

Auto-calibrage à partir de deux vues et de la projection d'un motif lumineux

T. Molinier et D. Fofi

Le2i UMR CNRS 5158
IUT Le Creusot
Université de Bourgogne
12 rue de la Fonderie
F-71200 Le Creusot, France

Auto-Calibrage à partir de Deux Vues et de la Projection d'Un Motif Lumineux

Thierry Molinier, David Fofi

Le2i UMR CNRS 5158

IUT Le Creusot – Université de Bourgogne

12, rue de la Fonderie 71200 Le Creusot

d.fofi@iutlecreusot.u-bourgogne.fr

Résumé - Nous présentons les premiers résultats obtenus, ainsi que les perspectives envisagées, pour l'auto-calibrage d'un capteur de vision en lumière structurée imperceptible. La méthode choisie est basée sur l'estimation du tenseur trifocal reliant les caméras et le projecteur. A partir de celui-ci, nous extrayons les matrices de projection des trois éléments. A l'aide de quelques hypothèses couramment admises (pixel carré, rétine orthogonale, point principale au centre de l'image – mais distance focale libre) nous montrons qu'il est possible d'obtenir une reconstruction métrique (à une similarité près) de la scène observée.

Mots Clefs – *Lumière structurée imperceptible, auto-calibrage, reconstruction métrique.*

I. INTRODUCTION

La reconstruction 3D par capteur de vision actif – appelée *vision en lumière structurée* – nécessite l'emploi d'une ou plusieurs images et d'un motif projeté sur la scène observée [1]. Malheureusement, cette projection empêche de pouvoir obtenir les informations de texture et de couleur que l'on pourrait désirer plaquer sur une reconstruction 3D. Le but de cet article est d'exposer le fonctionnement de notre système trinoculaire qui utilise le principe de la lumière structurée imperceptible [6][10] pour contourner ce problème, ainsi que les premiers résultats obtenus grâce à des hypothèses prises sur les matrices de calibration du capteur.

Notre objectif principal est l'auto-calibrage du capteur et la reconstruction 3D dans un espace métrique, garantissant suffisamment de degrés de liberté au système pour pouvoir être utilisé dans le cadre d'une stratégie de vision active. A terme, nous voulons utiliser l'information de texture et de couleur fournie par le capteur pour obtenir une reconstruction dense et texturée de la scène.

Notre article sera découpé de la façon suivante : la deuxième section décrit le principe du capteur trifocal en lumière structurée imperceptible, les inconvénients et les contraintes qu'il impose ; la troisième section expose les raisons du choix du tenseur trifocal et son utilité, ainsi que les méthodes d'estimation que nous avons implémentées ; la quatrième section explique la méthode suivie pour atteindre la reconstruction métrique et la cinquième section présente nos premiers résultats. Notre conclusion décrit l'orientation prise par nos travaux et leurs développements futurs.

II. LUMIERE STRUCTUREE IMPERCEPTIBLE

A. Principe et conception du capteur

Notre capteur trifocal se compose de deux caméras et d'un projecteur (Fig. 1). La source lumineuse projette, à fréquence élevée, un motif structuré (diapositive ou image) suivi de son motif complémentaire, de sorte que la résultante forme une lumière uniforme (Fig. 2). L'une des caméras, que nous appellerons dans la suite de l'article *caméra 1*, est synchronisée à la projection du motif et dispose d'un temps d'intégration rapide ; la seconde caméra (*caméra 2*) est configurée avec un temps d'intégration plus long. En conséquence, la caméra 1 voit la scène avec le motif projeté alors que la caméra 2 voit la scène sous une illumination uniforme.

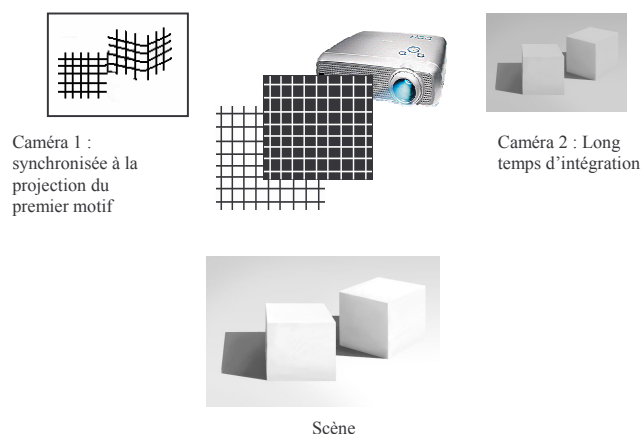


Fig. 1 Principe du dispositif.

