2]

En photogrammétrie, de nombreux systèmes d'équation sont polynomiaux, dont les équations exprimant la condition de coplanarité, ou celle de colinéarité. Une des méthodes les plus connues et employées pour résoudre les systèmes polynomiaux est la résolution analytique qui exploite f ainsi que ses dérivées, p. ex. la méthode de Newton qui est itérative, ou encore la minimisation locale de f . L'inconvénient majeur de ces méthodes est du au fait qu'elles nécessitent des solutions approchées servant de valeurs initiales, parfois difficiles à trouver. Une autre façon de trouver les solutions d'un système polynomial est d'avoir recours à une résolution algébrique.

Les techniques algébriques subdivisent

le problème de la résolution en deux étapes : la première consiste à transformer le système en un ou plusieurs systèmes équivalents mais mieux adaptés et qui constituent ce que l'on peut appeler une solution algébrique. La deuxième étape consiste, dans le cas où K est un sous-corps des complexes, à calculer les valeurs numériques des solutions à partir de la solution algébrique. Le calcul de la solution algébrique se décompose elle-même généralement en plusieurs étapes. La première est le plus souvent le calcul d'une base de Gröbner ; aussi de nombreux chercheurs ont assimilé pendant plusieurs années le calcul des bases de Gröbner à la résolution des systèmes polynomiaux [4].

Généralisant la division euclidienne, le calcul de PGCD et l'élimination de Gauss, les bases de Gröbner furent développées initialement par B. Buchberger dans les années 1960 [1] [15].

La solution algébrique [16] peut prendre diverses formes, chacune ayant ses avantages et ses inconvénients. La forme la mieux adaptée au calcul des solutions numériques est la RUR (Représentation Univariée Rationnelle) qui consiste en une équation en une variable $f(t) = 0$ (ou t est souvent une variable auxiliaire) et en l'expression des autres variables comme fractions rationnelles en t (quotients de deux polynômes). À partir de la RUR, le calcul des valeurs numériques des solutions revient à calculer les racines d'un polynôme univarié (à une seule variable). Ce n'est toutefois pas aussi simple qu'il n'y paraît car c'est généralement un polynôme de degré élevé ayant de très grands coefficients [14].

Les étapes de l'algorithme

Rappelons ici qu'une phase critique qui intervient au cours des opérations permettant l'orientation relative de deux images est celle qui consiste à identifier, sur deux photographies différentes, les deux points qui sont les images d'un même point du terrain. Pour cette partie du travail, il n'est pas besoin que ce point terrain dispose de coordonnées connues, par contre il est indispensable

que l'on puisse garantir qu'un point pris dans une image est bien homologue d'un point pris dans l'autre, c'est-à-dire que ces deux images correspondent à un même point du terrain. Pour automatiser cette phase, on commence par extraire de chaque image des points faciles à pointer, et logiquement, si ces points ont été bien choisis, on parvient ensuite à mettre en correspondance la plupart des points d'une image avec ceux de l'autre dans la zone vue en stéréoscopie : on appelle ces points les points d'intérêt. Bien évidemment il est important d'éviter les points mal définis (p. ex. pris le long d'une bordure), ceux qui n'ont pas de définition géométrique stable (p. ex. bordure d'une ombre, intersection de lignes qui ne sont pas dans un même plan), etc. L'automatisation de l'extraction des points d'intérêt est directement liée à l'emploi d'images numériques, et l'un des premiers outils utilisés a été publié par C. Harris en 1986 [6] : son détecteur est basé sur l'extraction automatique de coins, et si assez rapidement ses insuffisances ont été connues, sa simplicité d'implémentation en a fait un outil très employé. Néanmoins, il a fallu attendre les résultats de deux décennies de recherche pour disposer d'une méthode réellement plus fiable, la méthode SIFT.

