

Mouvement par Reconstruction et Reprojection

G. Toscani, O.D. Faugeras

INRIA
Rocquencourt
France

Résumé

L'estimation du mouvement est un problème crucial en vision artificielle: dans l'analyse de scènes, la prédiction du mouvement, le contrôle de processus industriels dynamiques.

Dans cet article nous nous posons le problème suivant: étant donné un certain nombre de points mis en correspondance entre deux images prises par une caméra en mouvement, retrouver la rotation et la direction de la translation associées au mouvement de la caméra.

Les techniques existantes donnent des résultats médiocres sur des images réelles, même pour un rapport signal sur bruit élevé.

La technique que nous présentons semble être très robuste: elle a été testée sur des images synthétiques et réelles et comparée à deux autres techniques et les résultats du test sont décrits.

1 Introduction

Il y a essentiellement deux approches pour déterminer le mouvement tri-dimensionnel d'un objet rigide et isolé à partir d'une séquence d'images: méthodes basées sur le **flux optique** et méthodes basées sur l'utilisation d'**éléments discrets**.

Baron [1] passe en revue les méthodes existantes pour l'estimation du **flux optique** et la déduction du mouvement tri-dimensionnel et de la structure tridimensionnelle.

En général les **éléments discrets** peuvent être des points, des droites ou des contours. L'hypothèse fondamentale de l'analyse basée sur les éléments discrets est que le problème de la correspondance entre ces éléments a déjà été résolu avant le début du calcul du mouvement.

Le problème d'estimer le déplacement à partir de **droites** a suscité récemment un vif intérêt [2,3,4,5] et les techniques proposées semblent être robustes [7].

Diverses techniques ont été proposées aussi en utilisant les **points**. Gennery [8] a proposé une méthode itérative qui utilise de réajustements au moindres carrés mais sans mentionner de résultats expérimentaux. Cohen [9] signale que cette technique est trop sensible au bruit.

Parmi les **techniques linéaires**, Longuet-Higgins [10] donne un algorithme pour retrouver les paramètres du mouvement dans le cas d'une solution exacte. Huang [11] met en garde (page 353) que les résultats expérimentaux sur des données synthétiques et réelles, mentionnés dans [12,13] ont indiqué qu'une résolution très élevée de l'image est nécessaire pour retrouver les paramètres du mouvement de façon raisonnablement précise (à 10% près).

Pour tenir compte du bruit, nous avons proposé [14,15] une technique linéaire qui utilise la représentation des rotations par les quaternions.

Ces deux dernières techniques ont été comparées à celle décrite dans cet article, qui s'est révélée toujours au moins aussi bonne et dans la plupart des cas nettement supérieure.

2 Le modèle de transformation perspective

Nous allons décrire brièvement le modèle de la camera utilisé dans la suite.

Abstract

Estimation of motion is a crucial problem in artificial vision: examples are scene analysis, prediction of motion, control of dynamic industrial processes.

The problem we address in this paper is the following: we consider a moving camera taking a picture of the same object at two different times. Given a number of point matches between the two images, the question is to recover the rotation and the direction of the translation of the camera.

Existing techniques are reported to perform poorly on real images, even when the signal to noise ratio is big.

The technique we present in this paper seems to be quite robust: it has been tested on both synthetic as well as real images and the test results are reported and compared with other two techniques.

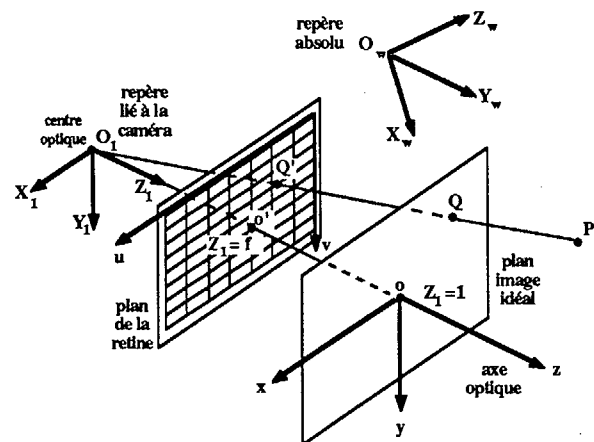


Figure 1 : Le modèle de la camera

En première approximation on considère que la transformation opérée par la caméra est une transformation perspective parfaite, de centre O_1 (le centre optique de la caméra) à une distance f (la distance focale) du plan focal (Figure 1).

