Singularités au service du mouvement et de la calibration.*

Diane Lingrand et Thierry Viéville
INRIA - Projet RobotVis
B.P.93
06902 Sophia Antipolis Cedex
{dlingran, vthierry}@sophia.inria.fr

Résumé Cet article s'insère dans un travail relatif aux singularités des scènes, des modèles de caméra et des mouvements au sein de séquences monoculaires non calibrées, dans le but d'obtenir le plus d'informations possible sur l'auto-calibration, le mouvement et la scène (objets géométriques, relations, structure 3D). Cette étude est un prolongement de [VL96] comprenant plus de cas grâce à l'introduction de code généré automatiquement et prévoyant dans un futur proche, l'introduction de toute contrainte.

Mots-Clés : séquences monoculaires, mouvements singuliers, auto-calibration

1 Etat de l'art

L'analyse du mouvement dans le cas de séquences d'images monoculaires non-calibrées a été étudiée par de nombreux auteurs se plaçant aussi bien du point de vue des correspondances de points, lignes ou motifs plans, utilisant une représentation discrète ou continue du mouvement et ceci, entre deux ou plusieurs images [MF92, VFL96, Fau95, Har94a, Sha94]. Ces études sont motivées par le fait qu'*on ne peut raisonablement pas considérer un système*

*Un résumé étendu en hypertexte est disponible à l'adresse suivante : `http://www.inria.fr/robotvis/personnel/dlingran/Arges/Orasis97/orasis97.html` *actif comme calibré*. Cependant, dans le cas général, il n'est pas possible d'auto-calibrer une caméra lors d'un zoom ou lors de modifications des paramètres intrinsèques à la caméra [VFL96], alors que certaines des équations présentent des singularités [Har94a, ZM94].

Observant ces faits, certains auteurs ont réanalysé le problème, considérant des déplacements ou modèles de caméra spécifiques [VFL96, TZM95, EV96, Stu97b, Stu97a]. Plus spécifiquement, rotations pures [Har94b, Ste95], translations pures, [SN94, VGMPVD94], variation de focale [PVGM95, EZV96], mouvements plans [ZF94, BZM94], rotations dont l'axe est orthogonal au plan de translation [BRZM95, Vié94, VFC95], modèle simplifié de caméra [VL96, LV96, CH94], ou, dans le cas planaire, étudié de façon très appronfondie par de nombreux auteurs, comme rapporté dans [VZR95] dans le cas non-calibré.

Par exemple, des rotations dont l'axe est orthogonal au plan de translation et l'angle connu ou des rotations pures [Vié94] permettent d'estimer les paramètres de calibration avec leur incertitude. Par la suite, on peut déterminer le type de mouvement optimal pour notre problème d'estimation.

Ces informations, collectées dans un cadre statistique comme dans [TZM95, VZR95, ZDFL94], nous permettent de déterminer quel type de déplacement ou de modèle de caméra va nous fournir le plus d'informations (couramment représentées par l'inverse de la matrice de covariance) sur la géométrie de la scène, le mouvement des objets et les paramètres de calibration.

2 Les équations du problème

2.1 Modèle de caméra et repère de référence

Nous utilisons le modèle standard "pinhole" pour notre caméra. Concernant la projection effectuée par la caméra, le modèle couramment employé et expérimentalement le plus adéquat est le modèle de projection perspective. Cependant, d'autres modèles tels que le modèle de projection paraperspective ou affine conviennent dans des cas précis et simplifient notoirement les équations [Fau93, LV96, CH94].

Les coordonnées sont exprimées par rapport à un repère de référence $\mathcal{R} = (C, \mathbf{x}, \mathbf{y}, \mathbf{z})$ *attaché à la rétine*, $\mathbf{z}$ étant aligné avec l'axe optique, $\mathbf{x}$ et $\mathbf{y}$ étant alignés avec les axes horizontaux et verticaux de l'image. Le plan rétinien est donc perpendiculaire à l'axe optique Cz.

Nous proposons ici un modèle général de projection pouvant englober les modèles utilisés dans la littérature :

$$Z\,\mathbf{m} = \underbrace{\begin{pmatrix} \alpha_u & \gamma & \lambda\, u_0 & (1-\lambda)\, u_0 \\ 0 & \alpha_v & \lambda\, v_0 & (1-\lambda)\, v_0 \\ 0 & 0 & \mu & (1-\mu) \end{pmatrix}}_{\mathbf{P}} \mathbf{M} \tag{1}$$

Dans ce modèle, on distingue deux types de paramètres, **les paramètres modaux** (λ et μ) et **les paramètres intrinsèques** (u_0, v_0 et γ).