La méthode SIFT de Lowe, 2004 (*Scale Invariant Feature Transform*) [11] permet d'obtenir des points d'intérêt dont la détermination est très peu sensible à des changements, même importants, de facteurs d'échelle et d'orientation, et aussi assez peu sensibles aux variations locales de radiométrie (différences d'éclairage, différences de point de vue, etc.), toutes sortes de situations rencontrées très fréquemment en vision par ordinateur et où le détecteur de Harris est souvent tenu en échec. Elle est basée, entre autres, sur la considération suivante : dans une image, appliquer un changement d'échelle à un détail dont la répartition de densités radiométriques présente une forme gaussienne fournit un détail identique, à un facteur d'échelle radiométrique près.

Le traitement SIFT commence, pour chaque image, par une convolution suc-

■ ■ ■ cessive avec des imagerie de gausiennes de tailles progressivement croissantes. Puis une soustraction entre ces images convoluées successives est effectuée, permettant d'identifier un premier ensemble de points candidats, déjà a priori très peu sensibles aux changements d'orientation et d'échelle. Un premier filtrage est alors effectué pour éliminer les candidats ayant un contraste faible, ou qui sont mal localisables car situés sur un linéament.

L'étape suivante consiste, pour chacun de ces candidats, à identifier une ou plusieurs directions particulières au sein des gradients effectués dans toutes les directions autour du point. On construit alors, en utilisant ces directions comme références, un descripteur des gradients de l'environnement immédiat du point, sorte de carte d'identité du point en généralement 128 valeurs numériques, extrêmement peu sensible à l'orientation et à l'échelle du voisinage, ainsi qu'aux différences d'éclairément comme on le comprend aisément.

Chaque point candidat ainsi équipé peut alors participer à une grande opération de mise en correspondance avec les candidats issus de l'autre image, avec tout l'outillage mathématique que l'on peut envisager actuellement pour que l'opération aille vite et soit fiable (QuadTree, RanSac [5], etc.), mais la base de cette rapidité tient au travail de

tri et d'orientation qui a été fait en amont. Sur des images de taille réduite (500 x 500 pixels), on peut généralement identifier plusieurs milliers de points d'intérêt, ce nombre dépend du réglage plus ou moins critique de nombreux paramètres, à toutes les étapes du calcul. Il est clair qu'avec un nombre aussi important de points trouvés, la mise en correspondance peut être aisément assortie de critères de qualité très stricts, et que même ainsi, il reste couramment plusieurs centaines de points d'intérêt correctement extraits.

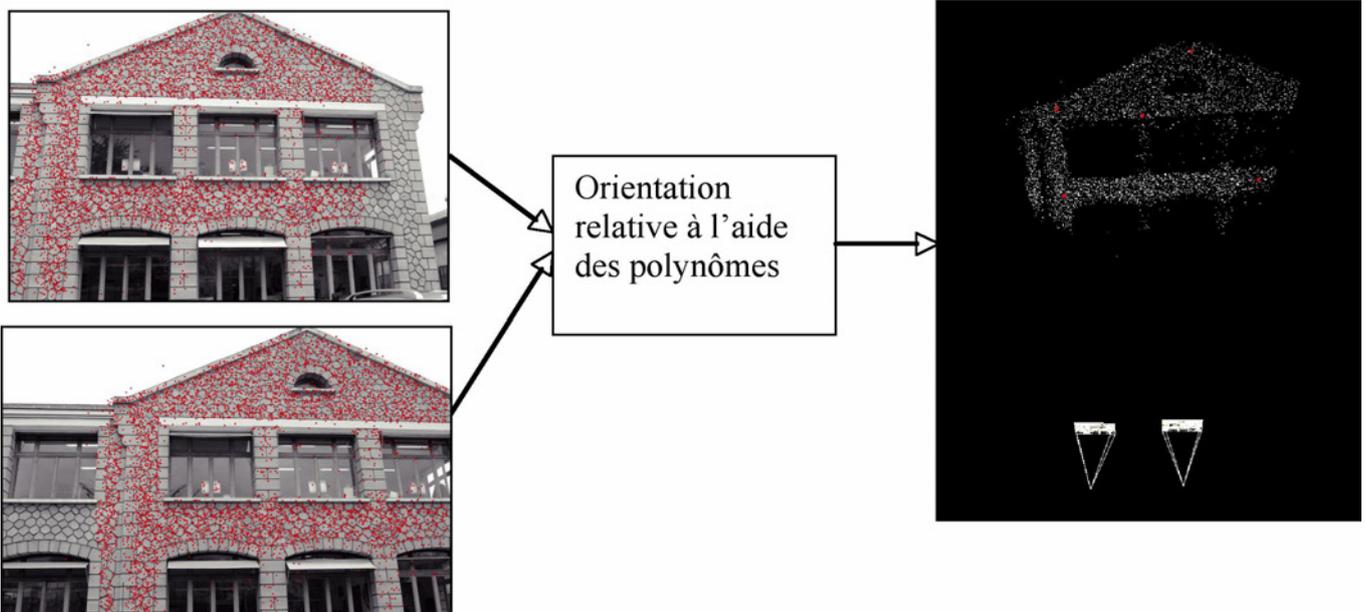
L'algorithme qui a été développé ici comporte donc trois étapes, tout d'abord une détection automatique des points de liaisons (ou points d'intérêt) avec la méthode SIFT, une estimation robuste basée sur la méthode RanSac, et enfin une estimation de l'orientation relative à l'aide des bases de Gröbner.