Le point image Q' , situé sur la rétine et dont on connaît les coordonnées en pixels, est la projection du point P . Dans [6] on montre comment on peut projeter Q' sur un plan image idéal situé à une distance 1 du centre optique et parallèle au plan de la rétine.

À tout point Q' , de coordonnées $[u, v]^T$ en pixels dans le repère 2D de la rétine, on peut associer un point Q , de coordonnées $[x, y, 1]^T$ dans le repère 3D lié à la caméra. Dans la suite nous ne considérerons que les projections sur ce plan image idéal.

3 Position du problème

Nous allons considérer le système de la Figure 2 où la caméra a été déplacée de sa première position 1 à sa deuxième position 2 par un déplacement D .

Il y a une infinité de façons de décomposer un déplacement D

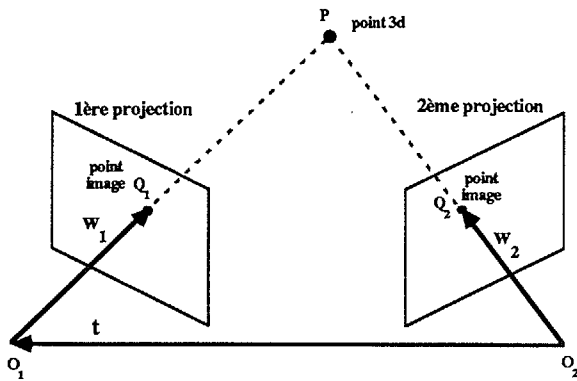


Figure 2 : Déplacement de la caméra

comme le produit d'une rotation et d'une translation. Nous allons adopter la décomposition définie par:

$$X_2 = R \cdot X_1 + t \tag{1}$$

où X_1 (resp. X_2) sont les coordonnées d'un point exprimées dans le repère lié à la première (resp. la deuxième) projection, R la matrice de rotation et t le vecteur translation. En posant $X_1 = [0, 0, 0]^T$ en 1 on a: $t = O_2 O_1$.

Il est bien connu qu'à partir d'un couple d'images stéréo et avec un tel modèle de la caméra on ne peut recouvrir le mouvement qu'à un facteur d'échelle près. Les paramètres à trouver ne sont donc que 5: 3 pour la rotation et 2 pour la translation.

Pour la rotation et la translation on va utiliser les représentations suivantes:

- Soit R une matrice de rotation et $r = [r_x, r_y, r_z]^T$ le vecteur associé à cette rotation: parallèle à l'axe de rotation et de norme égale à l'angle de rotation α . On a la représentation exponentielle suivante:

$$R = e^{\tilde{r}}$$

où \tilde{r} est la matrice définie en 7 dans l'annexe.

- Soit $t = [t_x, t_y, t_z]^T$ le vecteur translation, qu'on peut supposer unitaire, et $s = [\vartheta, \varphi]^T$, où ϑ et φ définissent la direction de la translation comme suit:

$$\begin{cases} t_x = \cos \varphi \cos \vartheta \\ t_y = \cos \varphi \sin \vartheta \\ t_z = \sin \varphi \end{cases} \begin{cases} -\pi \leq \vartheta \leq \pi \\ -\pi/2 \leq \varphi \leq \pi/2 \end{cases}$$

Les paramètres à estimer sont donc les composantes de r et s .

4 Reconstruction et reprojection

Soient Q_1 et Q_2 les projections effectives d'un point 3D P sur les deux plans image, avec $w_1 = [x_1, y_1, 1]^T$ et $w_2 = [x_2, y_2, 1]^T$ comme en Figure 2. Étant donnée un estimé de la rotation et de la translation, R_e et t_e , nous pouvons reconstruire un estimé P_e du point P .

