Auto-calibrage de caméras et reconstruction 3-D à partir d'images

Peter Sturm

INRIA Rhône-Alpes
655 Avenue de l'Europe, 38330 Montbonnot St Martin
Peter.Sturm@inrialpes.fr
http://www.inrialpes.fr/movi/people/Sturm

Résumé. Nous passons en revue différents scénarios de reconstruction 3-D, qui se distinguent selon le nombre d'images utilisées, si les caméras sont calibrées ou non et si la scène observée est connue ou non.

1 Introduction

Dans ce manuscript, nous essayons de regrouper quelques résultats marquants de ce que l'on pourrait appeler la "vision 3-D par ordinateur". C'est une branche commune entre les domaines très liés de la photogrammétrie et de la vision par ordinateur ; elle étudie comment obtenir des informations tri-dimensionnelles, à partir d'image bi-dimensionnelles. Le plus souvent, le but recherché est l'obtention de mesures sur des objets, voire la création de modèles complets, par exemple pour la visualisation photoréaliste. D'autres applications vont s'intéresser à localiser la caméra (ou par exemple le robot qui porte la caméra) dans un environnement ou de mesurer un déplacement effectué.

La vision 3-D propose toute une gamme de solutions pour de tels problèmes d'estimation. Nous restreignons ce manuscript à la considération exclusive des parties concernant l'estimation d'entités géométriques, tout en soulignant que par exemple la création de modèles photoréalistes est bien plus complexe…

Les principaux objets de recherche du domaine sont le développement d'algorithmes numériques et la preuve théorique des conditions minimales pour l'existence de solutions. Ces recherches sont d'un côté motivées par des problèmes pratiques concrets, de l'autre côté souvent par la volonté de résoudre des problèmes en utilisant de moins en moins d'informations.

Dans ce manuscript, nous ne voulons donner qu'un très bref aperçu de quelques problèmes de base qui ont été résolus, sans prétendre à un traitement exhaustif ni donner une bibliographie complète. Pour une étude plus approfondie, nous renvoyons le lecteur aux ouvrages récents [4, 2].

Dans la section 2, nous donnons une définition générale du problème de la vision 3-D. La section 3 contient une description de quelques problèmes résolus.

2 Description du problème général

Nous considérons un ensemble d'images d'une scène (dans ce manuscript, nous supposons que la scène est statique). Notre but principal est la détermination d'informations tri-dimensionnelles (e.g. distances,

positions), à partir des images, qui, elles, sont bi-dimensionnelles. Les images dépendent de la structure de la scène, mais aussi des caractéristiques des caméras qui les ont acquises [1]. Une définition possible du problème général de la vision 3-D est donc la modélisation de la scène *et* des caméras, c'est-à-dire l'estimation des paramètres définissant leur géométrie.

La principale source d'informations pour résoudre ce problème est l'ensemble des images ; plus particulièrement, c'est surtout la *mise en correspondance* des images, c'est-à-dire l'identification des mêmes objets ou primitives géométriques dans différentes images, qui fournit des indices sur la structure de la scène et le positionnement des caméras. Les algorithmes que nous allons présenter dans la suite, utilisent des *points image*, mis en correspondance au préalable, par une méthode quelconque (automatiquement ou manuellement). Les points sont les primitives les plus utilisées, grâce à leur manipulation algébrique aisée et au fait que la plupart des images réelles permettent d'extraire beaucoup de "points d'intérêt" (e.g. points de contour). D'autres primitives, telles des droites ou des courbes, sont également utilisées par beaucoup d'approches, mais leur traitement dépasserait le cadre de ce manuscript.

La structure de la scène est donc modélisée ici par un ensemble de points 3-D. Quant à la géométrie d'une caméra, nous distinguons sa *géométrie externe* – sa position et orientation – de sa *géométrie interne* – l'ensemble de ses *paramètres internes*, décrivant ses propriétés optiques et autres, tels la distance focale ou la taille des pixels [2]. Une caméra dont la géométrie interne est connue, sera appelée *calibrée*. Voici donc les ingrédients de notre problème général : structure de la scène, géométrie interne des caméras et leur géométrie externe (c'est-à-dire leur positionnement relatif). Selon les connaissances a priori sur ces entités, différents problèmes de modélisation se déclinent. Quelques-uns des plus importants sont traités dans la section suivante.

