

Réalisation d'un système de stéréovision mobile routier

■ Nicolas JANVIER

Dans le cadre de l'enrichissement des bases de données routières, l'élaboration d'un système de lever topométrique mobile s'impose comme étant la solution incontestable. Ainsi, le Laboratoire de Topométrie de l'Ecole Polytechnique Fédérale de Lausanne développe le projet Photobus permettant de déterminer précisément la géométrie d'une route. Mon sujet de travail de diplôme s'imbrique dans cette problématique et se propose de réaliser un système de stéréovision mobile de précision submétrique, en s'inscrivant dans une stratégie low cost et open source. Sa singularité réside principalement dans l'utilisation de capteurs d'imagerie basés sur la technologie CMOS. Le concept est l'acquisition de clichés géoréférencés afin d'autoriser une restitution photogrammétrique pour localiser les éléments d'intérêts (c'est-à-dire l'axe, les panneaux de signalisation et la largeur des voies), à partir d'un système GPS bi-antenne.

Le Laboratoire de Topométrie de l'Ecole Polytechnique Fédérale de Lausanne met au point un système de lever topométrique mobile pour déterminer la géométrie d'une route : le Photobus. Ce projet s'inscrit dans le cadre de l'enrichissement des bases de données routières, via la collecte d'informations telles que la position précise de l'axe de la route, des panneaux de signalisation, ainsi que la détermination du nombre et de la largeur des voies. Ce travail de fin d'études s'inscrit dans le cadre de ce projet et consiste en la réalisation d'un système de stéréovision mobile de précision submétrique en respectant une stratégie de prix bas pour la partie matériel, et en privilégiant l'utilisation de logiciels issus de la communauté libre d'Internet.

Un système de stéréovision mobile est un cas particulier de "mobile mapping system" puisqu'il résulte de la fusion d'une composante localisation, via un système GPS bi-antenne dans ce travail, et d'une composante imagerie, soit ici deux caméras de type CMOS (Complementary Metal Oxyde Semiconductor).

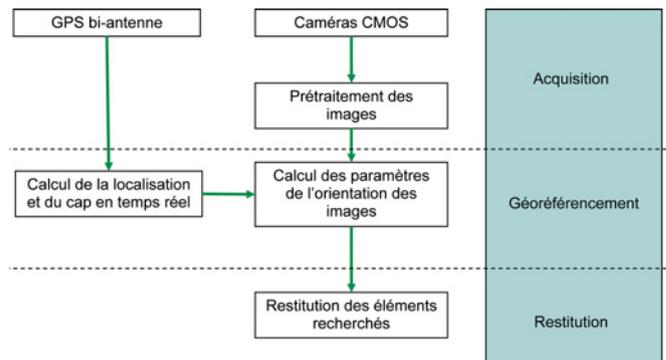


Figure 1 : Principe de fonctionnement du système de stéréovision mobile.

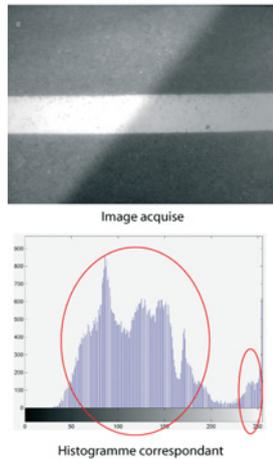
L'investigation a sensiblement été orientée sur la composante imagerie, et la stratégie open source a nécessité la conception et la réalisation de programmes pour l'orientation des images et la restitution des objets et ce, à partir de bibliothèques librement accessibles sur Internet.

Généralités sur les capteurs d'images numériques

Un capteur d'images numériques exploite l'effet photoélectrique de manière à mesurer l'intensité du flux lumineux incident. Les deux types de capteur existants sont les capteurs CCD (Charge-Coupled Device) et les capteurs CMOS. La différence primordiale entre ces deux technologies réside dans l'étape de conversion charge tension. Pour un capteur CCD, cette phase est effectuée juste avant la diffusion de l'image vers l'extérieur, tandis que pour un capteur CMOS, elle a lieu au niveau même du pixel.