Pour R_e et t_e donnés, les points image Q_1 et Q_2 donnent deux droites ρ_1 et ρ_2 dans l'espace qui, en général, ne s'intersectent pas, comme le montre la Figure 3. Un estimé raisonnable de P est l'intersection de la perpendiculaire commune aux deux droites et ρ_1 . Soit $Q'_2(R_e, t_e)$ la projection de P_e sur le deuxième plan image, avec $w'_2 = [x'_2, y'_2, 1]^T$. L'idée est de chercher la rotation et la translation qui minimisent les distances image $|Q'_2(R_e, t_e) Q_2|$ pour tous les points mis en correspondance.

Le critère à minimiser est donc le suivant:

$$C(R, t) = \sum |Q'_2(R, t) Q_2|^2 \tag{2}$$

Ceci peut être fait soit en calculant les dérivées et en utilisant une méthode de gradient, soit en utilisant directement une technique de différences finies.

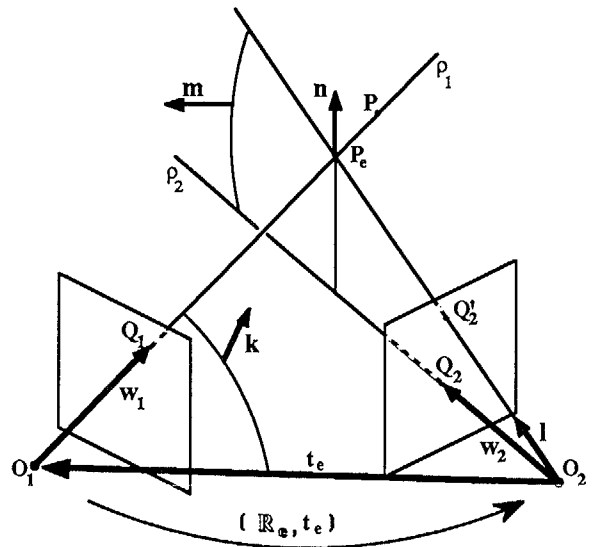


Figure 3 : Reconstruction et reprojection

4.1 Dérivation du critère

Soient:

- $[P_{ex}, P_{ey}, P_{ez}]^T$ le vecteur $O_2 P_e$ (coordonnées du point P_e dans le repère lié au deuxième plan image),
- n le vecteur de la perpendiculaire commune aux deux droites ρ_1 et ρ_2 défini par:

$$n = w_2 \wedge R_e \cdot w_1$$

- m le vecteur normal au plan passant par O_2 et contenant la perpendiculaire commune défini par:

$$m = w_2 \wedge n$$

- k le vecteur normal au plan passant par O_1, O_2 et P défini par:

$$k = t_e \wedge R_e \cdot w_1$$

- l un vecteur directeur de la droite passant par O_2 et P_e défini par:

$$l = k \wedge m = [l_x, l_y, l_z]^T$$

On a:

$$x'_2 = \frac{P_{ex}}{P_{ez}} = \frac{l_x}{l_z}$$

$$y'_2 = \frac{P_{ey}}{P_{ez}} = \frac{l_y}{l_z}$$

On peut donc réécrire le critère 2 de la façon suivante:

$$C(R_e, t_e) = \sum \left(\left(x_2 - \frac{l_x}{l_z} \right)^2 + \left(y_2 - \frac{l_y}{l_z} \right)^2 \right)$$

Dérivons par rapport à r et s :

$$\frac{\partial C}{\partial \bullet} = 2 \sum \left[\left(x_2 - \frac{l_x}{l_z} \right) \left(\frac{l_x \frac{\partial l_x}{\partial \bullet} - l_z \frac{\partial l_x}{\partial \bullet}}{l_z^2} \right) + \left(y_2 - \frac{l_y}{l_z} \right) \left(\frac{l_y \frac{\partial l_x}{\partial \bullet} - l_z \frac{\partial l_y}{\partial \bullet}}{l_z^2} \right) \right]$$

où \bullet représente le vecteur r ou s .