3 Aperçu de quelques problèmes de modélisation 3-D

Nous décrivons quelques-uns des problèmes de base. Tous ces problèmes sont résolus, mais leur mise en œuvre demande parfois une certaine expertise. Pour la plupart des problèmes, nous esquissons quelques conditions minimales pour leur solution.

3.1 Calibrage

Il est bien connu que l'on peut calibrer une caméra à partir d'une ou plusieurs images d'un *objet de calibrage* dont la structure est *connue* (e.g. connaissance des coordonnées des points d'intérêt dans un repère quelconque), voir la figure 1 pour des exemples. Le calibrage permet d'estimer la géométrie interne et celle externe de la caméra, quoique souvent, seule la géométrie interne est recherchée (e.g. pour une caméra qui se déplace au cours d'une application ultérieure). Ici, la géométrie externe définit la position et l'orientation de la caméra *par rapport à l'objet de calibrage*. Par conséquent, un système de plusieurs caméras statiques peut être calibré complètement en plaçant un objet connu dans le champs de vue commun des caméras : le placement relatif d'une caméra par rapport aux autres peut être déterminé en combinant les informations sur le positionnement relatif des caméras par rapport à l'objet.

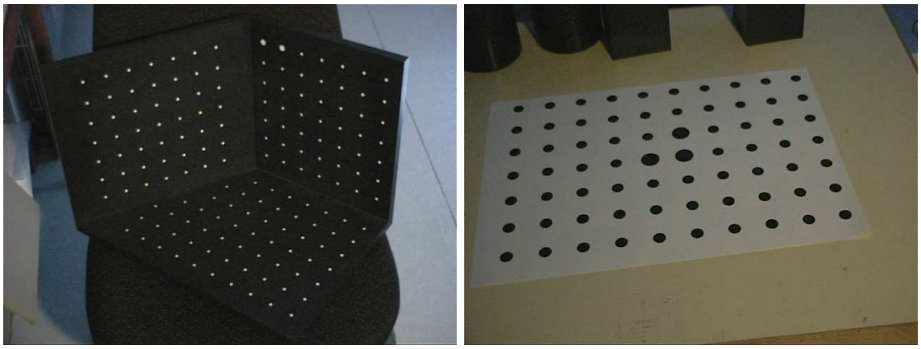
Un calibrage complet à partir d'une seule image requiert que l'objet de calibrage soit *tri-dimensionnel*.

1. Des images réelles dépendent bien sûr de beaucoup d'autres facteurs, comme l'illumination et la réflectance de surfaces, mais ici, nous nous concentrons sur les aspects purement géométriques de la formation d'images.

2. Dans les figures suivantes, le modèle de caméra sténopé (projection perspective) est illustré ; beaucoup d'autres modèles existent : des modèles plus simples, tels diverses classes de projections parallèles, et des modèles plus complexes modélisant des déviations de la projection réelle du modèle perspectif, voir e.g. [6].

Pourtant, la connaissance partielle de la géométrie interne (e.g. données de constructeur) permet parfois de faire pareil avec une image d'un objet *plan*. Même sans aucune connaissance a priori sur la géométrie interne, un calibrage complet est possible à partir de plusieurs images d'un objet plan, prises de différents points de vue [8, 12].

Notons qu'il existe bien d'autres moyens de (partiellement) calibrer une caméra, par exemple en prenant des images de sphères.