Ces deux technologies possèdent un rendement quantique similaire dans la gamme des visibles, de l'ordre de 90%. Le CCD est incontestablement la référence pour les applications requérant un faible niveau de bruit, une grande résolution d'image et pour les applications à faible éclairage. Outre un prix et un encombrement moindres à matrice pixellaire équivalente, la technologie CMOS confère aux images acquises un contenu plus riche, mais au détriment de la qualité. Elle autorise aussi un accès indépendant à chaque pixel par une technique d'adressage x-y et permet ainsi de focaliser l'acquisition sur des régions d'intérêt. Enfin, cette technologie offre la possibilité de rassembler l'ensemble des fonctions d'une caméra dans un seul boîtier et permet même de réaliser des capteurs "intelligents" qui embarquent des fonctionnalités de traitement d'image. Elle présente une vitesse d'acquisition élevée

Capteur CMOS logarithmique



Capteur CCD

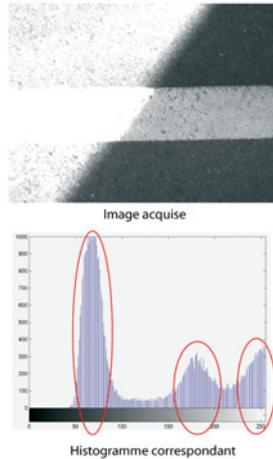


Figure 2 : Comparaison d'une image acquise avec un capteur CMOS logarithmique et un CCD.

et le traitement peut être effectué dans la foulée. La technologie CMOS logarithmique est parfaitement adaptée au Photobus puisqu'elle possède une bonne réaction à l'illumination en évitant la saturation. Ceci est particulièrement utile en lever routier où les conditions d'éclairément varient fortement entre les zones d'ombre et de soleil.

Etude du capteur d'imagerie Ethernam

L'Ethernam, fabriquée par Neuricam, est un système de vision complet combinant les fonctions d'acquisition et de traitement d'images dans un boîtier compact. Les images brutes ou interprétées sont transmises à un ordinateur client via une connexion Ethernet 10 Mbps. Une interface série RS-232 et trois lignes I/O opto-isolées autorisent une acquisition d'image synchronisée avec le temps GPS. Cette caméra est équipée d'un capteur monochromatique CMOS de taille VGA. La réponse du pixel est une fonction logarithmique de l'intensité lumineuse, conférant au capteur une gamme dynamique de six décades (120 dB). Comme pour tout capteur

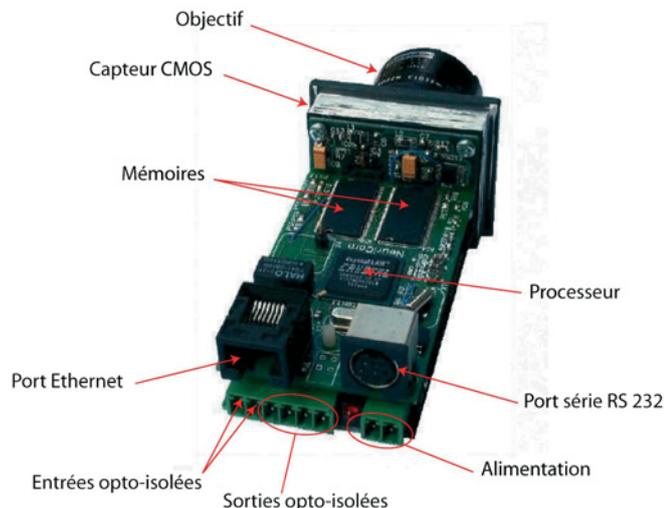


Figure 3 : Capteur d'imagerie Ethernam.

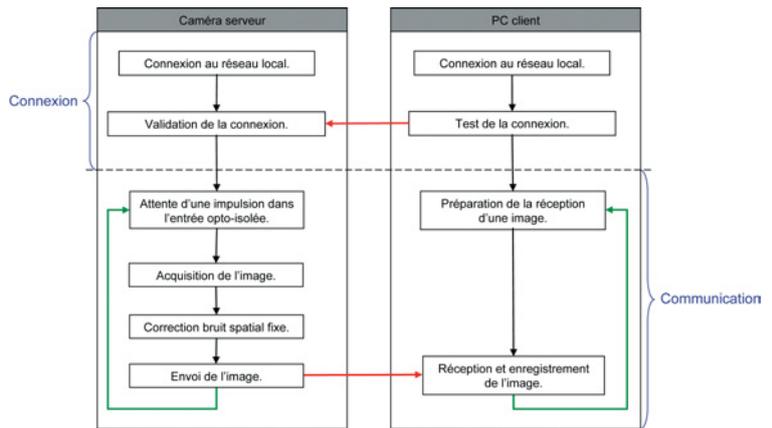


Figure 4 : Architecture et interaction des programmes serveur et client.

CMOS, les images brutes sont fortement entachées du bruit spatial fixe. La correction peut directement être effectuée sur la caméra avant le transfert vers l'ordinateur client.