Il ne reste donc qu'à dériver l , en tenant compte de 8 et de 9:

$$\frac{\partial l}{\partial r} = \frac{\partial}{\partial r} (k \wedge m) = \tilde{m} \frac{\partial m}{\partial r} - \tilde{m} \frac{\partial k}{\partial r} \tag{3}$$

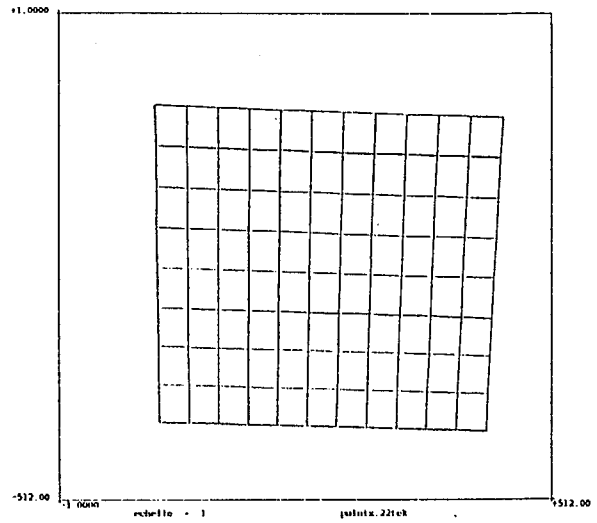
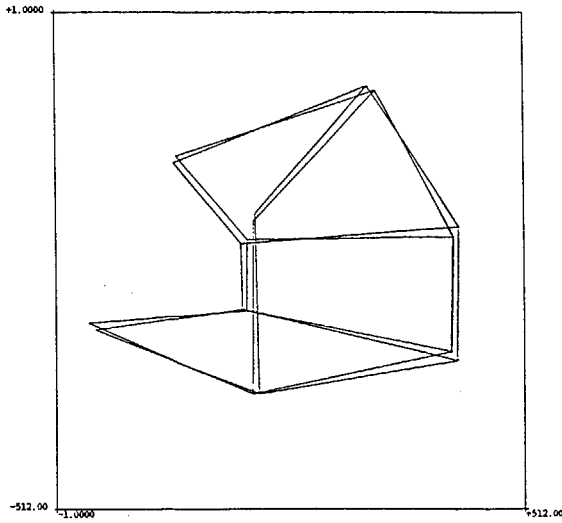


Figure 4 : Les données synthétiques parfaites et bruitées utilisées dans le test décrit en table 1.

Figure 5 : Les droites extraites de la mire de calibration utilisées dans le test décrit en table 2.

$$\frac{\partial \mathbf{k}}{\partial \mathbf{r}} = \frac{\partial}{\partial \mathbf{r}} (\mathbf{t}_e \wedge \mathbf{R}_e \cdot \mathbf{w}_1) = \tilde{\mathbf{t}}_e \frac{\partial \mathbf{R}_e \mathbf{w}_1}{\partial \mathbf{r}} = \tilde{\mathbf{t}}_e \mathbf{K}(\mathbf{R}_e, \mathbf{w}_1) \quad (4)$$

$$\frac{\partial \mathbf{m}}{\partial \mathbf{r}} = \frac{\partial}{\partial \mathbf{r}} (\mathbf{w}_2 \wedge \mathbf{n}) = \tilde{\mathbf{w}}_2 \frac{\partial \mathbf{n}}{\partial \mathbf{r}} \quad (5)$$

$$\frac{\partial \mathbf{n}}{\partial \mathbf{r}} = \frac{\partial}{\partial \mathbf{r}} (\mathbf{w}_2 \wedge \mathbf{R}_e \cdot \mathbf{w}_1) = \tilde{\mathbf{w}}_2 \frac{\partial \mathbf{R}_e \mathbf{w}_1}{\partial \mathbf{r}} = \tilde{\mathbf{w}}_2 \mathbf{K}(\mathbf{R}_e, \mathbf{w}_1) \quad (6)$$

En remplaçant 4, 5 et 6 dans 3 on obtient:

$$\frac{\partial \mathbf{l}}{\partial \mathbf{r}} = \left(\tilde{\mathbf{k}} \tilde{\mathbf{w}}_2^2 - \tilde{\mathbf{m}} \tilde{\mathbf{t}}_e \right) \mathbf{K}(\mathbf{R}_e, \mathbf{w}_1) + \left(\tilde{\mathbf{t}}_e \mathbf{R}_e \mathbf{w}_1 \tilde{\mathbf{w}}_2^2 - \tilde{\mathbf{w}}_2^2 \mathbf{R}_e \mathbf{w}_1 \tilde{\mathbf{t}}_e \right) \mathbf{K}(\mathbf{R}_e, \mathbf{w}_1)$$

Dérivons maintenant l par rapport à la translation.