Le principe psychophysique de la lumière structurée imperceptible repose sur l'effet visuel de papillotement lors de la projection d'un motif et de son complémentaire : en dessous de la FCF (Fréquence Critique de Fusion, fixée à 75 Hz selon [13]), l'œil perçoit l'alternance entre les deux images. Au-delà de cette fréquence critique, l'œil ne voit plus que la fusion du motif et de son complémentaire, soit une lumière uniforme.

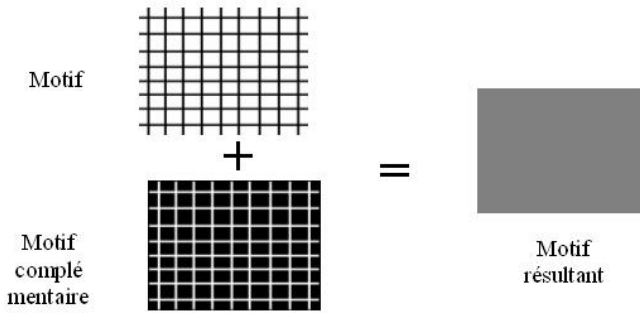


Fig.2 Le motif projeté et son complémentaire.

B. Avantages et contraintes

Un capteur de vision en lumière structurée imperceptible permet d'allier les avantages de la lumière structurée (reconstruction, mise en correspondance) à ceux de la vision classique (analyse de la texture, des couleurs, etc....) pour, finalement, obtenir une reconstruction 3D de la scène et le plaquage des textures et des couleurs sur les surfaces reconstruites. En outre, c'est un capteur actif non-invasif si l'on fixe une fréquence de papillotement supérieure à la FCF. D'où de nombreuses applications possibles en robotique, mesures anthropométriques, métrologie, profilométrie, contrôle industriel, etc.

Cependant, ce capteur impose quelques contraintes :

1. l'impossibilité de trouver des points de correspondance entre l'image du projecteur et la caméra 2 qui perçoit la scène sous une lumière uniforme ;
2. le capteur étant composé d'éléments hétérogènes (caméras et projecteur), dont les paramètres intrinsèques sont généralement différents, les méthodes d'auto-calibrage classiques ne sont pas applicables ;
3. un déplacement du projecteur provoque le glissement du motif sur les surfaces mesurées (et donc une perte des points 3D), nous sommes donc limités à trois vues (deux images et une projection) pour l'algorithme d'auto-calibrage.

C. Méthode proposée

La méthode complète que nous proposons pour obtenir une reconstruction 3D avec l'information de texture est décrite par la Fig. 3. Nous voulons, à partir de l'estimation du tenseur trifocal effectuée hors-ligne (pour contourner la contrainte 1), extraire les matrices de projection des deux caméras et du projecteur. En posant des hypothèses, détaillées plus loin, permettant de réduire le nombre d'inconnues du système, nous souhaitons déterminer à partir de celles-ci les paramètres intrinsèques des caméras et du projecteur ainsi que l'équation du plan à l'infini pour reconstruire métriquement la scène. La fonction de transfert de points du tenseur permet, à l'aide des correspondances entre l'image capturée par la caméra 1 et le

motif, de calculer les coordonnées du correspondant dans l'image capturée par la caméra 2 et ainsi d'en déduire la couleur/texture du point 3D observé (en ligne, et en évitant la contrainte 1).