Les paramètres modaux $\lambda \in \{0,1\}$ et $\mu \in \{0,1\}$ déterminent le modèle de projection :

	λ	μ
Projection perspective pure	1	1
Projection faiblement perspective (i.e. affine ou orthographique)	0	0
Projection para-perspective	1	0

2.2 Utilisation de points comme primitives

On représente un point 3D M par le vecteur des coordonnées homogènes $\mathbf{M} = (X, Y, Z, \mathcal{U})^T$, où $\mathcal{U} = 1$ pour les points euclidiens et $\mathcal{U} = 0$ pour les points de fuite ou points à l'infini, i.e. direction de droite.

On représente un point 2D de la rétine, de coordonnées (u, v), par le vecteur 3D des coordonnées homogènes: $\lambda\ \mathbf{m} = \lambda\ \vec{Cm} = \lambda\ (u, v, 1)^T$, correspondant aux droites d'une direction donnée passant par le centre optique (espace projectif 2D).

2.3 Représentation de déplacements rigides

Nous nous intéressons au mouvement d'objets rigides dans la scène et au mouvement propre de la caméra. Cela signifie que les éléments de la scène vont effectuer un mouvement rigide paramétrisé par une matrice de rotation R, pouvant être exprimée par $\mathbf{I} + \left[\frac{\tilde{\mathbf{r}}+\frac{1}{2}\tilde{\mathbf{r}}^2}{1+\frac{\mathbf{r}^T\mathbf{r}}{4}}\right]$ où $\mathbf{r} = [r_0 r_1 r_2]^T$, l'axe de rotation étant $\frac{\mathbf{r}}{||\mathbf{r}||}$ et l'angle 2 $\arctan(\frac{||\mathbf{v}||}{2})$ (pour plus de détails, voir [LV96]), et un vecteur de translation $\mathbf{t} = [t_0 t_1 t_2]^T$, ce qui donne pour un point 3D M :

$$\mathbf{M}' = \mathbf{R}\,\mathbf{M} + \mathbf{t} \tag{2}$$

2.4 Equations du cas général

Considérant deux images consécutives :

$$\exists(\lambda,\lambda',X,Y,Z,U)/\left\{\begin{array}{lcl}\lambda\,\mathbf{m} & = & \mathbf{P}\,\mathbf{M}\\ \lambda'\,\mathbf{m}' & = & \mathbf{P}'\,\mathbf{M}\end{array}\right. \tag{3}$$

Ce qui peut également s'écrire :

$$\exists(\lambda,\lambda',X,Y,Z,U)/\left|\begin{array}{ccc}\mathbf{m} & \begin{pmatrix}0\\0\\0\end{pmatrix} & \mathbf{P}\\ \begin{pmatrix}0\\0\\0\end{pmatrix} & \mathbf{m}' & \mathbf{P}'\end{array}\right| = 0$$

On peut remarquer qu'il s'agit d'une forme bilinéaire en $\mathbf{m}$ et $\mathbf{m}'$, ce qui montre que ce résultat peut se mettre sous une forme plus habituelle, utilisant une généralisation de la matrice fundamentale :

$$\mathbf{m}'^T\,\mathbf{F}(P,P')\,\mathbf{m} = 0$$

3 Les cas particuliers

Nous ne traitons dans le cadre de ce papier que les cas particuliers relevant de la modélisation de la caméra et du mouvement de la caméra et de la scène. Les cas relevant des particularités de la scène ne sont pas étudiés dans cet article; on supposera les points dans une position générale (non-coplanaires) ainsi que le mouvement autre qu'une rotation pure, ceci afin d'utiliser exclusivement la matrice $\mathbf{F}$.