On a:

$$\frac{\partial \mathbf{l}}{\partial \mathbf{s}} = \frac{\partial \mathbf{l}}{\partial \mathbf{t}} \frac{\partial \mathbf{t}}{\partial \mathbf{s}}$$

$$\frac{\partial \mathbf{t}}{\partial \mathbf{s}} = \begin{bmatrix} -\cos \varphi \sin \vartheta & -\sin \varphi \cos \vartheta \\ \cos \varphi \cos \vartheta & -\sin \varphi \sin \vartheta \\ 0 & \cos \varphi \end{bmatrix}$$

En dérivant l comme précédemment on obtient facilement:

$$\frac{\partial \mathbf{l}}{\partial \mathbf{t}} = \tilde{\mathbf{m}} \mathbf{R}_e \tilde{\mathbf{w}}_1$$

5 Resultats

Nous comparons la technique de reconstruction et reprojection avec une version améliorée de l'algorithme de Longuet-Higgins [10] décrite dans [7] et l'algorithme linéaire proposé dans [14] et développé dans [7,15].

Pour les données synthétiques on a utilisé un objet à forme de maison (Figure 4) constitué de 10 points, et les résultats sont reportés dans la Table 1.

Pour les données réelles on a considéré les points utilisés dans le processus de calibration (Figure 5). La rotation et translation de référence sont donc celles calibrées; les résultats sont reportés dans la Table 2.

ϑ	t	δp	av	sd	ϑ'	$\vartheta\%$	α	β	
5	10	1	57.8	7.2	5.89	17.8	24.59	26.24	(1)
					5.36	7.2	11.48	27.58	(2)
					5.21	4.2	2.60	4.12	(3)
5	50	1	91.8	14.3	15.18	203.6	42.21	4.53	(1)
					5.72	14.5	6.78	1.78	(2)
					5.25	5.0	3.31	0.78	(3)
5	100	1	139.6	31.1	11.03	120.7	37.31	3.09	(1)
					5.74	14.8	6.30	0.82	(2)
					5.25	5.1	3.38	0.38	(3)
5	10	5	57.7	8.1	6.06	21.2	47.34	72.43	(1)
					5.04	0.8	16.94	72.61	(2)
					5.14	2.8	8.61	71.91	(3)
5	50	5	91.5	14.7	20.01	300.2	115.43	24.59	(1)
					8.37	67.3	37.77	11.48	(2)
					5.99	19.8	10.12	4.31	(3)
5	100	5	139.4	31.1	55.90	1018	121	23.2	(1)
					9.09	82	22.3	5.06	(2)
					6.17	23.4	11.81	2.11	(3)
15	50	5	189.1	20.4	34.6	130.7	35.6	6.68	(1)
					19.5	30.3	8.5	7.84	(2)
					15.05	0.14	2.78	5.05	(3)
15	100	5	232.3	30.3	25.15	67.6	24.98	4.19	(1)
					19.73	31.6	5.37	1.93	(2)
					15.24	1.6	2.63	2.64	(3)

axe de rotation: (1,1,1)		direction de la translation: (1,0,0)	
ϑ	= angle de rotation (deg)	$\vartheta\%$	= erreur angle de rotation (en %)
t	= module de la translation exacte (cm)	α	= angle entre l'axe de rotation calculé et celui de référence
δp	= bruit de mesure (pixels)	β	= angle entre la direction de la translation calculée et celle de référence
av	= déplacement moyen image (pixels)	(1)	= Longuet-Higgins modifié
sd	= écart type du déplacement image (pixels)	(2)	= technique linéaire (quaternions)
ϑ'	= angle calculé de la rotation	(3)	= reconstruction et reprojection