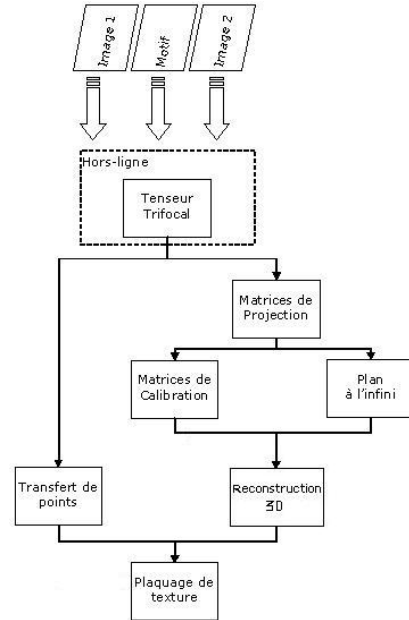


Fig. 3 Méthode pour la reconstruction 3D

III. LE TENSEUR TRIFOCAL

A. Principe et méthodes de calcul

La géométrie épipolaire (Fig. 4) traduit les relations de correspondance entre les projections sur plusieurs plans d'un point de l'espace 3D.

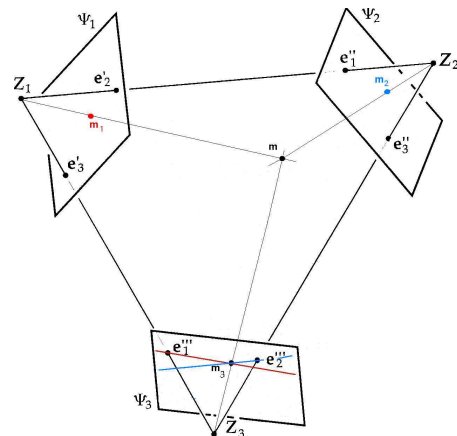


Fig.4 Géométrie épipolaire pour trois vues.

Le tenseur trifocal est une matrice $3 \times 3 \times 3$ qui décrit la géométrie épipolaire pour trois vues comme la matrice fondamentale le fait pour deux vues. Il se compose de 27 éléments mais ne comporte que 18 degrés de liberté car il doit satisfaire aux contraintes épipolaires, les relations de trilinearités, qui permettent de relier par correspondance les points et les lignes entre eux : 3 points de correspondance, 3 lignes, 2 points et une ligne, etc. La Fig. 4 montre qu'à partir de 2 points correspondants sur deux plans, on peut tracer les deux droites épipolaires correspondant sur le troisième plan et ainsi déterminer le troisième point de correspondance : c'est la *fonction de transfert de points* que nous évoquons dans la Fig. 3.

La difficulté est donc de calculer un tenseur homogène, *i.e.* qui respecte ses 18 degrés de liberté. Notre travail a tout d'abord consisté à implémenter et évaluer trois méthodes que l'on trouve dans la littérature (Fig. 5) :

- la première méthode, la plus simple, est appelée *solution directe linéaire* [11]. Son calcul est très simple mais de sa simplicité découle son imperfection. Les éléments du tenseur doivent satisfaire 8 contraintes internes, pour représenter un tenseur trifocal valide. Ces contraintes en général ne sont pas satisfaites par la solution linéaire directe. Un autre inconvénient est que la solution linéaire directe ne réduit pas les erreurs dans les mesures originales de point et de ligne, appelées erreurs de projection ou erreurs résiduelles ;
- la deuxième méthode est basée sur un calcul de moindre carré ajusté par un jeu de poids variable à chaque itération [3] ;
- la dernière méthode, la plus compliquée, utilise la méthode RANSAC : au lieu d'utiliser le maximum de données afin de réduire les erreurs de reprojection, on utilise au contraire le jeu minimum de points, six, pour calculer le tenseur, et on renouvelle le calcul du tenseur un nombre de fois conséquent, puis on détermine le meilleur tenseur à l'aide d'un critère, le MLE, Maximum likelihood Estimation ou estimation du maximum de voisinage [12].

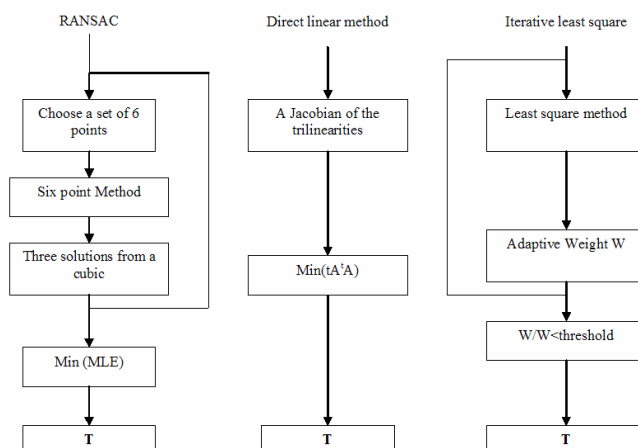


Fig. 5 Différentes méthodes d'estimation du tenseur trifocal.