Résultats

Tout d'abord, plusieurs séries de points présentant des situations géométriques différentes ont été simulées, afin de valider le bon comportement des algorithmes dans des situations différentes, et y compris celles réputées difficiles (en particulier base très courte, ou points plus ou moins coplanaires). Ensuite, deux séries d'acquisitions d'images de terrain sur des bâtiments

ont été effectuées, puis traitées, tout d'abord la façade de l'école des Mines de Paris, puis l'un des bâtiments de l'IGN à Saint-Mandé. Nous présentons ici un exemple tiré de cette dernière acquisition, avec deux images de 3008x2000 pixels. Il s'agit donc d'un cas où les points sont coplanaires, un cas réputé défavorable et non fonctionnel pour les algorithmes de cinq points existants. Sur ces images, ont été reportés en rouge les 5562 points SIFT (parmi 5 616 points extraits a priori) qui ont été retenus après l'appariement automatique et le filtrage RanSac. Le résultat du traitement avec les algorithmes présentés ici apparaissent sur l'image ci-dessous à droite : l'orientation relative a été obtenue à partir de cinq de ces points, tirés au hasard parmi la population ainsi filtrée, la seule contrainte qui a été appliquée porte sur la distance minimale qui a été imposée entre ces cinq points : dans chaque image, ces points respectent une distance minimale de 200 pixels entre eux, ce qui ne représente pas une contrainte sévère vu la taille de chaque image.

Dans l'image de droite, compte tenu de l'orientation relative qui a été calculée, les points d'intérêt sont désormais en 3D, les points de prise de vue calculés en même temps apparaissent à leur place exacte. Les 5 points ayant servi au calcul ont été représentés en rouge.



Perspectives

Ce travail a été développé afin de parvenir à un tableau d'assemblage automatique, permettant une aide à l'acquisition d'images sur le terrain. En outre, il ouvre de nombreuses perspectives pour accélérer la phase d'orientation relative d'images, rapprochant ainsi les méthodes de photogrammétrie de celles, orientées vers les applications temps réel, de la communauté de la vision par ordinateur. Rappelons ici que l'originalité du procédé tient à ce que le calcul de l'orientation relative est obtenu sans l'aide d'aucune information externe, et qu'elle est donc entièrement automatique. Et par rapport aux méthodes équivalentes développées jusqu'ici, celle-ci peut fonctionner avec le nombre le plus petit possible permis par la théorie, soit 5 points identifiés comme homologues dans les deux images, et fonctionne encore avec des points coplanaires, comme l'exemple présenté ci-dessus le montre bien. ●

Remerciements

Nous tenons à remercier ici Erwan Houzay et Bertrand Cannelle, du Laboratoire MATIS, pour l'aide apportée à l'acquisition des images ayant servi aux tests et aux calculs de mise en œuvre de la méthode SIFT.