Table 1 : Données synthétiques



	θ	t	av	sd	θ'	$\theta\%$	α	β	
I2S	3.48	8	13.4	6.2	3.36	3.3	0.96	0.93	(1)
					3.37	3.1	0.75	0.91	(2)
					3.47	0.2	0.81	0.70	(3)
I2S	5.32	44	20.6	5.2	14.64	175	58.1	7.25	(1)
					5.13	3.6	0.41	0.76	(2)
					5.14	3.5	0.66	0.31	(3)
I2S	4.69	37	24.9	7.2	4.60	2.0	4.16	2.18	(1)
					4.58	2.4	2.40	2.17	(2)
					4.81	2.5	0.53	0.92	(3)
Pul	9.71	53	26.2	16.4	9.55	1.5	3.10	0.52	(1)
					9.68	0.2	0.32	0.02	(2)
					9.69	0.2	0.80	0.008	(3)
Pul	9.15	49	20.7	12.2	9.43	3.0	1.36	0.22	(1)
					9.06	1.0	0.61	0.22	(2)
					9.03	1.3	0.22	0.23	(3)
Pul	8.27	45	18.8	12.4	8.13	1.6	2.74	0.31	(1)
					8.14	1.5	0.65	0.01	(2)
					8.29	0.25	0.02	0.02	(3)

θ = angle de rotation (deg)	α = angle entre l'axe de rotation calculé et celui de référence
t = module de la translation de référence (cm)	β = angle entre la direction de la translation calculée et celle de référence
av = moyenne du déplacement image (pixels)	(1) = Longuet-Higgins modifié
sd = écart type du déplacement image (pixels)	(2) = technique linéaire (quaternions)
θ' = angle de la rotation calculée	(3) = reconstruction et reprojection
$\theta\%$ = erreur angle de rotation (en %)	I2S = caméra I2S CCD
	Pul = caméra Pulnix CCD

Table 2 : Données réelles

6 Annexe

On adopte les notations et conventions suivantes:

- la dérivée d'un scalaire f par rapport à un vecteur $\mathbf{a} = [a_x, a_y, a_z]^T$ est un vecteur colonne:

$$\frac{df}{d\mathbf{a}} = \left[\frac{df}{da_x}, \frac{df}{da_y}, \frac{df}{da_z} \right]^T$$

- la dérivée d'un vecteur \mathbf{b} par rapport à un vecteur \mathbf{a} est une matrice:

$$\frac{d\mathbf{b}}{d\mathbf{a}} = \begin{bmatrix} \frac{db_x}{da_x} & \frac{db_x}{da_y} & \frac{db_x}{da_z} \\ \frac{db_y}{da_x} & \frac{db_y}{da_y} & \frac{db_y}{da_z} \\ \frac{db_z}{da_x} & \frac{db_z}{da_y} & \frac{db_z}{da_z} \end{bmatrix}$$

- pour tout vecteur $\mathbf{a} = [a_x, a_y, a_z]^T$, on note par $\tilde{\mathbf{a}}$ la matrice

$$\tilde{\mathbf{a}} \stackrel{\text{def}}{=} \begin{bmatrix} 0 & -a_z & a_y \\ a_x & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \quad (7)$$

On a le résultat suivant:

$$\frac{d}{dc} (\mathbf{a} \wedge \mathbf{b}) = \tilde{\mathbf{a}} \frac{d\mathbf{b}}{dc} - \tilde{\mathbf{b}} \frac{d\mathbf{a}}{dc} \quad (8)$$