L'Ethernam est entièrement programmable, sa gestion se découpe en trois parties : communication, acquisition et traitement de l'image. L'interaction de la caméra avec l'ordinateur fonctionne selon le principe du LAN (Local Area Network ou réseau local). Deux programmes interagissent : un serveur pour l'acquisition et le traitement des images par la caméra, et un client pour l'enregistrement des images sur l'ordinateur. L'architecture des programmes serveur et client est similaire et composée de deux séquences : une routine de connexion et une boucle de communication, comme illustrée sur la figure 4. Le traitement effectué se limite à la correction du bruit spatial fixe afin de conserver une cadence d'acquisition des images élevée.

Enfin, l'étalonnage des deux caméras a été effectué à l'aide de Camera Calibration Toolbox for Matlab de J.-Y. Bouguet, gratuitement disponible sur Internet¹. L'avantage de cette boîte d'outil est qu'elle ne nécessite pas de polygone d'étalonnage puisqu'elle exploite les images d'un damier de carrés de côtés connus. La mise en œuvre est ainsi simplifiée, mais l'estimation des paramètres, notamment celle de la focale, est moins optimal qu'avec un polygone car les points connus sont disposés sur une surface

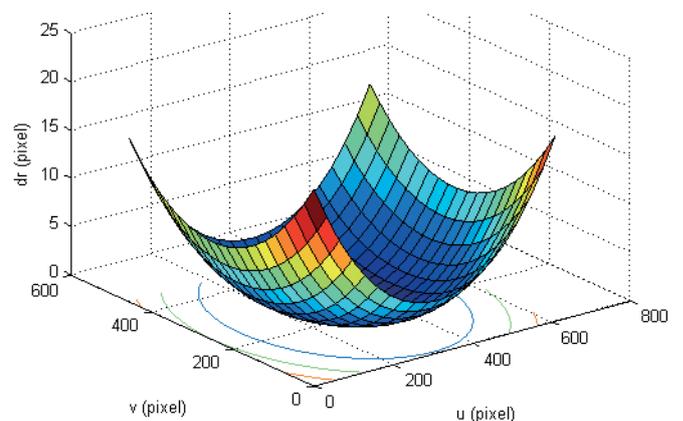


Figure 5 : Modélisation de la distorsion pour une Ethernam avec CCT for Matlab de J.-Y. Bouguet.

■ ■ ■ et non dans un volume. Les prises de vue du damier sont réalisées à différentes distances et sous plusieurs angles afin d'estimer au mieux les paramètres intrinsèques. Le calcul des éléments de l'étalonnage est basé sur l'exploitation de la relation entre les coordonnées image et terrain des points mesurés, soit l'équation de colinéarité. En connaissant les deux coordonnées il est alors possible de calculer les corrections à apporter pour améliorer l'exactitude de la relation. La distorsion des deux objectifs est très importante en bord d'image, puisque de l'ordre d'une quinzaine de pixels. Des comparaisons avec d'autres solutions de calibration, comme BINGO, ont été menées afin de vérifier la qualité de l'étalonnage réalisé.

Conception et réalisation du système de stéréovision mobile

Le système de stéréovision mobile est une plateforme regroupant les deux capteurs d'imagerie et le dispositif de localisation afin de collecter des couples d'images géoréférencées pour restituer les objets d'intérêt. Une telle opération nécessite la connaissance des paramètres d'orientation interne et externe. Les premiers ont été déterminés lors de l'étalonnage des caméras et sont constants au cours du temps. Les seconds doivent être collectés au cours du déplacement du système. Le géoréférencement direct des images sans points d'appui est basé sur l'extrapolation des paramètres d'orientation et conditionne directement la qualité de la restitution. Il passe par une localisation dynamique précise des caméras à partir des données du GPS. Ceci nécessite une synchronisation parfaite des capteurs d'imagerie et de localisation sur une base de temps commune. Le principe de synchronisation repose sur l'émission d'une impulsion par le port série d'un ordinateur vers un synchronisateur qui divise le signal en trois impulsions qui vont simultanément déclencher les deux acquisitions d'images et la datation par le récepteur GPS.

