

Génération de cartes de disparité denses à partir d'images panoramiques cylindriques haute définition

Laurent SMADJA, Erwan BIGORGNE, Ryad BENOSMAN, Jean DEVARs

Laboratoire des Instruments et Systèmes d'Ile-de-France

Université Paris VI

BP 164, Tour 22-12 2ème étage

4 Place Jussieu, 75252 Paris Cedex 05, France

laurent.smadja@lis.jussieu.fr, erwan.bigorgne@lis.jussieu.fr

benosman@ccr.jussieu.fr, devars@ccr.jussieu.fr

Résumé – Cette étude porte sur le développement, l'utilisation d'un capteur panoramique cylindrique à très haute définition, et les cartes de profondeur obtenues à partir de deux vues de ce capteur.

Partant d'images calibrées, plusieurs étapes sont réalisées: détection et appariement de points d'intérêt, détermination de la SFM, appariement dynamique des lignes épipolaires de l'image.

Les résultats obtenus sont les cartes de disparité denses de 2048 * 3600 pixels du panorama étudié.

Abstract – This paper deals with creating and using a high definition cylindrical sensor, and disparity maps resulting from two scenes of this imaging system.

Starting from calibrated images, several steps are processed: feature detection and matching, SFM determination, dynamic matching of epipolar lines.

Results are dense 2048 * 3600 disparity maps corresponding to the observed panorama.

1 Introduction

La structure tridimensionnelle d'une scène peut être retrouvée à partir de plus de deux images prises à des endroits différents. Ce domaine, connu sous le nom de *Stéréovision* a été largement étudié depuis plus de 20 ans [9] et il existe aujourd'hui des algorithmes de reconstruction 3D denses à partir de séquences d'images [10]. Ces reconstructions sont très précises et s'appuient sur un mode opératoire parfaitement balisé mais restent dans le cadre de l'image perspective [8].

D'autre part, les capteurs panoramiques constituent de nos jours un des champs de recherche les plus actifs de la communauté. De par leur champ de vue à 360°, ils sont de plus en plus utilisés dans le guidage de robots ou la télésurveillance. Mais leur faible définition les empêchent définitivement d'obtenir des reconstructions précises. Cependant, de nombreux chercheurs appliquent à présent l'ensemble de la théorie développée dans le cadre des cameras classiques aux capteurs panoramiques [11]

2 Présentation du capteur

Le capteur développé est basé sur l'utilisation d'une caméra linéaire Tri-CCD de 2048 pixels, mise en rotation sur un axe vertical par un moteur à courant continu. Celui-ci envoie régulièrement, typiquement tous les dixièmes de degré, un signal à la caméra déclenchant l'acquisition d'une colonne. En concaténant les colonnes acquises au cours d'une rotation complète, on obtient une image panoramique RGB au format 2048*3600. La

caméra utilisée est une Jai CV L-103, barrette linéaire de 2048 pixels qui délivre des images RGB à un débit maximal de 2400 images par seconde, ce qui est largement plus que nécessaire. Elle est couplée à une Coreco Pc DIG comme carte d'acquisition et l'objectif choisi est un Nikon F-Or 28/80 mm, généralement réglé sur 28 mm. La fréquence du signal envoyé par le moteur conditionne la durée d'intégration de la barrette Tri-CCD; il est donc choisi en fonction des conditions expérimentales d'acquisition. La barrette est mise en rotation selon un axe passant par son centre focal, qu'il faut donc déterminer. Un procédé optique particulièrement simple dans sa mise en oeuvre a été développé. Cette méthode est très précise et permet de placer le centre optique de la caméra sur l'axe de rotation du capteur, et ainsi de connaître le rayon du cylindre image.