(a) (b)

FIG. 1: *(a) Un objet de calibrage spécialement manufacturé. (b) Un objet de calibrage plan très simple (produit à l'aide d'une imprimante laser).*

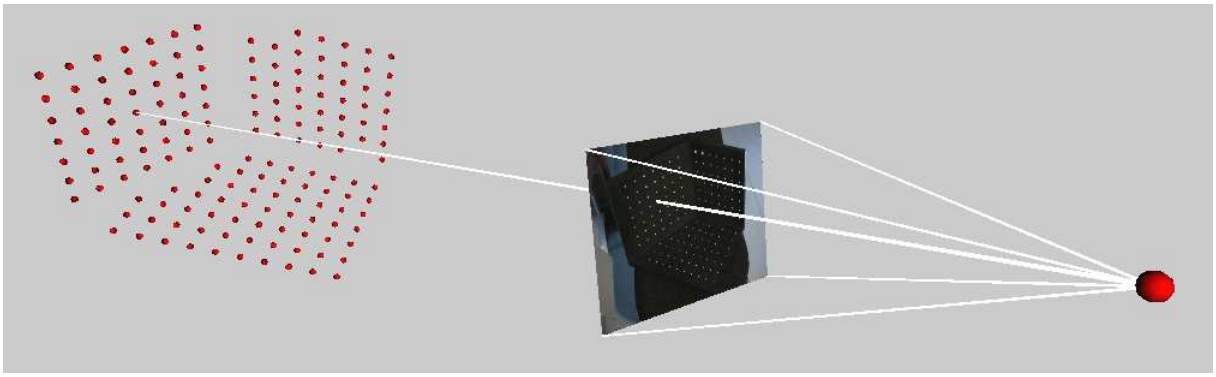


FIG. 2: *La position et l'orientation de la caméra (relatives au modèle de l'objet de calibrage) ainsi que sa géométrie interne ont été retrouvées par calibrage. La figure montre un rayon de projection associé à une des cibles de l'objet de calibrage.*

Conditions minimales. Une seule image d'un objet tridimensionnel suffit ; plus précisément, les projections de 6 points en position générale (non coplanaires etc.) sont requises au minimum. L'utilisation d'un objet plan, requiert typiquement deux ou plusieurs images pour un calibrage complet. Il faut savoir qu'il existe quelques classes de prises de vues singulières. Par exemple, une condition nécessaire est de prendre des images de différentes positions, mais aussi avec des orientations différentes de la caméra (il faut la tourner entre des prises de vue). Une description plus complète des singularités est donnée dans [8].

3.2 Calcul de pose

Il s'agit d'un sous-problème du calibrage, et concerne une caméra dont la géométrie interne est connue (par exemple par calibrage). Le but est, à partir d'une image d'un objet connu, d'estimer le positionnement relatif entre la caméra et cet objet (souvent appelé la *pose*). La résolution de ce problème est surtout intéressante pour le cas d'objets plans : bien qu'une seule image ne suffit pas pour effectuer un calibrage complet, la pose, elle, peut être déterminée ! Le calcul de pose, peut donc être très utile dans toute application où il faut localiser la caméra par rapport à la scène, par exemple en réalité augmentée.

Conditions minimales. Au moins 3 points de l'objet connu doivent être visibles. Dans ce cas, deux solutions plausibles peuvent être obtenues. Avec plus de points (même s'ils sont coplanaires avec les 3 premiers), il n'en reste plus qu'une seule solution plausible.

3.3 Triangulation

Un autre problème classique est plus ou moins la réciproque du calibrage. Etant donné plusieurs images d'une scène, prises par des caméras dont la géométrie complète (interne et externe) est connue, il s'agit d'estimer la structure de la scène. Le principe de la triangulation (à ne pas confondre avec la triangulation de surfaces) est illustré dans la figure 3.