La réalisation d'une structure rigide garantissant un repositionnement subcentimétrique des caméras CMOS sur le toit du véhicule est indispensable pour fixer les positions relatives des deux antennes et des deux caméras. Afin de réduire les effets de la distorsion, les caméras sont inclinées de façon à ramener la zone d'intérêt au centre de l'image. La base entre les deux caméras influence directement la précision de la restitution, mais elle est

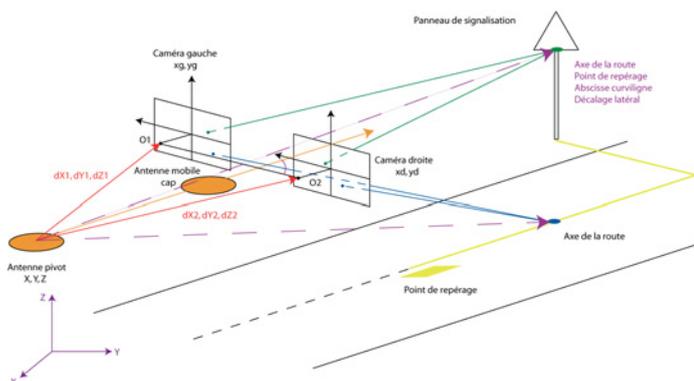


Figure 6 : Système de stéréovision mobile.

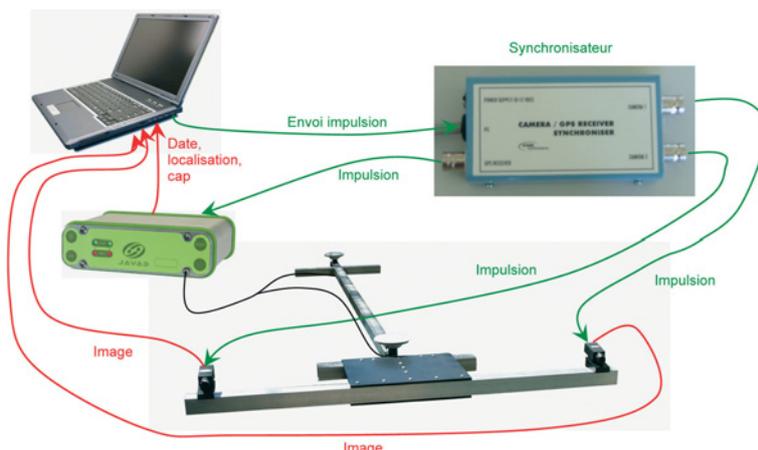


Figure 7 : Fonctionnement du système de stéréovision mobile conçu.

limitée par la largeur du véhicule et doit respecter la législation routière. La mise au point est réalisée à la distance du plan dans lequel sont situés les objets à restituer. Pratiquement, la profondeur de champ est déterminée en tenant compte de la diversité des scènes à observer variant d'une simple route à une 2x3 voies. Il est ensuite aisé d'effectuer la mise au point en fonction du premier et du dernier plan net. Le démarrage du système demande une bonne organisation. Le récepteur GPS doit être mis en marche en premier pour récupérer des observations d'initialisation et s'assurer d'un nombre de satellites visibles suffisant. Ensuite les caméras doivent être démarrées et connectées au LAN² séparément. Enfin le synchronisateur est activé et le bon fonctionnement du dispositif peut être vérifié. La vitesse du véhicule lors des deux tests réalisés oscillait entre 30 et 40 km/h tandis que la cadence d'acquisition est fixée à 2 images par seconde. Il est primordial de garder une vitesse constante pour un espacement entre les couples d'image régulier. Il existe en effet une interaction entre la vitesse du système, la cadence d'acquisition et l'espacement entre deux couples successifs. Le nombre de couples sur lesquels un point sera visible, donc la qualité de restitution, dépend de la vitesse du véhicule.

Géoréférencement direct des images

Le concept du géoréférencement direct autorise l'acquisition d'images géoréférencées, c'est-à-dire que chaque cliché se voit attribuer indépendamment des autres ses paramètres de localisation et d'orientation. Dans le cadre de ce travail, le géoréférencement direct est basé sur la transformation repère antenne - repère image, qui est fixe au cours du déplacement du véhicule grâce à la structure. Cette transformation implique l'expression des positions des deux antennes GPS dans les repères images. Les deux antennes définissent un repère, dénommé repère antenne, dont l'origine est l'antenne arrière, le plan xOy est horizontal, l'axe Ox dirigé selon la

(1) http://www.vision.caltech.edu/bouguetj/calib_doc/

(2) Local Network Area ou réseau local, ici le protocole d'accès est Ethernet.