3 Détection, appariement

L'ensemble de l'algorithme est réalisé sur la base d'une collection de points d'intérêt correctement appariés. Leur détection repose sur la sélection des maxima locaux d'une mesure caractéristique de la courbure locale du champs de gradient [2]. Cette mesure consiste à calculer les moyennes pondérées des produits vectoriels entre les vecteurs gradient au point considéré avec ses voisins, une mesure maximisée au niveau des coins.

$$m = \sum_{j \in V_i} M_j \left\| \overrightarrow{\text{grad}(P_i)} \wedge \overrightarrow{\text{grad}(P_j)} \right\|^2 \quad (1)$$

où P_i est le point considéré, V_i est le voisinage de P_i , et M_j correspond au coefficient d'un masque de pondération.

La mesure normalisée obtenue est donnée par :

$$k = \frac{I_x^2 \langle I_y^2 \rangle + I_y^2 \langle I_x^2 \rangle - 2I_x I_y \langle I_x I_y \rangle}{\langle I_x^2 \rangle + \langle I_y^2 \rangle} \quad (2)$$

où I_x et I_y sont les dérivées partielles de l'image $I(x,y)$ et $\langle . \rangle$ symbolise le produit de convolution par un masque de voisinage.

Afin d'apparier ces points de manière robuste, le signal couleur au voisinage de chacun de ces points est caractérisé à l'aide des moments de Zernike. Une normalisation adéquate de ces moments permet d'obtenir une signature présentant une invariance tant géométrique (invariance aux transformations du groupe des similitudes, quasi invariant du groupe des transformations projectives) que radiométriques (pour des transformations localement affine sur chaque bande du signal couleur) [1],[3].

L'appariement est essentiellement basé sur le seuillage et le classement des distances de Mahalanobis entre les vecteurs de caractéristiques. La matrice de covariance requise est calculée à partir d'un jeu adéquat de points d'intérêt. L'ensemble du processus inclut une condition de localité : deux points sont appariés si un nombre nécessaire de leurs plus proches voisins image sont suffisamment proches dans l'espace des caractéristiques. La figure 1 montre le résultat d'un tel appariement sur un détail des deux images panoramiques.

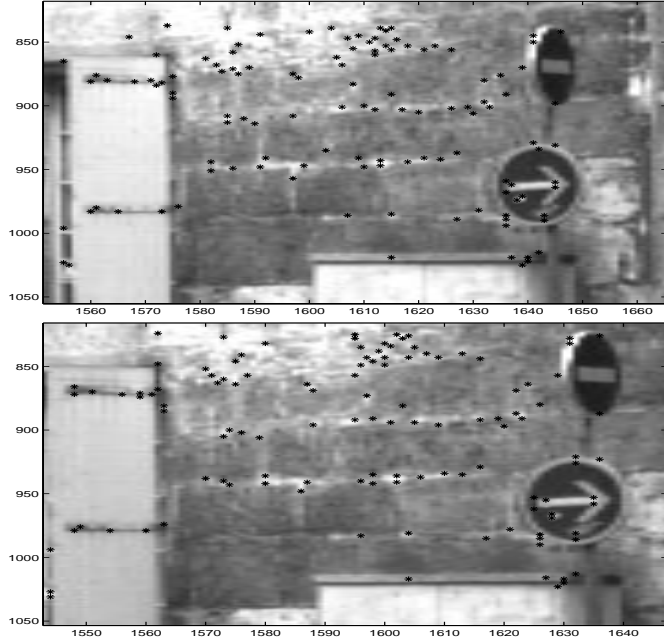


FIG. 1: Un couple d' "images" appariées

Cependant, compte tenu de la géométrie particulière du capteur (projection cylindrique), un fort taux de faux positifs (20%) reste à éliminer.