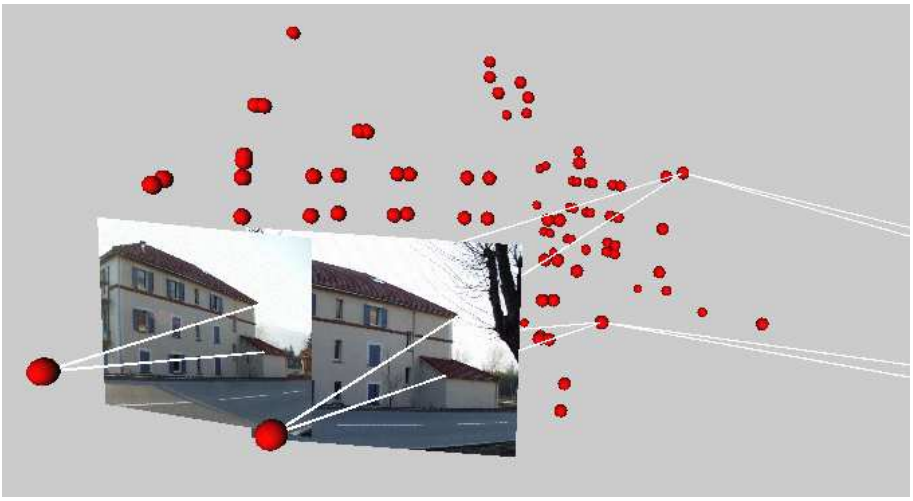


FIG. 3: *Triangulation de deux points d'une scène, à partir de 4 images (seules 2 sont montrées).*

Conditions minimales. Les seules conditions nécessaires pour la triangulation sont des centres optiques distincts et que le point à trianguler ne se trouve pas sur la droite entre les deux centres optiques.

3.4 Estimation du mouvement

Un autre problème classique concerne une caméra en mouvement, donc la géométrie interne est connue (par exemple par un calibrage préalable). Dans ce cas, le mouvement de la caméra peut être déterminé à partir d'images, même si la scène observée est *inconnue*. Ensuite, nous nous trouvons dans le cas de figure du paragraphe précédent, donc l'estimation de la structure de la scène est possible.

Pour être précis, le mouvement ne peut être estimé qu'à une échelle globale près, c'est-à-dire que seules des distances relatives, et non absolues peuvent être calculées. L'estimation du mouvement peut être formulée comme un problème géométrique comme illustré dans les figures 4 à 6 : la figure 4 montre deux images, prises au cours d'un déplacement d'une caméra. Nous supposons que la géométrie *interne* de la caméra est connue. Ceci permet en effet de calculer des lignes de vue pour chaque image (voir la figure 5). L'estimation du mouvement revient alors à positionner les deux images, l'une par rapport à l'autre, tel que les lignes de vue correspondant se coupent dans l'espace (voir la figure 6).

FIG. 4: *Estimation du mouvement : deux images prises au cours d'un déplacement.*

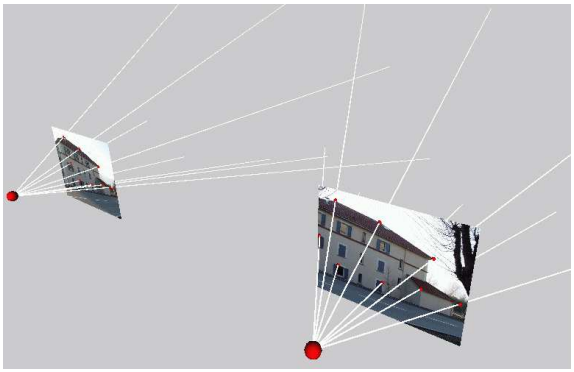


FIG. 5: *Estimation du mouvement : lignes de vue.*

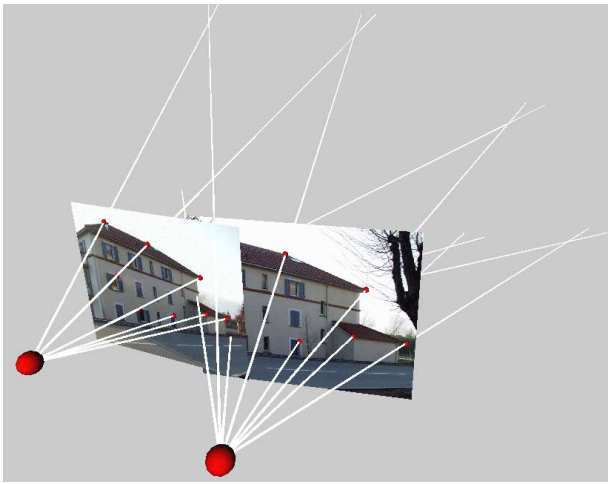


FIG. 6: *Estimation du mouvement : les lignes de vues se correspondants se coupent.*

Conditions minimales. La plupart des approches requièrent que la scène observée soit tri-dimensionnelle. Dans ce cas, le mouvement peut être estimé à partir de 5 points en position générale, mis en correspondance entre deux images (il existent 10 solutions, mais avec plus de 5 points il n'en reste plus qu'une seule).