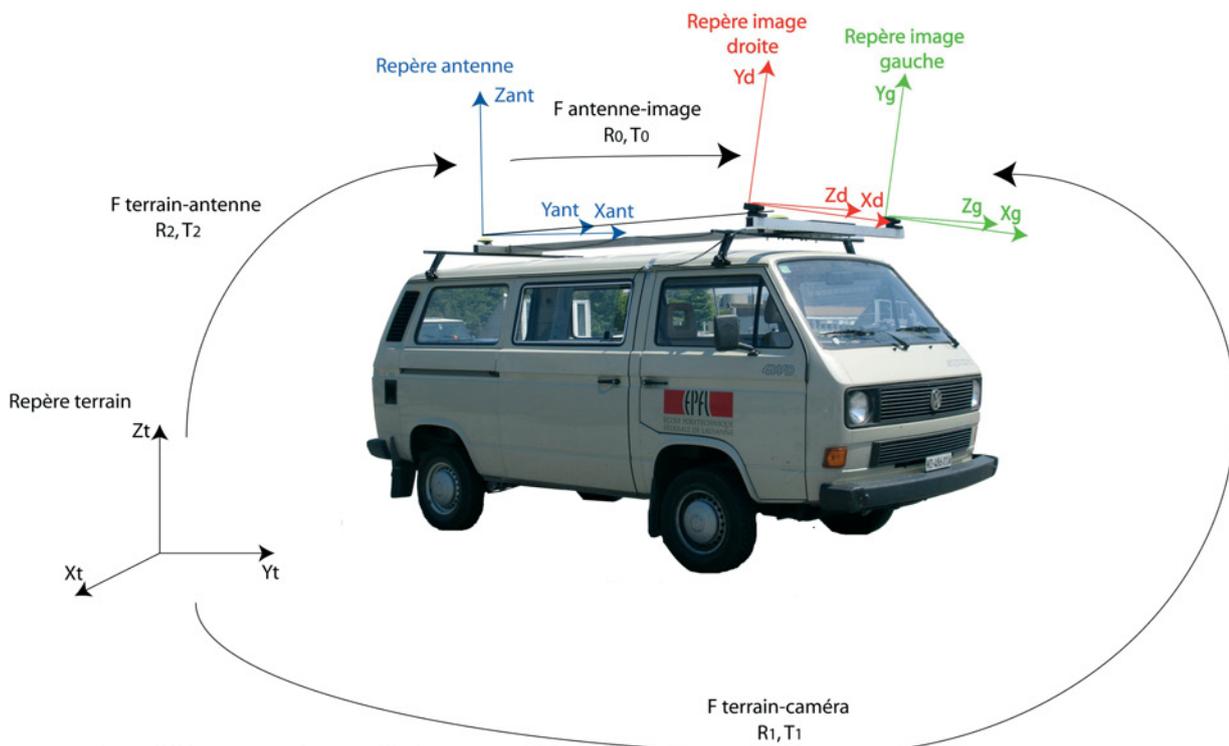


Figure 8 : Les différents repères utilisés.

droite définie par les deux antennes et l'axe Oz vertical. L'axe Oy complète le repère de façon à ce qu'il forme une base orthogonale directe. Les deux repères images sont donc fixes par rapport au repère antenne. Cette transformation sera alors déterminée lors d'un calibrage et ensuite utilisée pour géoréférencer les images acquises en la combinant avec la transformation entre le repère terrain, afin d'obtenir la localisation et l'orientation des images dans le repère terrain à partir des coordonnées des deux antennes.

Les passages entre ces repères sont régis par des transformations 3D qui peuvent être modélisées de la façon suivante :

$$\text{Repère antenne - caméra } F_0 : \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{caméra}} = R_0 \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{antenne}} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} = R_0 \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{antenne}} + T_0$$

$$\text{Repère caméra - terrain } F_1 : \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{terrain}} = R_1 \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{caméra}} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} = R_1 \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{caméra}} + T_1$$

$$\text{Repère terrain - antenne } F_2 : \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{antenne}} = R_2 \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{terrain}} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} = R_2 \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{terrain}} + T_2$$

Le principe de détermination de la fonction de géoréférencement consiste à réaliser des mesures GPS et acquérir des couples d'images simultanées devant un polygone composé de cibles connues en coordonnées. En combinant l'acquisition de mesures GPS avec des prises d'image du polygone, il est possible de calculer la transformation repère caméra - repère terrain (F1), grâce à l'équation de colinéarité qui fournit la matrice de rotation entre le repère antenne et terrain ainsi que les coordonnées du sommet dans le repère terrain. Ensuite, la transformation entre le repère terrain et celui des antennes GPS est aisément obtenu en résolvant l'équation F2,

après avoir exprimées les coordonnées terrain des antennes dans le repère antenne :

$$X_{\text{antenne arri\ere}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}; X_{\text{antenne avant}} = \begin{bmatrix} \sqrt{(X_{\text{terrain arri\ere}} - X_{\text{terrain avant}})^2 + (Y_{\text{terrain arri\ere}} - Y_{\text{terrain avant}})^2} \\ 0 \\ Z_{\text{terrain arri\ere}} - Z_{\text{terrain avant}} \end{bmatrix}$$