- soit \mathbf{R} une matrice de rotation et \mathbf{r} le vecteur associé à cette rotation. En posant: $f(\alpha) = \frac{\sin \alpha}{\alpha}$ et $g(\alpha) = \frac{1 - \cos \alpha}{\alpha^2}$ et en utilisant la formule de Rodrigues:

$$\mathbf{R} = \mathbf{I} + f(\alpha)\tilde{\mathbf{r}} + g(\alpha)\tilde{\mathbf{r}}^2$$

où \mathbf{I} est la matrice identité, pour tout vecteur \mathbf{x} on a:

$$\begin{aligned} \frac{d}{d\mathbf{r}} (\mathbf{R} \cdot \mathbf{x}) &= \frac{f'(\alpha)}{\alpha} \tilde{\mathbf{r}} \mathbf{x} \mathbf{r}^T + \frac{g'(\alpha)}{\alpha} \tilde{\mathbf{r}}^2 \mathbf{x} \mathbf{r}^T - \\ &\quad - f(\alpha)\tilde{\mathbf{x}} - g(\alpha) \left(\tilde{\mathbf{r}} \mathbf{x} + \tilde{\mathbf{r}} \mathbf{x} \right) = \\ &\stackrel{\text{def}}{=} \mathbf{K}(\mathbf{R}, \mathbf{x}) \end{aligned} \quad (9)$$

7 Remerciements

Ce travail à été financé en partie par le projet Esprit P940.

References

- [1] BARON J., 1984, "A Survey of Approaches for Determining Optical Flow, Environmental Layout and Egomotion", University of Toronto, RBCV-TR-84-5, November 1984.
- [2] LIU Y., HUANG T.S, 1986, "Estimation of Rigid Body Motion Using Straight Line Correspondence", Proc. Workshop on Motion: Representation and Analysis, May 7-9, 1986, Charleston, South Carolina, pp 47-52.
- [3] LIU Y., HUANG T.S, 1986, "Estimation of Rigid Body Motion Using Straight Line Correspondence, Further Results", Proc. ICPR 1986, Paris, pp 306-307.
- [4] MITICHE A., SEIDA S., AGGARWAL J.K., 1986, "Line Based Computation of Structure and Motion Using Angular Invariance", Proc. Workshop on Motion: Representation and Analysis, May 7-9, 1986, Charleston, South Carolina, pp 175-180.
- [5] MITICHE A., SEIDA S., AGGARWAL J.K., 1986, "Interpretation of Structure and Motion Using Straight Line Correspondence, Further Results", Proc. ICPR 1986, Paris, pp 1110-1112.
- [6] FAUGERAS O.D., TOSCANI G., 1987, "Camera Calibration for 3D Machine Vision", Proc. International Workshop on Machine Vision and Machine Intelligence, Tokyo, Japan, February 2-5, 1987, pp 240-247.
- [7] FAUGERAS O.D., LUSTMAN F., TOSCANI G., 1987, "Motion and Structure from Motion from Point and Line Matches", Proc. First International Conference on Computer Vision, London, U.K., June 8-11, 1987.
- [8] GENNERLY D.B., 1979, "Stereo-Camera Calibration", Proc. Image Understanding Workshop, November 1979, pp. 101-108.
- [9] COHEN R.R., FEIGENBAUM E.A., Editors, 1982, "Handbook of Artificial Intelligence", Heuris Tech Press, William Kaufmann, Inc..
- [10] LONGUET-HIGGINS H.C., 1981, "A Computer Algorithm for Reconstructing a Scene from two Projections", Nature, Vol 293, p 133-138, Sep 1981.
- [11] HUANG T.S., 1986, "Determining Three-Dimensional Motion and Structure from Two Perspective Views", Handbook of Pattern Recognition and Image Processing, Academic Press, New York.
- [12] FANG J.-Q., HUANG T.S., 1984, "Solving Three-Dimensional Small-Rotation Motion Equations: Uniqueness, Algorithms, and Numerical Results", Computer Vision, Graphics, and Image Processing, 26, pagg. 183-206.
- [13] FANG J.-Q., HUANG T.S., 1984, "Some Experiments on Estimating the 3-D Motion Parameters of a Rigid Body from Two Consecutive Image Frames", IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI-6 (5), pp 547-554.
- [14] FAUGERAS O.D., TOSCANI G., 1986, "The Calibration Problem for Stereo", Proc. Computer Vision and Pattern Recognition, Miami, pp. 15-20.
- [15] TOSCANI G., FAUGERAS O.D., 1987, "Structure and Motion from Two Noisy Perspective Views", Proc. IEEE International Conference on Robotics and Automation, Raleigh, North Carolina, USA, March 30-April 3, 1987.