Il existe des approches spécialisées à des scènes planes. Avec la connaissance de la planarité, 4 ou plus de points mis en correspondance, permettent d'estimer le mouvement, à 2 solutions plausibles près.

3.5 Modélisation non métrique

Ce paragraphe et le suivant, répondent à la question ce qui est possible de faire avec des caméras *non calibrées*, si la scène observée est *inconnue*. Au début des années 90, il a été montré que même dans ce cas-ci, la structure de la scène et la géométrie des caméras peuvent être déterminées à partir des seules images (et des correspondances entre elles) [1, 3]. Le principal inconvénient étant que celles-ci ne peuvent a priori être déterminées qu'à une transformation non-métrique près. Selon les conditions de prise de vue, on peut obtenir par exemple une *reconstruction affine* (si les caméras peuvent être modélisées par des projections parallèles) ou une *reconstruction projective* (pour des caméras perspectives). On pourra alors être sûr que l'estimation de la structure de la scène approche la réalité à une transformation inconnue près, du type donné, mais l'on ne saura pas dire plus. Ces types de reconstruction ne permettent pas de mesurer des distances ou des angles, mais il est possible de calculer des propriété invariantes au type de transformation concerné : une reconstruction affine, par exemple, permet de calculer des rapports de distances de 3 points collinéaires.

Malgré cet inconvénient, ce résultat n'est pas seulement d'un intérêt purement théorique. Tout d'abord, notons que d'un problème avec typiquement des centaines d'inconnues (l'ensemble des coordonnées des points 3-D et des paramètres des caméras), il n'en restent plus que quelques-uns – les paramètres de la transformation non-métrique. Nous verrons, dans le paragraphe suivant, qu'il y a beaucoup de moyens de combler ce manque et d'obtenir ainsi une reconstruction métrique.

Conditions minimales. Il faut disposer d'au moins 2 images, prises de points de vue différents. La mise en correspondance de 7 points dans 2 images, ou de 6 points dans 3 images, permet l'obtention d'une reconstruction projective de la scène et des caméras.

3.6 Auto-calibrage

Nous considérons le même scénario qu'au paragraphe précédent et supposons donc qu'une reconstruction non-métrique a pu être obtenue. Afin d'estimer la transformation non-métrique (projective ou affine) manquante, toute connaissance sur la structure de la scène ou la géométrie des caméras peut en principe être utilisée. Ce qui est probablement le moins contraignant pour des applications pratiques est d'introduire de faibles connaissances ou hypothèses sur la géométrie interne des caméras. Les informations utiles peuvent être :

- la connaissance de quelques paramètres internes, surtout de ceux qui ne varient en principe pas ou peu (e.g. la taille des pixels) ;
- la certitude que la géométrie d'une caméra, entre deux prises de vue, ne change pas ou seulement partiellement (e.g. seule la distance focale change, à cause de l'utilisation du zoom).

Selon le nombre d'images utilisées, le nombre de connaissances et d'autres facteurs, la transformation non-métrique peut être complètement ou partiellement estimée. L'estimation de cette transformation

implique celle de la géometrie interne des caméras, c'est pourquoi on parle de *calibrage en ligne* ou *auto-calibrage*.

Ayant effectué l'auto-calibrage, la structure *métrique* de la scène est connue (après application de la transformation estimée à la reconstruction non-métrique initiale), ainsi que la géométrie interne des caméras. L'estimation de leur géométrie externe (ou bien, l'estimation du mouvement) ne pose alors plus de problème.