Les éléments recherchés sont alors obtenus en composant les transformations F1 et F2 :

$$R_0 \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{caméra}} + T_0 = R_2 \cdot \left(R_1 \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{\text{caméra}} + T_1 \right) + T_2 \quad \text{soit} \begin{cases} R_0 = R_2 \cdot R_1 \\ T_0 = R_2 \cdot T_1 + T_2 \end{cases}$$

Ainsi au moment de la restitution, chaque image sera géoréférencée à l'aide des formules suivantes :

$$X_S^{\text{terrain}} = X_{\text{arri\ere}}^{\text{terrain}} + R_2^{-1} \cdot T_0$$

$$R = R_2^{-1} \cdot R_0$$

La transformation F2 dépend de la position du véhicule au moment de la mesure, elle est donc calculée pour chaque couple d'images. Il convient de remarquer que le roulis et le tangage sont négligés. Du fait de la position des deux antennes, le roulis n'est pas accessible, mais comme le dévers de la route varie de 2.5% à 5%, l'erreur commise est minime puisqu'inférieure au centimètre. Les pentes rencontrées usuellement sur les routes sont de l'ordre de 6%, soit un tangage de 3.5°. Le cadre de ce travail étant l'évaluation de la stéréovision mobile CMOS, les acquisitions seront réalisées en milieu découvert et relativement plat, les pentes seront donc largement inférieures à 6%. Le tangage sera alors de l'ordre de 1°, il peut donc être négligé.

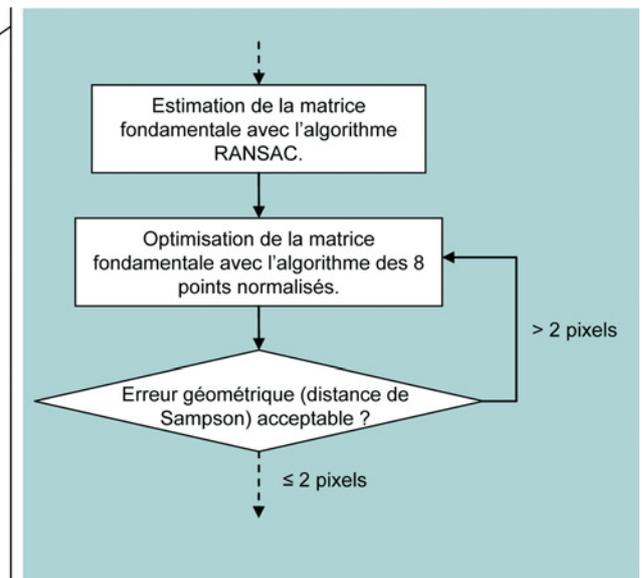
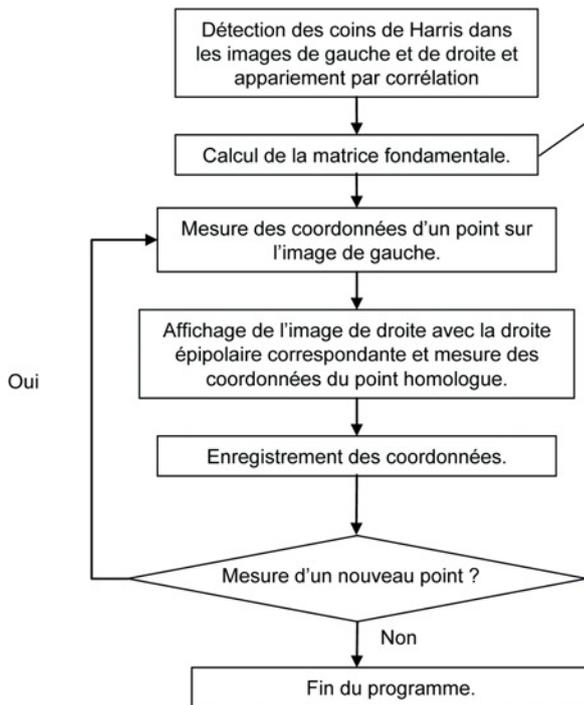


Figure 9 : Organigramme du module de mesure.

