

Approche d'Estimation du Mouvement en Temps Réel par Vision Monoculaire.

M. ELARBI BOUDIHIR, S. KHELFI, M. BENMOHAMED, A. RAHMOUN
Intelligent Control and Systems Laboratory
Institut d'informatique. Route d'Oran, W: Sidi Bel-Abbès 22000.
Tel/Fax: (213) 757 77 38. Email: Elarbi-boudiher@mailcity.com

Résumé : L'objectif principal de l'analyse du mouvement est l'estimation des composantes 3-D du déplacement d'un objet sous observation. Nous avons développé un système de vision pour un robot manipulateur ayant pour tâche la reconnaissance, la préhension et le classement d'objets se déplaçant sur un tapis roulant. Dans ce cadre, on présente une méthode de segmentation spatiale de séquences d'images qui s'appuie sur une phase de reconnaissance de formes et de prédiction des déplacements par estimation du flot optique en utilisant un algorithme de filtrage spatio-temporel. En outre, on a amélioré de façon importante les performances du processus d'estimation du déplacement au niveau du filtrage. Les résultats expérimentaux montrent la robustesse de l'approche proposée.

Mots clés: *Vision Dynamique, Estimation du Mouvement, segmentation d'Images, Poursuite d'Objets en 3D, Flot Optique.*

1. Introduction

Actuellement, le besoin d'effectuer l'analyse du mouvement en temps réel pour la navigation des robots, la manipulation et le suivi des objets et d'autres applications a suscité un intérêt dans l'estimation des composantes 3-D d'un mouvement en temps réel (temps réel veut dire au moins 20 estimations par seconde). Les séquences d'images numériques offrent un support naturel pour l'analyse du mouvement. L'aspect principal de cette analyse est l'estimation des composantes 3D du déplacement des objets[1]. Ceci s'applique souvent pour résoudre les problèmes de reconstruction 3D des objets, poursuite et préhension des objets, navigation des robots mobiles, etc. L'entité mathématique généralement retenue pour porter l'information du mouvement correspond au champ des vitesses apparentes 2D, projection sur le plan image du champ des vecteurs vitesse 3D de la scène en mouvement. L'estimation du champ des vitesses 2D est obtenue par analyse des variations spatio-temporelles de la fonction luminance dans la séquence[9]. L'analyse du mouvement comporte généralement deux phases : La première concerne l'estimation du mouvement dans le plan image. Ceci fournit un ensemble de vecteurs de déplacements plans plus ou moins dense. La seconde phase s'appuie sur ces vecteurs pour reconstruire des vecteurs de déplacements pouvant s'interpréter en mouvement 3D[3][7]. Ceci implique toujours l'utilisation des modèles de scènes ou d'objets pour effectuer une interprétation. A l'opposé, la première phase doit pouvoir être effectuée plus en aveugle en se limitant à quelques hypothèses simples pour s'adapter à des variétés de situations possibles résultant des diverses applications.

En robotique, la détection et l'estimation de mouvement participent à des tâches passives comme la surveillance, la trajectographie de cible, mais aussi à des tâches actives comme la navigation et l'évitement d'obstacles. Le mouvement est aussi à la base des techniques de compression vidéo dont l'idée directrice est, à partir d'une image donnée, de ne conserver de la suivante que ce qui a changé. Les techniques de détection et d'estimation de mouvement touchent ainsi le domaine très vaste des images numériques pour des applications telles que les réseaux ou la télévision numérique.

Différentes approches ont été proposées pour obtenir un ensemble plus ou moins dense de vecteurs de déplacements plans entre deux images successives d'une séquence. Elles se répartissent en deux

familles : les méthodes différentielles et les méthodes de mise en correspondance de structures caractéristiques.

Pour notre application, l'objectif est d'estimer le déplacement des objets en mouvement sur un tapis roulant pour leur préhension et rangement par un robot manipulateur. L'orientation des objets est supposée aléatoire, ainsi une étape de reconnaissance des formes basée sur des critères a priori connus est nécessaire. Par conséquent, on a combiné une approche de haut niveau composée de la méthode de reconnaissance de formes, avec une approche de bas niveau basée sur un algorithme de filtrage passe-bandes spatio-temporel.

2. Phase d'Identification

La méthode d'identification que nous avons utilisé, n'est pas universelle mais plutôt rapide et simple. L'objectif fixé ici est de reconnaître et de localiser un objet dans chaque image d'une séquence puis à partir de deux images successives déterminer les valeurs des paramètres de son déplacement. L'objet suivi est plan et son modèle est défini par un ensemble de chaînes fermées de segments orientés (polygones)[1]. Ce modèle contient à la fois des informations nécessaires pour la reconnaissance et la localisation (coordonnées des arêtes exprimées dans un repère lié à l'objet).

Toutes ces caractéristiques sont topologiques et donc indépendantes. La tâche envisagée par le robot manipulateur se trouve caractérisée par les points suivants :

- L'utilisateur n'a initialement aucune connaissance a priori de la scène
- L'acquisition des images est faite à l'aide d'une caméra fixe perpendiculaire au plan du tapis roulant.
- Les objets ne sont modélisés que par des contours fermés.
- Les objets ne doivent pas subir des déformations lors de l'acquisition la séquence.

Le système procède par une recherche automatique basée sur une phase d'apprentissage et une phase d'identification. Notre objectif consiste à essayer d'obtenir le contour fermé le plus fidèle possible d'un objet. Par conséquent, suivre un contour peut être défini comme étant l'action de chaîner un certain nombre de points délimitant une région. Le suivi de contour introduit donc la notion de voisinage[10]. Les voisinages les plus couramment employés sont le voisinage de 4 points et le voisinage de 8 points. Pratiquement le suivi s'effectue par un balayage ligne par ligne afin de détecter le premier point de contour de l'objet, qui constituera le pivot pour la recherche du voisin proche (au sens du voisinage de 8). Ces deux points déterminent alors la première direction du codage. Le critère d'arrêt de cet opérateur est la détection d'un point de contour considéré suffisamment proche du point d'origine.