4 Détermination de la SFM et contrainte épipolaire

Le problème initial, posé par Faugeras et Maybank [5], est de déterminer le déplacement rigide entre deux vues perspectives d'une même scène, afin d'estimer une reconstruction de cette scène. Cet article pose les bases du problème de *Structure From Motion (SFM)* et introduit pour la première fois la *matrice essentielle E*, comme contrainte fondamentale dans l'étude des couples d'images stéréo. Cette matrice encode la contrainte de coplanarité (*la contrainte épipolaire*) inhérente à la projection d'un même point physique M sur deux plans image différents. Les points M_1 et M_2 , images de M doivent vérifier l'équation :

$$M_2^T E M_1 = 0 \quad (3)$$

Exprimée en coordonnées pixelliques, la matrice E devient la *matrice fondamentale F*, qui exprime évidemment la même contrainte mais dans le cadre d'images non calibrées. Beaucoup de chercheurs se sont concentrés sur les algorithmes de détermination robuste de la matrice F et une étude très complète peut être trouvée dans [12]. Cependant, l'ensemble de ces travaux s'applique aux caméras classiques et une adaptation aux procédés panoramiques s'effectue progressivement.

Dans [4], Chang et Hebert proposent une implémentation de la SFM dans le cadre de capteurs catadioptriques calibrés. S'appuyant sur la même démarche, notre approche s'applique au capteur cylindrique: les images sont calibrées, ce qui signifie que la focale $K_v f$ (exprimée en pixels) et la position de la projection du centre optique v_0 sur la barrette sont connues. Les coordonnées de points image u_i, v_i appariés nous renvoient alors des positions physiques M_i métriques sur les deux cylindres image dans leur propre repère *via* la transformation :

$$M_i = \begin{cases} x_i = K_v f \cos(K_u u_i) \\ y_i = K_v f \sin(K_u u_i) \\ z_i = v_0 - v_i \end{cases} \quad (4)$$

Ces couples de points doivent vérifier la contrainte 3, que l'on peut linéariser puis résoudre en première approximation par *Direct Linear Transform (DLT)*.

4.1 DLT

Chaque couple de points vérifiant l'équation 3 fournit une contrainte linéaire de la forme :

$$U_i e = 0 \quad (5)$$

où :

- U_i est un vecteur ligne à 9 coordonnées, produit tensoriel des points M_{i1} et M_{i2} .
- e correspond aux 9 éléments de la matrice essentielle E réorganisés en un vecteur colonne.

En concaténant tous les vecteurs U_i relatifs à tous les couples, on obtient une matrice U qui vérifie

$$U e = 0$$

Pour un nombre de couples supérieur à 8, le système est sur-déterminé [7] et se résout en choisissant le vecteur singulier associé à la plus petite valeur singulière de U [12].



FIG. 2: Un couple d'images panoramiques et quelques courbes épipolaires

Cependant, cette méthode est très sensible au bruit et une détection des points mal appariés durant la première phase s'avère nécessaire. On peut alors introduire des critères de différentes natures (algébriques et géométriques) permettant une pondération qui tend à annuler l'influence des U_i issus de points mal appariés.

4.2 RANSAC

La technique du RANSAC [6] est une méthode statistique permettant, par l'utilisation de *samples* aléatoires, d'estimer une transformation optimale entre deux jeux de points appariés et d'éliminer les faux appariements. Les paramètres épipolaires ie les translation, rotation entre les deux repères image sont finalement recalculés à partir de la base de points correctement appariés.

4.3 Contrainte épipolaire

Les lignes épipolaires de nos images cylindriques ayant été définies grace aux paramètres estimés par SFM, il nous faut à présent appairier l'ensemble des points de chaque image. La programmation dynamique nous permet d'appairier les lignes épipolaires deux à deux tout en autorisant des disparités assez élevées. Ce procédé est relativement coûteux en taille mémoire mais fournit des résultats très intéressants, particulièrement au niveau des forts gradients.

5 Résultats

Nous présentons ici un couple d'images panoramiques cylindriques haute définition, les courbes épipolaires obtenues après la détermination de la SFM et une carte de disparité au format des images d'origine.