■ ■ ■ Identification des caractéristiques d'une route par restitution photogrammétrique

La restitution implique la mesure des coordonnées image des points d'intérêt. Elles sont ensuite introduites dans les équations inverses de la colinéarité pour l'obtention des coordonnées terrain :

$$\begin{cases} X = X_S + (Z - Z_S) \frac{r'_{11}x + r'_{12}y - r'_{13}c}{r'_{31}x + r'_{32}y - r'_{33}c} \\ Y = Y_S + (Z - Z_S) \frac{r'_{21}x + r'_{22}y - r'_{23}c}{r'_{31}x + r'_{32}y - r'_{33}c} \end{cases}$$

avec (X,Y,Z) les coordonnées terrain du point mesuré, (x,y) les coordonnées image du point mesuré, (X_S,Y_S,Z_S) les coordonnées terrain du sommet perspectif de la caméra, c la focale et r'_{ij} les termes de la matrice de rotation de l'image considérée.

Tout d'abord, un module de mesure des coordonnées image a été développé en utilisant la géométrie épipolaire afin de faciliter l'identification des points d'intérêt dans l'image de droite. Sa mise en œuvre s'appuie sur deux boîtes d'outils librement accessibles sur Internet : *MATLAB Functions for Computer Vision and Image Analysis*³ de Peter Kovesi et *Structure and Motion Toolkit for Matlab*⁴ de Philip Torr. La mise en œuvre de la géométrie épipolaire repose sur un calcul de la matrice fondamentale à partir des coins de Harris dans les images de gauche et de droite, qui sont ensuite appariés via une comparaison basée sur le maximum de corrélation. La matrice fondamentale est ensuite estimée à l'aide de l'algorithme de RANSAC, qui permet

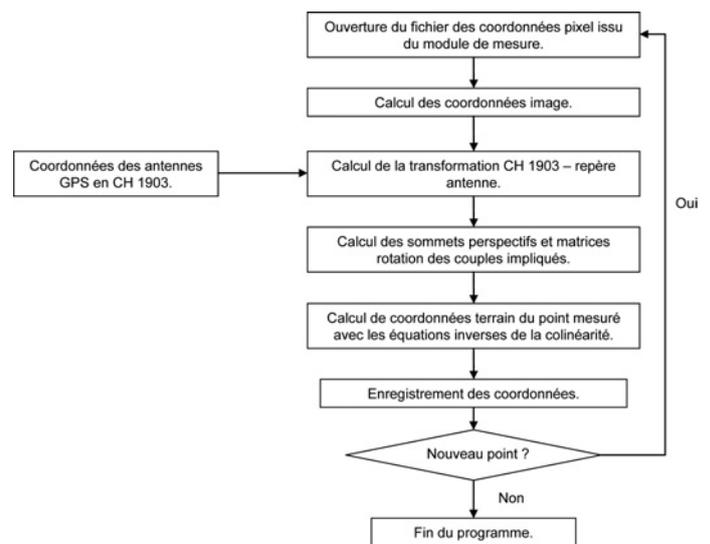


Figure 10 : Organigramme du module de restitution.

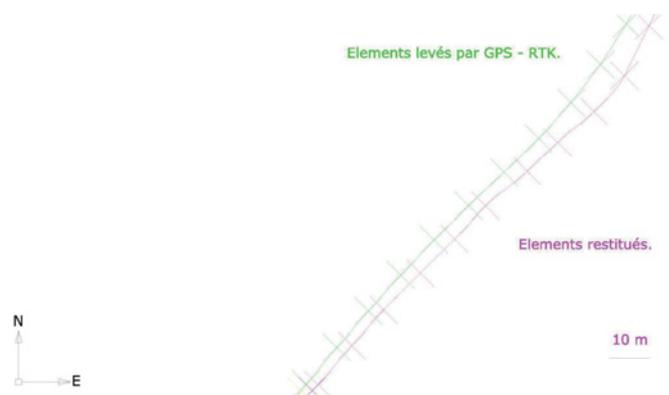


Figure 11 : Plan des axes obtenus par levé GPS-RTK et par restitution.

(3) <http://www.csse.uwa.edu.au/~pk/research/matlabfns/>

(4) <http://research.microsoft.com/~phil/torr/>

d'éliminer les erreurs d'appariement et d'obtenir un premier résultat fiable. Le calcul est ensuite affiné de façon itérative avec l'algorithme des 8 points (Longuet - Higgins, revu par Hartley - 1997), le critère d'itération étant une distance de Sampson inférieure à deux pixels.