3.1 Cas particuliers relevant de la modélisation de la caméra

Nous nous intéressons ici aux paramètres modaux et intrinsèques. Nous avons déjà vu les valeurs possibles pour les paramètre modaux, ce qui se résume à trois cas. Concernant les paramètres intrinsèques, plusieurs cas sont à considérer :

γ, **mesurant la non-orthogonalité du plan CCD à l'axe optique,** peut raisonnablement être négligé dans la plupart des cas ou au pire être considéré constant

α_u et α_v, les tailles horizontale et verticale des pixels conduisent à plusieurs cas possibles :

- α_u et α_v sont constantes
- α_u et α_v sont constantes et connues
- le rapport $\frac{\alpha_u}{\alpha_v}$ est constant et connu

u_0 et v_0, les coordonnées du centre optique peuvent être considérées comme :

- constantes mais inconnues
- constantes mais connues (par exemple fixées au centre de l'image)

3.2 Cas particuliers relevant du mouvement de la caméra (ou de la scène)

Mouvement continu - mouvement discret Dans une séquence d'images, si le mouvement est faible par rapport à la cadence des images, on peut approximer le mouvement à son premier ordre, ce qui concerne uniquement la rotation : $\mathbf{R} = \mathbf{I} + \tilde{\mathbf{r}}$

Autres cas matrice de rotation R : nullité d'une ou de plusieurs composantes de $\mathbf{r}$

vecteur de translation t : nullité d'une ou deux composantes de $\mathbf{t}$

parallélisme : $\mathbf{r} \perp \mathbf{t} \Leftrightarrow \mathbf{r}.\mathbf{t} = 0$

orthogonalité : $\mathbf{r} \parallel \mathbf{t} \Leftrightarrow \exists\lambda \ / \ \mathbf{r} = \lambda\mathbf{t}$

3.3 Contraintes : compatibilité - redondance - complétude

L'ensemble des contraintes est formé des contraintes dites atomiques car élémentaires (vues précédemment) et des contraintes dites moléculaires constituées par des "ET" logiques entre contraintes atomiques ceci selon certaines règles : on élimine les combinaisons contradictoires et les combinaisons redondantes en ne considérant que le résultat de la combinaison et on impose des conditions d'adéquation à une réalité physique (choix obligatoire d'un modèle de caméra, d'un mode de rotation et non-nullité de la translation lors de l'estimation de $\mathbf{F}$)

D'autres part, les contraintes ne sont pas forcément suffisantes pour résoudre le problème. Dans ce cas, on peut, soit se limiter aux éléments déterminables, soit ajouter des contraintes "par défaut" afin d'effectuer la résolution complète ce que l'on a choisi dans cet article.

3.4 Utilisation des particularités pour :

- déterminer le mode de projection de la caméra

- déterminer les paramètres de calibration (ceux qui sont constants)

- déterminer le type de mouvement et si possible ses caractéristiques (direction de translation, axe de rotation, angle de rotation)

- éventuellement, segmenter les points en ensembles de points rigidement liés

4 Implémentation

En entrée, nous avons deux images et une liste de contraintes à partir desquelles nous allons déterminer, en sortie, la cohérence ou non des contraintes (la non-cohérence arrête l'algorithme à ce stade), un coefficient d'adéquation des données aux contraintes et les valeurs des paramètres du problème non connus en entrée.

Dans chaque image, nous extrayons des points d'intérêts correspondant aux points de plus grande courbure à l'aide du détecteur de coins de Harris [HS88] et recherchons les correspondances entre les points des deux images.

La saisie des contraintes s'effectue par l'intermédiaire d'une interface Java. La combinaison des contraintes et la vérification de compatibilité s'effectue ensuite par le logiciel de calcul formel MAPLE[1] qui, à partir de ces contraintes et à l'aide la librairie Mascotte[2], génére le code C des routines de minimisation portant ainsi uniquement sur les paramètres libres et du calcul d'erreur par l'intermédiaire des matrices fondamentales ou homographies. Ce code est ensuite compilé avec le code pré-existant.

1. `http://www.maplesoft.com/`
2. `http://www.inria.fr/robotvis/personnel/vthierry/mascotte/main.html`

4.1 Corps de l'algorithme

- Choix des contraintes (par un utilisateur, voir automatiquement, de façon exhaustive).

- Test sur la compatibilité des contraintes. Si oui, continuer, si non, revenir au choix des contraintes.

- Détermination des paramètres libres et liés - Mise en équations réduites.

- Minimisation à l'aide des données du critére d'adéquation en faisant varier les paramètres libres.