B. Implémentation pratique

Il ne peut y avoir de points de correspondance entre les images du projecteur et de la caméra 2. Pour le calcul du tenseur trifocal et pour avoir une bonne précision sur la correspondance des points entre les trois plans, les deux caméras doivent percevoir le motif projeté (voir Fig. 6). Une fois la mise en correspondance réalisée, on peut estimer le tenseur trifocal à l'aide des méthodes décrites dans la section précédente.

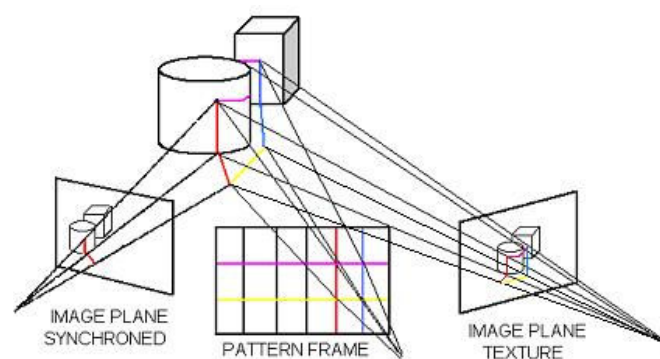


Fig. 6 Capteur en mode "hors ligne".

Un tenseur trifocal homogène est très pratique car il renferme presque toutes les informations utiles, entre autres les matrices de projection, et toute la géométrie épipolaire. On peut en extraire les matrices de projection, mais aussi les épipoles et les matrices fondamentales, même celles entre les images du projecteur et de la caméra 2, alors que l'on ne dispose d'aucun point de correspondance. L'autre avantage du tenseur trifocal est qu'il permet la correspondance entre points, alors que la matrice fondamentale se contente de correspondance entre points et lignes.

IV. RECONSTRUCTION METRIQUE

A. Modélisation des caméras et du projecteur

Le modèle de la caméra et du projecteur est décrit par la Fig. 7. On s'aperçoit rapidement qu'ils sont caractérisés par deux types de paramètres : les paramètres *intrinsèques*, qui ne dépendent que de la caméra/projecteur (point principal, distance focale), regroupés à l'intérieur d'une matrice dite de *calibration*, et ceux qui dépendent du monde extérieur (position relative des caméras, position par rapport au repère-monde), les paramètres *extrinsèques*.

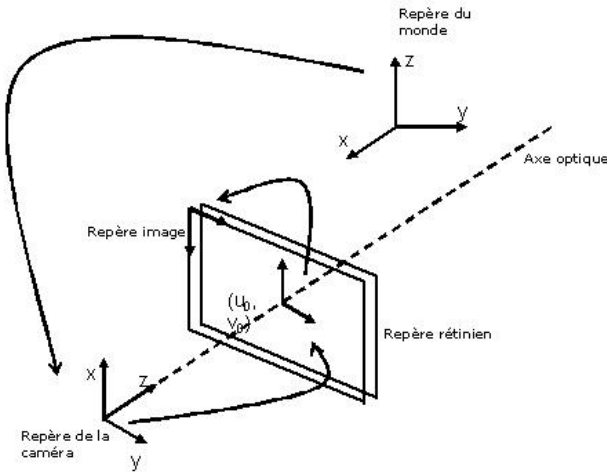


Fig. 7 Modèle de caméra/projecteur.

L'équation suivante décrit la projection perspective d'un point 3D sur le plan image :

$$\mathbf{m} \propto \mathbf{P}\mathbf{M}$$

où \propto représente une égalité à un facteur d'échelle près, ou égalité projective ; \mathbf{P} est une matrice 3×4 décrivant la projection ; $\mathbf{M} = [X \ Y \ Z \ 1]^T$ et $\mathbf{m} = [x \ y \ 1]^T$ sont respectivement les vecteurs de coordonnées homogènes du point 3D et de son image.