Ensuite, un module de restitution a été conçu pour appliquer aux coordonnées précédemment mesurées, les équations inverses de la condition de la colinéarité en fonction des données de géoréférencement direct. Il est indispensable de s'assurer que les paramètres fixes dans la fonction de géoréférencement respectent bien la géométrie des deux caméras déterminée par la structure les accueillant.

Les premiers résultats mettent en évidence une erreur importante de 4 à 5 mètres dans le sens latéral. Mais la précision n'a pas pu être vraiment estimée puisqu'un problème de communication est venu perturber le test. En effet un décalage au niveau de la numérotation des images a été remarqué, l'image de gauche numéro 882 correspond à l'image de droite numéro 875. Un phénomène d'embouteillage a vraisemblablement eu lieu au niveau de la connexion Ethernet.

Conclusion

L'étude des caméras CMOS logarithmiques montrent qu'elles sont parfaitement adaptées à la capture d'image en extérieur, notamment en zones soumises à de brusques variations d'éclairément. De plus, ces capteurs intelligents autorisent la réalisation de traitements embarqués afin de soulager l'ordinateur hébergeant le logiciel de mobile mapping de certaines tâches de post-traitement. De part leur coût, elles permettent de respecter pleinement la tactique de prix bas.

La conception des programmes nécessaires à la mise en place du processus photogrammétrique a pu en partie s'appuyer sur des boîtes d'outils performantes et gratuitement disponibles sur Internet. L'étalonnage des caméras et la détermination de la géométrie épipolaire ont été réalisés par ce biais avec satisfaction. Les logiciels de mesure et de restitution ont été spécialement développés en utilisant les techniques de la photogrammétrie. La stratégie d'*open source* est ainsi respectée.

La conception du dispositif d'acquisition a débouché sur l'élaboration d'une structure accueillant les capteurs et la détermination de la méthode de synchronisation. Le géoréférencement a ensuite pu être réalisé à l'aide d'un système GPS bi-antenne exploitant le concept de station de référence mobile.

Les éléments essentiels dans l'élaboration d'un système de stéréovision mobile bon marché et efficace sont une parfaite synchronisation des capteurs d'imagerie et de localisation, un étalonnage des caméras optimal, ainsi qu'un géoréférencement direct précis. Ces deux dernières conditions ont été remplies. ●

Bibliographie

- Chong. A. K. (2004)** *Optimizing the accuracy of a low-cost photogrammetric motion and study system*, University of Otago, New Zeland, The Photogrammetric Record 19(108) : 296-310, December.
- Gontran. H. (2005)** *Thèse en préparation*, Laboratoire de Topométrie, Ecole Polytechnique Fédérale de Lausanne.
- Gontran. H., Skaloud. J. et Gilléron. P.-Y. (2004)** *Photobus : en route vers la cartographie mobile en temps réel*, Laboratoire de Topométrie, Ecole Polytechnique Fédérale de Lausanne, Géomatique Suisse, septembre.
- Heikkilä. J., Silvén. O. (1997)** *A four step camera calibration procedure with implicit image correction*, Infotech Oulu and Department of Electrical Engineering, University of Oulu, Finland, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97), San Juan, Puerto Rico, p1106-1112.
- Luhmann. T. (2000)** *Nahbereichsphotogrammetrie*, Grundlagen, Methoden und Anwendungen, Wichmann, Heidelberg, Deutschland.
- Torr. P. H. S. (2002)** *A Structure and Motion Toolkit in Matlab "Interactive Adventures in S and M"*, Technical Report MSR-TR-2002-56, june.
- Notices d'utilisation Etherecam** *User Manual, Programmer Manual*, axLinux - cdb89712.

Contact

Nicolas JANVIER
nico.janvier@wanadoo.fr

ABSTRACT

Within the framework of the improvement of road databases, the design of a mobile mapping system asserts oneself to be the undisputable solution. In this way, the Geodetic Engineering Laboratory at the Swiss Federal Institute of Technology (EPFL) has developed the Photobus project to determine the geometry of a road with accuracy. The topic of my Master's Thesis tackles this problem and consists in carrying out a mobile stereovision system with a submetric accuracy, within the scope of a low cost and open source strategy. Its singularity resides in the use of image sensors based on CMOS technology. The concept is the acquisition of georeferenced pictures so as to allow a photogrammetric plotting of the elements of interest (i.e the road axis and signs, as well as the width of the lanes). The adopted strategy leads to realize the programs needed to plot, by using open source toolboxes if possible.