4.2 Minimisation - Critère utilisé

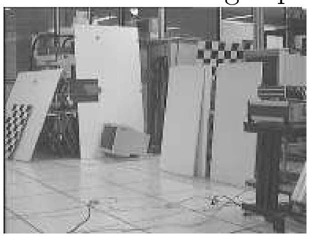
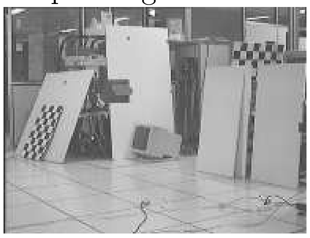
N'ayant pas encore résolu le problème de la détermination automatique d'un estimateur des paramètres, nous utilisons l'algorithme "Nonlinear Least-Squares" qui, d'autre part, fournit de bons résultats.

Comme expliqué par Zhang dans [ZDLF94], un critère efficient est la moyenne des distances euclidiennes entre chaque point et sa droite épipolaire correspondante, pondérée par un poids correspondant à la précision de l'appariement, en réalité, l'inverse de la variance de la précision de l'appariement.

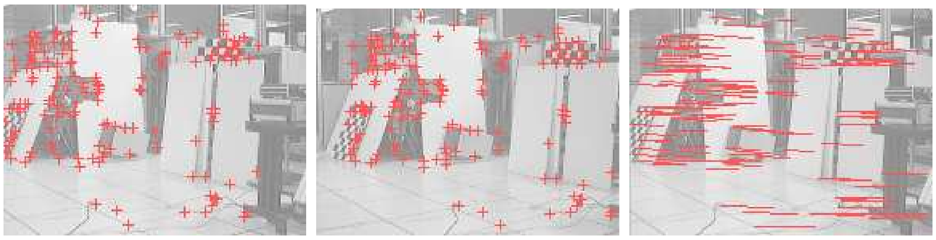
5 Résultats

Voici les résultats correspondants à un mouvement réel de translation pure sur le rail effectué par Argès. On compare les résidus obtenus pour plusieurs séries d'hypothèses sur le mouvement.

Voici les images prises par Argès :

et les 154 points détectés sur ces images :



cas particulier	résidu
rotation nulle, zoom fixe, γ nul	1.00368
translation en x pure	0.00485771
translation en z pure	1.00368
rotation en x, translation en z, γ nul	987.827
zoom et translation purs	0.955463

Ces résultats, ainsi que ceux reportés dans la version hypertexte[3] de ce papier, confirment 2 points:

- *le meilleur résultat correspond au mouvement effectué :* cette méthode permet donc de retrouver à quel type de mouvement on a à faire

- *le cas de translation pure fournit un résultat moins bon que translation en x pure :* moins on a de paramètres, plus la méthode est précise

6 Application

L'intérêt immédiat de ce travail est de permettre de recueillir plus d'informations sur la caméra, la scène et les déplacements. On peut ensuite envisager de se placer volontairement et successivement dans des cas particuliers pour obtenir un maximum d'informations et une reconstruction "hiérarchique": Le robot effectue un déplacement particulier pour déterminer tel ou tel paramètre qu'il utilisera ensuite lors d'un autre déplacement pour reconstruire la scène, ... les critères permettant de déterminer de façon automatique l'adéquation du mouvement aux cas particuliers.

3. `http://www.inria.fr/robotvis/personnel/dlingran/Arges/Orasis97/orasis97.html`

Bibliographie

Références

[BRZM95] Paul A. Beardsley, Ian D. Reid, Andrew Zisserman, and David W. Murray. Active visual navigation using non-metric structure. In *Proceedings of the 5th International Conference on Computer Vision* [icc95], pages 58–64.

[BZM94] Paul Beardsley, Andrew Zisserman, and David Murray. Navigation using affine structure from motion. In Eklundh [Ekl94], pages 85–96.

[CH94] Stéphane Christy and Radu Horaud. Euclidian shape and motion from multiple perspective views by affine iterations. Technical Report 2421, INRIA Rhones-Alpes, December 1994.

[Ekl94] J-O. Eklundh, editor. volume 800-801 of *Lecture Notes in Computer Science*, Stockholm, Sweden, May 1994. Springer-Verlag.

[EV96] Reyes Enciso and Thierry Viéville. Self-calibration from four views with possibly varying intrinsic parameters. *Image and Vision Computing Journal*, 1996.

[EZV96] R. Enciso, A. Zisserman, and T. Viéville. An affine solution to the euclidean calibration while using a zoom lens. In *Workshop ALCATECH, 21-27, Danemark*, July 1996.

[Fau93] Olivier Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint.* MIT Press, 1993.

[Fau95] Olivier Faugeras. Stratification of 3-D vision: projective, affine, and metric representations. *Journal of the Optical Society of America A*, 12(3):465–484, March 1995.