Contact

Mahzad KALANTARI

Laboratoire MATIS
Méthodes d'Analyse et de Traitement
d'Images pour la Stéréorestitution
2 - 4, avenue Pasteur
94 165 Saint-Mandé Cedex
mahzad.kalantari@ign.fr

Franck JUNG

Franck.Jung@equipement.gouv.fr

Bibliographie

[1] Buchberger, B. (1965). Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal. Mathematical Institute. Austria, University of Innsbruck. PhD.

[2] Cox, D., Little, J., O'Shea, D., 1997. Ideals, Varieties, and Algorithms. Springer-Verlag New York.

[3] Cox, D., Little, J., O'Shea, D., 1998. Using Algebraic Geometry, Springer-Verlag New York.

[4] Faugère, J. C. (1999). "A New Efficient Algorithm for Computing Grobner bases (F_4)". Journal of Pure and Applied Algebra 139.

[5] Fischler, M., Bolles, R., 1981. Random Sample Consensus: a Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography, Commun. Assoc. Comp. Mach. 24:381-395.

[6] Harris, C. and Stephens, M. 1988. A combined corner and edge detector. In Fourth Alvey Vision Conference, Manchester, UK, pp. 147-151

[7] Hartley R., Zisserman, A., 2000. Multiple View Geometry in Computer Vision, Cambridge University Press, ISBN 0-521-62304-9.

[8] Kasser M, Egels Y., 2001. Photogrammétrie Numérique. Hermes-Sciences.

[9] Kruppa, E., 1913. Zur Ermittlung eines Objektes aus zwei Perspektiven mit Innerer Orientierung, Sitz.-Ber. Akad.Wiss.,Wien, Math. Naturw. Kl., Abt. IIa., 122:1939-1948.

[10] Longuet-Higgins, H., 1981. A Computer Algorithm for Reconstructing a Scene from Two Projections, Nature, 293(10):133-135.

[11] Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. Int. Journal of Computer Vision 60(2), 91-110.

[12] Nister, D., 2004. An Efficient Solution to the Five-Point Relative Pose Problem, IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(6):756-770.

[13] Philip, J., 1996. A Non-Iterative Algorithm for Determining all Essential Matrices Corresponding to Five Point Pairs, Photogrammetric Record, 15(88):589-599.

[14] Rouillier, F. (1999). Solving zero-dimensional systems through the rational univariate representation. Journal of Applicable Algebra in Engineering, Communication and Computing 9.

[15] Eric Schost. cours de mathématique de l'école Polytechnique

[16] Systèmes Polynomiaux, Arithmétiques, Calculs Efficaces et Sûrs. Projet SPACE <http://www.spaces.lip6.fr/>

[17] Stewénius, H., Engels, C., and Nister, D. (2006). Recent developments on direct relative orientation. ISPRS Journal of Photogrammetry.

ABSTRACT

The calculation of the extrinsic parameters of a camera from several views with the help of the points of interests is the basis of all photogrammetric applications as well as computer vision ones. The major inconvenience of the photogrammetric approach, based on the condition of colinearity, is due to the non linearity of the problem, and thus approached values are necessary. In terrestrial applications, these initial values are unknown, and for that reason the community of computer vision since several years was interested to the calculation of the extrinsic parameters of the camera using linear equations, based on the classical condition of coplanarity, removing the need for approached values. In a case where the internal parameters of the camera are known, the calculation of the extrinsic parameters is often performed with help the so-called essential matrix. A lot of approaches have been finalized to solve this matrix. The most known are the algorithms said of the 8, 7 or 6 points. But the condition of coplanarity only has 5 degrees of liberty. Nister has proposed a polynomial resolution of the essential matrix, requesting only 5 points, but this approach has been reported as not working with coplanar points. In this article we present therefore a new method of direct resolution of the rotation and the translation from 5 points. Our method doesn't use the essential matrix, requires a polynomial resolution, using the bases of Gröbner, and behaves correctly with coplanar points. This method is also suitable for panoramic images, which is a significant comparative advantage.