D'autres contraintes utiles pour l'estimation de la transformation non-métrique manquante, sont :

- des connaissances partielles sur la structure de la scène (une connaissance complète permettrait un calibrage complet), telles la perpendicularité de droites ou de plans, le rapport de deux distances, etc. Ces connaissances sont très utiles pour des modélisations interactives puisqu'elles sont souvent faciles à définir par un utilisateur. De telles connaissances permettent parfois même d'obtenir la structure de la scène et la géométrie interne d'une caméra simultanément, à partir d'une seule image (voir le paragraphe suivant) !
- des connaissances sur la géométrie externe des caméras. Par exemple, il peut être connu qu'une caméra ne bouge que dans un plan (e.g. une caméra montée sur un véhicule) ou que les images proviennent d'une tête stéréo rigide (deux caméras fixées l'une par rapport à l'autre).

Bien entendu, des connaissances de différents types peuvent en principe être utilisées conjointement.

Conditions minimales. Elles dépendent fortement du nombre des paramètres internes connus a priori et du mode de variation des paramètres (e.g. si le zoom ou la mise au point automatique sont utilisés). Si par exemple toute la géométrie interne d'une caméra est connue, à l'exception de la distance focale, celle-ci peut être déterminée à partir de seulement 2 images d'une scène inconnue. Avec moins de connaissances a priori, il faut typiquement acquérir entre 3 et des dizaines d'images pour un bon résultat. Ces images doivent être prises de points de vue suffisamment "variés" : si les images sont prises par une caméra au cours d'un déplacement trop "régulier" (par exemple, une translation), l'auto-calibrage est souvent sous-contraint [7].

La plupart des approches pratiques supposent que la scène observée est tri-dimensionnelle. Il a pourtant été prouvé que même une scène plane (e.g. un mur) *inconnue* permet de faire l'auto-calibrage [9] !

3.7 D'autres problèmes

Reconstruction 3-D à partir d'une seule image. Beaucoup de scènes d'environnements humains présentent des "régularités" qui peuvent facilement être explitées pour leur reconstruction : elles sont principalement composées de surfaces planes, qui sont le plus souvent orientées verticalement ou horizontalement, etc. Ces contraintes géométriques sont faciles à définir de manière *interactive*, même à partir d'une seule image. La combinaison de ces contraintes permet souvent de créer un modèle grossier d'une scène [10] (voir un exemple dans la figure 7.

Capteurs non classiques. Actuellement, beaucoup de recherches sont faites sur des "caméras omnidirectionnelles", voir e.g. [5]. Ces caméras ont des champs de vue souvent hémisphériques, d'où leur application pour la surveillance vidéo. Elles permettent une estimation du mouvement plus précise qu'avec des caméras classiques ; elles sont donc également très intéressantes pour des applications en robotique mobile. Un exemple d'une caméra omnidirectionnelle est montré dans la figure 8.

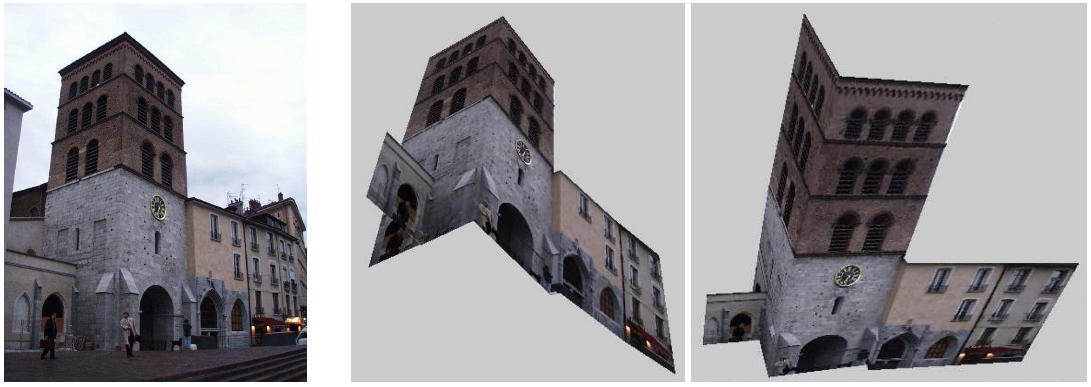


FIG. 7: *Une image et deux rendus du modèle 3-D construit à partir de cette image.*