Dans un espace métrique, on a :

$$\mathbf{P}_M = \mathbf{K}[\mathbf{R} | -\mathbf{R}\mathbf{t}]$$

où \mathbf{R} et \mathbf{t} sont respectivement la matrice de rotation et le vecteur de translation du repère monde vers le repère caméra et \mathbf{K} est la matrice de calibration, de la forme :

$$\mathbf{K} = \begin{bmatrix} f_x & s & u_x \\ 0 & f_y & u_y \\ 0 & 0 & 1 \end{bmatrix}$$

f_x et f_y représentent les distances focales divisées par la taille des pixels en longueur et largeur ; (u_x, u_y) sont les

coordonnées du point principal et s est le facteur de *skew* (s est égale à zéro, les pixels sont orthogonaux).

Ce modèle décrit également la géométrie du projecteur. Celui-ci peut en effet être vu comme une caméra opérant « à l'envers » en inversant le sens de la ligne de vue (de la scène vers le plan image pour la caméra, du plan image vers la scène pour le projecteur).

B. Méthode pour l'auto-calibrage

La Fig. 8 détaille le principe général de la reconstruction 3D par stratification des géométries.

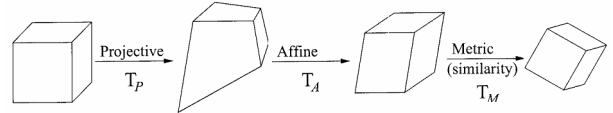


Fig. 8 Stratification de la reconstruction 3D.

La première étape [2], le calcul projectif, est assez simple : on peut soit extraire directement les matrices de projection du tenseur [3][11], soit les recalculer par un calcul proche de celui employé par la méthode itérative pour l'obtention du tenseur trifocal. Les deux autres étapes sont plus complexes, en particulier le calcul de la matrice \mathbf{T}_M , car il s'agit de déterminer le plan à l'infini, alors que \mathbf{T}_A revient à calculer les paramètres intrinsèques de la caméra. La connaissance de la matrice de calibration \mathbf{T}_A d'une caméra est nécessaire pour passer d'une reconstruction projective à affine. On peut choisir de dissocier les deux calculs (équations de Kruppa [4][7] pour \mathbf{T}_A , puis méthode de Hartley [5] pour le plan à l'infini) ou alors on peut utiliser l'image de la quadrique absolue ω^* [8][9] : la projection de la quadrique absolue sur les plans images code à la fois la conique absolue ω et le plan à l'infini Π_∞ (Fig. 9). Puisque la conique absolue est invariante aux translations et rotations, elle ne dépend donc que des paramètres intrinsèques de la caméra ; donc, calculer la conique absolue revient à déterminer la matrice de calibration.

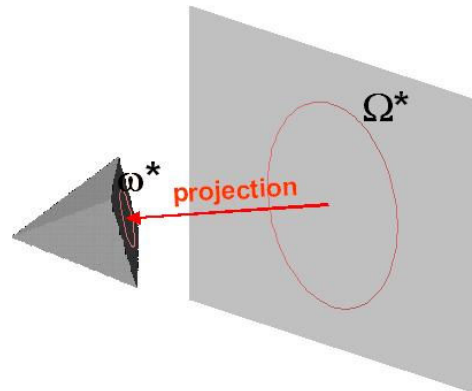


Fig. 9 Projection de la quadrique absolue.

Soit une expression de la quadrique absolue (le lecteur peut se référer à [9] pour plus de détails sur la suite de ce paragraphe) :

$$\Omega^* = \begin{bmatrix} \mathbf{K}\mathbf{K}^T & -\mathbf{K}\mathbf{K}^T a \\ -a^T \mathbf{K}\mathbf{K}^T & a^T \mathbf{K}\mathbf{K}^T a \end{bmatrix}$$

où \mathbf{K} représente la matrice de calibration et a les paramètres de l'équation du plan à l'infini ($\Pi_\infty = [a^T \ 1]^T$). La matrice de passage de l'espace projectif à l'espace métrique s'exprime alors sous la forme :

$$\mathbf{T}_{P \rightarrow M} = \begin{bmatrix} \mathbf{K}^{-1} & 0 \\ a^T & 1 \end{bmatrix}$$