[Har94a] Richard Hartley. Projective reconstruction and invariants from multiple images. *PAMI*, 16(10):1036–1040, 1994.

[Har94b] Richard Hartley. Self-calibration from multiple views with a rotating camera. In Eklundh [Ekl94], pages 471–478.

[HS88] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. 4th Alvey Vision Conf.*, pages 189–192, 1988.

[icc95] Boston, MA, June 1995. IEEE Computer Society Press.

[icp94] Jerusalem, Israel, October 1994. Computer Society Press.

[LV96] Diane Lingrand and Thierry Viéville. Dynamic foveal 3D sensing using affine models. In *13th ICPR*, volume 1, pages 810–814, August 1996.

[MF92] S. J. Maybank and O. D. Faugeras. A theory of self-calibration of a moving camera. *The International Journal of Computer Vision*, 8(2):123–152, August 1992.

[PVGM95] M. Pollefeys, L. Van Gool, and T. Moons. Euclidean 3d reconstruction from stereo sequences with variable focal lenghts. In *Proceedings of the 2nd Asian Conference on Computer Vision*, volume II, page 6, Singapore, December 95.

[Sha94] Amnon Shashua. Projective structure from uncalibrated images: structure from motion and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):778–790, 1994.

[SN94] A. Shashua and N. Navab. Relative affine structure: Theory and application to 3D reconstruction from perspective views. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, Seattle, WA, June 1994. IEEE.

[Ste95] Gideon P. Stein. Accurate internal camera calibration using rotation with analysis of sources of error. In *Proceedings of the 5th International Conference on Computer Vision* [icc95].

[Stu97a] P. Sturm. Critical motion sequences and conjugacy of ambiguous Euclidean reconstructions. In *Proceedings of the 10th Scandinavian Conference on Image Analysis, Lappeenranta, Finland*, 1997. submitted.

[Stu97b] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, 1997. submitted.

[TZM95] Phil Torr, Andrew Zisserman, and Steven Maybank. Robust detection of degenerate configurations for the fundamental matrix. In *Proceedings of the 5th International Conference on Computer Vision* [icc95], pages 1037–1042.

[VFC95] T. Viéville, P.E.D.S. Facao, and E. Clergue. Computation of ego-motion using the vertical cue. *Machine Vision and Applications*, 8(1):41–52, 1995.

[VFL96] Thierry Viéville, Olivier D. Faugeras, and Quang-Tuan Luong. Motion of points and lines in the uncalibrated case. *The International Journal of Computer Vision*, 17(1):7–42, January 1996.

[VGMPVD94] L. Van Gool, T. Moons, M. Proesmans, and M. Van Diest. Affine reconstruction from perspective image pairs obtained by a translating camera. In *Proceedings of the International Conference on Pattern Recognition* [icp94].

[Vié94] T. Viéville. Autocalibration of visual sensor parameters on a robotic head. *Image and Vision Computing*, 12, 1994.

[VL96] Thierry Viéville and Diane Lingrand. Using singular displacements for uncalibrated monocular visual systems. In *4th ECCV*, volume 2, pages 207–216, April 1996.

[VZR95] Thierry Viéville, Cyril Zeller, and Luc Robert. Using collineations to compute motion and structure in an uncalibrated image sequence. *The International Journal of Computer Vision*, 18(2), 1995.

[ZDFL94] Zhengyou Zhang, Rachid Deriche, Olivier Faugeras, and Quang-Tuan Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78(1-2):87–119, 1994. Appeared in October 1995, also INRIA Research Report No.2273, May 1994.

[ZDLF94] Z. Zhang, R. Deriche, Q.-T. Luong, and O. Faugeras. A robust approach to image matching: Recovery of the epipolar geometry. In *Proc. International Symposium of Young Investigators on Information\Computer\Control*, pages 7–28, Beijing, China, February 1994.

[ZF94] Cyril Zeller and Olivier Faugeras. Applications of non-metric vision to some visual guided tasks. In *Proceedings of the International Conference on Pattern Recognition* [icp94], pages 132–136. A longer version in INRIA Tech Report RR2308.

[ZM94] Andrew Zisserman and Steven Maybank. A Case Against Epipolar Geometry. In Joseph L. Mundy, Andrew Zissermann, and David Forsyth, editors, *Applications of Invariance in Computer Vision*, volume 825 of *Lecture Notes in Computer Science*. Springer-Verlag, 1994.