Références

- [1] C. Achard, E. Bigorgne, and J. Devars. A sub-pixel and multispectral corner detector. *Proc. 15th Int. Conference on Pattern Recognition*, 971-974, September 2000.
- [2] E. Bigorgne, C. Achard, and J. Devars. An invariant local vector for content-based image retrieval. *Proc. 15th Int. Conference on Pattern Recognition*, pages 1019-1022, September 2000.
- [3] E. Bigorgne, C. Achard, and J. Devars. A local color descriptor for efficient scene-object recognition. *Proc. 11th Int. Conference on Image Analysis and Processing*, September 2000.
- [4] P. Chang and M. Hebert. Omni-directional structure from motion. *IEEE Workshop on Omnidirectional Vision (OMNIVIS)*, September 2000.
- [5] O. Faugeras and S. Maybank. Motion from point matches: multiplicity of solutions. *Int. Journal of Computer Vision*, 4:225-246, 1990.
- [6] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Communi-*

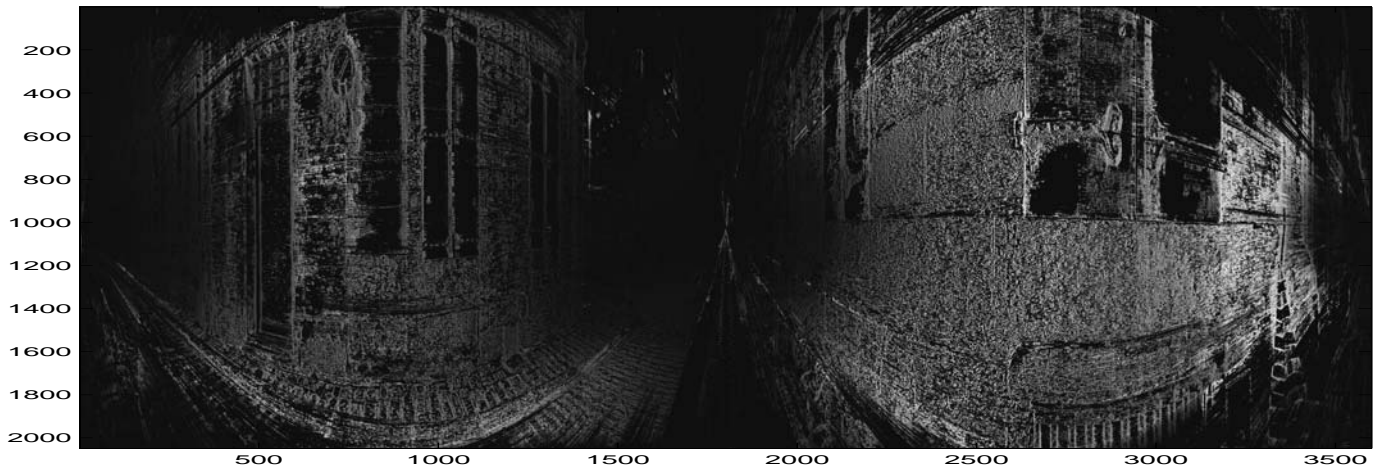


FIG. 3: La carte de disparité obtenue à partir de ce couple

- cation Association and Computing Machine, 24(6), 381-395, 1981.*
- [7] R. Hartley. In defence of 8-point algorithm. *Proc. 5th Int. Conference on Computer Vision, 1064-1070, 1995.*
- [8] R. Hartley and A. Zissermann. Multiple view geometry in computer vision. *Cambridge University Press, 2000.*
- [9] H. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature 293, 133-135, 1981.*
- [10] M. Pollefeys. Self-calibration and metric reconstruction from uncalibrated images sequences. *PHD Thesis, 1999.*
- [11] T. Svoboda. Central panoramic cameras design, geometry, egomotion. *PHD Thesis, 1999.*
- [12] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International Journal of Computer Vision, 27(2), 161-198, 1998.*