Le fait que ω_i^* soit la projection de la quadrique absolue se traduit par l'équation :

$$\omega_i^* \propto \mathbf{P}_i \Omega^* \mathbf{P}_i^T \propto \mathbf{K}_i \mathbf{K}_i^T$$

où ω_i^* représente l'image de la quadrique absolue ; Ω^* , la quadrique absolue ; \mathbf{P}_i , les matrices de projection (dans un espace projectif) des caméras et du projecteur ; et les \mathbf{K}_i , les matrices de calibration, contenant les paramètres intrinsèques, des caméras et du projecteur.

Le critère suivant est à minimiser pour résoudre le problème en \mathbf{K}_i et en a :

$$\min \sum_{i=1}^n \left\| \frac{\mathbf{K}_i \mathbf{K}_i^T}{\|\mathbf{K}_i \mathbf{K}_i^T\|_F} - \frac{\mathbf{P}_i \Omega^* \mathbf{P}_i^T}{\|\mathbf{P}_i \Omega^* \mathbf{P}_i^T\|_F} \right\|_F^2$$

où l'opérateur $\|\cdot\|_F$ représente la norme de Frobenius.

Résoudre les équations directement se révèle difficile car nous disposons de plus d'inconnues que d'équations (le nombre de vues se limite à trois). Nous avons fait quelques hypothèses, couramment admises, pour simplifier l'expression de ω_i^* :

- la rétine est orthogonale (le facteur *skew* est nul) ;
- les pixels sont carrés : la distance focale est la même dans les deux directions de la lentille ($f_x = f_y$) ;
- le point principal est au centre de l'image (il faudra déplacer le centre de chaque image sur l'origine du plan image).

Après simplification, l'expression de l'image de la quadrique se réduit à :

$$\omega_i^* = \begin{bmatrix} f_i^2 & 0 & 0 \\ 0 & f_i^2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Trois vues sont donc suffisantes pour estimer les paramètres intrinsèques f_i des caméras et du projecteur et l'équation du plan à l'infini.

V. RESULTATS EXPERIMENTAUX

Nous avons testé chaque algorithme de la reconstruction pour trouver les étapes peu robustes de notre méthode de manière à pouvoir y remédier. Nous nous sommes, dans un premier temps, contenté d'effectuer ces tests sur des données simulées, plus facilement observables et commandables.

Pour éprouver la robustesse des algorithmes, nous appliquons sur les données un bruit gaussien de moyenne nulle et de variance σ .

A. Tenseur trifocal

Nous avons d'abord testé la justesse du tenseur, à savoir si les points sur les plans se trouvent bien à l'intersection de leurs épipolaires correspondants (voir Fig. 10).

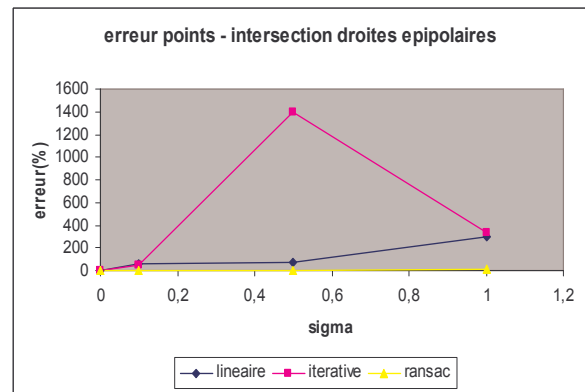


Fig. 10 Erreur de géométrie épipolaire.

Les méthodes linéaire directe et itérative dégénèrent rapidement, résultat attendu pour la première et moins pour la seconde. La méthode RANSAC résiste mieux au bruit, mais pas encore suffisamment. C'est pourquoi nous avons tenté d'améliorer cette méthode d'estimation : nous n'utilisons plus le calcul des matrices projectives à partir du tenseur trifocal que comme initialisation d'un algorithme récursif qui vise à diminuer les erreurs de reprojection. Les résultats sont maintenant très satisfaisants (Fig. 11).