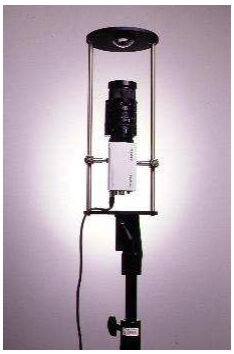


FIG. 8: *Un capteur omnidirectionnel, consistant d'une caméra et d'un miroir parabolique.*

Les capteurs omnidirectionnels se distinguent des caméras "classiques" surtout en ce qui concerne leur géométrie interne. Une fois celle-ci estimée, on peut en principe résoudre les mêmes tâches, souvent en utilisant les mêmes méthodes (voir par exemple la figure 9 qui montre un exemple d'une reconstruction interactive, à partir d'une seule image omnidirectionnelle).

Scènes dynamiques. Tout au long de ce manuscript, nous avons considéré des scènes statiques, connues ou inconnues. Naturellement, beaucoup de chercheurs s'intéressent à des scènes dynamiques, dont la modélisation est bien sûr plus complexe. Beaucoup d'approches existantes procèdent par une segmentation préalable de la scène en des composantes exhibant des mouvements indépendants, suivie par la reconstruction séquentielle de celles-ci. Actuellement, quelques recherches très intéressantes sur la modélisation simultanée des différentes composantes, voient le jour [11]. Il s'agit de travaux de nature plutôt théoriques pour le moment, et les mouvements des différentes composantes sont supposés être rigides.

D'autres travaux considèrent des mouvements non rigides, ou "semi-rigides", comme le mouvement du corps humain : il s'agit d'un mouvement articulé, qui est combiné avec le mouvement non rigide de la chaire et des vêtements.

FIG. 9: *Une image omnidirectionnelle et deux rendus du modèle 3-D construit à partir de cette image.*

4 Conclusion

Nous espérons que ce manuscript permet de cerner les grandes lignes de ce qui est faisable en vision 3-D et surtout, qu'il peut représenter une base pour des questions plus détaillées de la part du lecteur intéressé.

Références

[1] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 563–578. Springer-Verlag, May 1992.

[2] O. Faugeras, Q.-T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images*. MIT Press, March 2001.

[3] R.I. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764, 1992.

[4] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, June 2000.

[5] *Proceedings of the IEEE Workshop on Omnidirectional Vision, Hilton Head Island, South Carolina*. IEEE Computer Society Press, June 2000.

[6] C.C. Slama, editor. *Manual of Photogrammetry, Fourth Edition*. American Society of Photogrammetry and Remote Sensing, Falls Church, Virginia, USA, 1980.

[7] P. Sturm. *Vision 3D non calibrée : contributions à la reconstruction projective et étude des mouvements critiques pour l'auto-calibrage*. Thèse de doctorat, Institut National Polytechnique de Grenoble, December 1997.

[8] P. Sturm and S. Maybank. On plane-based camera calibration: A general algorithm, singularities, applications. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, pages 432–437, June 1999.

[9] B. Triggs. Autocalibration from planar scenes. In *Proceedings of the 5th European Conference on Computer Vision, Freiburg, Germany*, 1998.

[10] M. Wilczkowiak, E. Boyer, and P. Sturm. Camera calibration and 3d reconstruction from single images using parallelepipeds. In *Proceedings of the International Conference on Computer Vision, Vancouver, Canada*, pages 142–148, 2001.

[11] L. Wolf and A. Shashua. On projection matrices $P^k \rightarrow P^2, k = 3, \cdots, 6$, and their applications in computer vision. In *Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada*, volume 1, pages 412–419. IEEE Computer Society Press, July 2001.

[12] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Proceedings of the 7th International Conference on Computer Vision, Kerkyra, Greece*, September 1999.