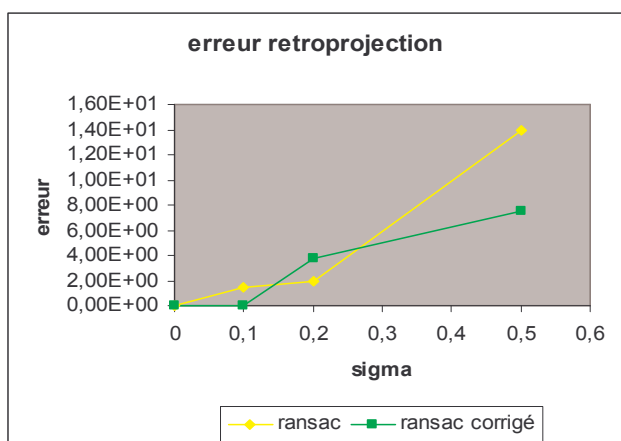


Fig. 11 : Comparaison de la robustesse de l'algorithme RANSAC et de sa version améliorée

B. Reconstruction projective

Nous avons ensuite testé les matrices de projection en calculant la reconstruction 3D projective puis en reprojétant ces points 3D pour voir s'ils correspondent bien avec les points projetés initialement (Fig. 12).

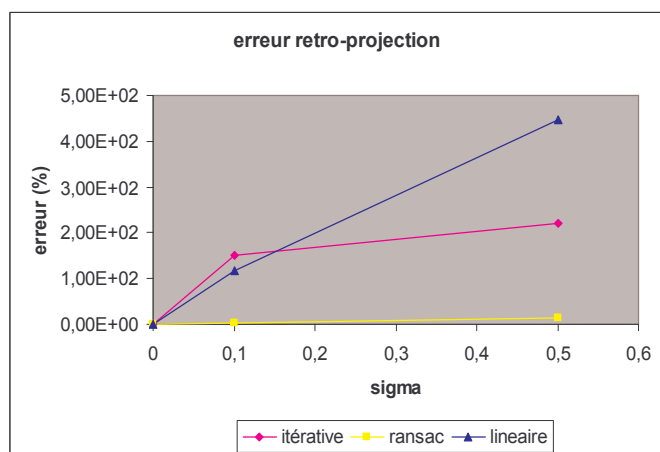


Fig. 12 Erreur de retro-projection.

La conclusion à la vue du graphique est la même que précédemment : les méthodes linéaires et itérative dégénèrent rapidement, alors que la méthode RANSAC résiste bien. La reconstruction projective obtenue est consistante et suffisamment précise pour alimenter l'étape de redressement métrique qui lui fait suite.

C. Reconstruction métrique

Enfin nous avons comparé les objets initiaux avec leur reconstruction 3D sans bruit.

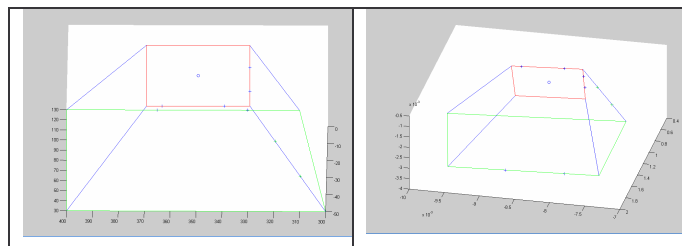


Fig. 13 Reconstruction 3D d'une pyramide.

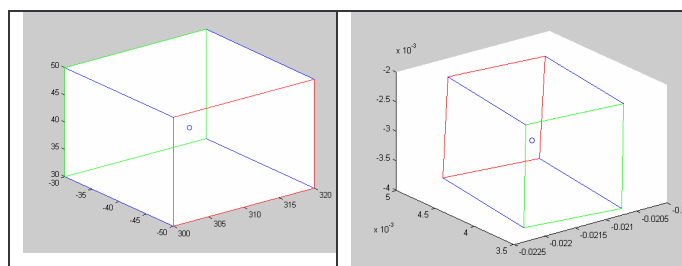


Fig. 14 Reconstruction 3D d'un cube.

On constate que les résultats sont bons : on reconnaît la forme du cube et de la pyramide et l'erreur mesurée sur les rapports de distance n'est que de 10%. Les approximations faites sur la matrice de calibration (pixels carrés et skew nul) expliquent en grande partie ce biais dans la reconstruction des points 3D.

VI. CONCLUSIONS

La reconstruction 3D à partir de trois vues se décompose en deux grandes étapes : le calcul du tenseur trifocal et l'auto-calibrage des caméras et du projecteur. Notre travail a d'abord consisté à implémenter trois des principales méthodes d'estimation du tenseur trifocal et d'en évaluer la robustesse. Il ressort que seul un algorithme robuste au bruit et capable de rejeter les points aberrants permet d'obtenir des résultats suffisamment satisfaisant pour alimenter l'étape d'auto-calibrage (même s'il faut encore améliorer la robustesse des algorithmes notamment en mettant des conditions sur la forme des matrices et du tenseur, comme par exemple la norme de Frobenius à 1).

Nous avons montré qu'une méthode d'auto-calibrage laissant la distance focale libre était applicable à un capteur de vision en lumière structurée imperceptible. Nous avons obtenu des résultats encourageants pour la reconstruction 3D à partir d'images de synthèse non bruitées, ce qui nous a permis de valider la méthode. Reste maintenant à intégrer tous ces algorithmes dans une séquence de traitements automatisés, à les éprouver sur des images réelles et à y adjoindre la fonction de transfert des points dans le but d'utiliser les informations de la seconde caméra pour plaquer les textures sur les surfaces reconstruites. La segmentation de l'image couleur et de l'image en lumière structurée, ainsi que la fusion des informations qu'elles contiennent, seront des points à étudier pour parvenir à l'automatisation totale du capteur.

RÉFÉRENCES

- [1] J. Battle, E. Mouaddib, J. Salvi, “Recent progress in coded structured light as a technique to solve the correspondence problem: a survey”, *Pattern Recognition*, 31(7), pp. 963-982, 1998.
- [2] O. Faugeras, “Stratification of 3-D vision: projective, affine, and metric representations”, *Journal of the Optical Society of America A*, 12(3), pp. 465-484, March 1995.
- [3] R. Hartley, “Lines and points in three views - an integrated approach”, *Proceedings of the ARPA IU Workshop*. DARPA, Morgan Kaufmann, 1994.
- [4] R. Hartley, E. Hayman, L. de Agapito, I. Reid, “Title camera calibration and the search for infinity”, *Proceedings of the International Conference on Computer Vision*, Kerkyra, Greece, 1999.
- [5] R. I. Hartley, E. Hayman, L. de Agapito, and I. D. Reid, “Camera calibration and the search for infinity”, *Proceedings of the International Conference on Computer Vision*, Kerkyra, Greece, 1999.
- [6] M. Livingston, “Vision-Based Tracking with Dynamic Structured Light for Video See-through Augmented Reality”, *PhD Thesis*, University of North-Carolina, 1998.
- [7] M. Lourakis, R. Deriche, “Camera self-calibration using the Kruppa equations and the SVD of the fundamental matrix: the case of varying intrinsic parameters”, *Rapport de Recherche INRIA*, N° 3911, Mars 2000.
- [8] M. Pollefeys, R. Koch, L. Van Gool, “Self-Calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters”, *International Journal of Computer Vision*, 32(1), pp. 7-25, 1999.
- [9] M. Pollefeys, “Self-calibration and metric 3D reconstruction from uncalibrated image sequences”, *PhD Thesis*, May 1999
- [10] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, H. Fuchs, “The Office of the Future: A Unified Approach to Image-Based Modeling and Spatially Immersive Displays”, *Proceedings of SIGGRAPH*, Orlando, Florida, 1998.
- [11] C. Ressel, “Geometry, constraints and computation of the trifocal tensor”, *PhD Thesis*, June 2003.
- [12] P. Torr, A. Zisserman, “Robust parameterization and computation of the trifocal tensor”, *Image and Vision Computing*, 15, pp. 591–605, 1997.
- [13] A. B. Watson, “Temporal Sensitivity”, *Handbook of Perception and Human Performance*, K. Boff, L. Kaufman and J. Thomas editors, New-York, 1986.