L'étape suivante sert à modéliser l'objet qui s'établira sous la forme d'un vecteur V_o (Périmètre, Aire, Rapport de finesse, Rayon minimum, Rayon maximum) soit un vecteur de 5 composantes. Ce vecteur de modélisation est le cœur du processus de reconnaissance de formes. Ces paramètres sont donnés par :

Périmètre

$$P = \sum_{i=0}^{i=n} d(x_i, x_{i+1} + 1) \quad (1)$$

où x_i et x_{i+1} sont deux points consécutifs du contour. La détermination de certains de ces paramètres nécessite le calcul des moments principaux d'inertie M_{pq} jusqu'à l'ordre 2 ($p+q=2$)

$$M_{pq} = \sum_i \sum_j i^q j^p f(i, j) \quad (2)$$

$f(i, j) \rightarrow 1$ pour l'objet

$f(i, j) \rightarrow 0$ pour le fond

la surface

$$S = M_{00} \quad (3)$$

Ainsi, le rapport de finesse est donné par :

$$R_f = \frac{4\pi * M_{00}}{(\text{périmètre})^2} \quad (4)$$

Le facteur d'échelle est un rapport de proportion qui permet de relier 2 objets, à l'origine identiques mais qui sous une suite d'acquisitions différentes apparaissent éventuellement avec des tailles différentes :

$$\lambda = \left(\frac{M'_{pq}}{M_{pq}} \right)^{\frac{1}{2+p+q}} \quad (5)$$

Le processus d'identification est établi par minimisation de l'erreur quadratique entre les composantes du vecteur objet V_o et le vecteur modèle V_m :

$$E = \sum_{i=1}^5 \|V_{oi} - V_{mi}\|^2 \quad (6)$$

3. Phase du Filtrage

La méthode de segmentation et de suivi temporel qu'on présente s'appuie sur les cohérences temporelles existantes entre les images successives d'une même séquence. Le principe de la méthode repose sur la considération de la séquence temporelle des images 2-D comme une image spatio-temporelle en 3-D. Ensuite des filtres tri-dimensionnels sont appliqués à au volume image dans l'échelle espace-temps. Donc, si les filtres sont choisis avec une orientation, une localisation, et une sensibilité par rapport à un intervalle de fréquences spatio-temporelles, le mouvement peut alors être estimé avec une bonne précision. Il existe deux stratégies de base pour l'utilisation du volume spatio-temporel 3-D :

- 1- La détection du mouvement en deux dimensions spatiales devient un problème de détection d'orientation du plan dans le volume spatio-temporel 3-D.
- 2- Le mouvement en 2-D d'un objet ayant une texture devient un problème de recherche de textures orientées 3-D dans le volume spatio-temporel.

La méthode du filtrage spatio-temporelle englobe les deux approches. En outre, son implémentation sur les systèmes de vision temps réel devient simplifiée et pratique grâce aux circuits vidéo dédiés. L'idée est illustrée par la figure 1 suivante :

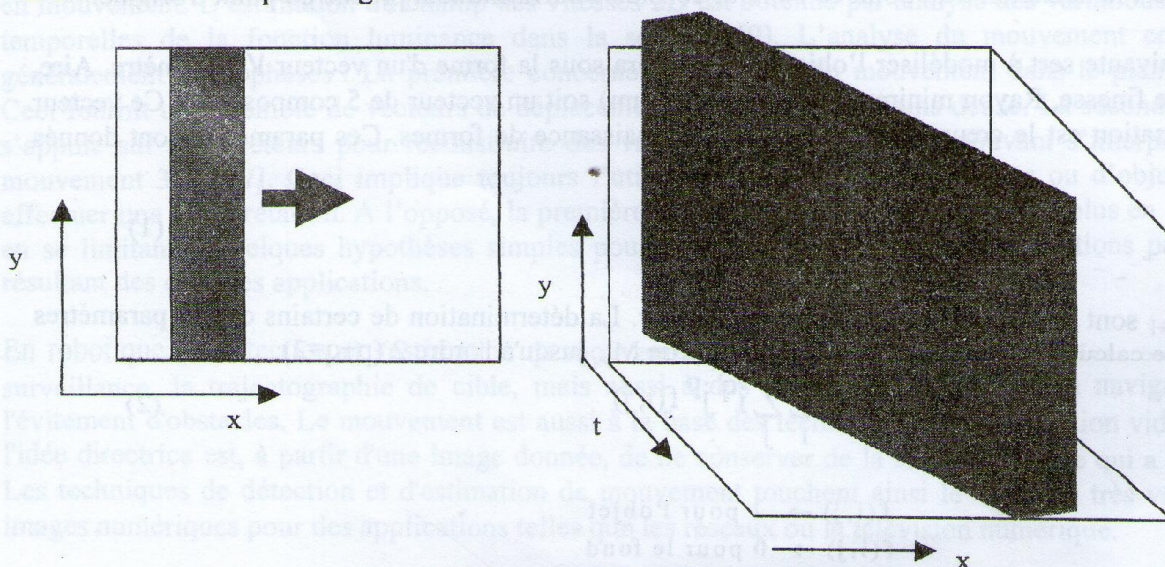


Figure1 :Mouvement d'un objet en espace-temps.

Un filtre 3-D orienté peut détecter l'orientation de la surface. Ainsi, on remarque que le mouvement dans le domaine spatial peut être identifié avec l'attribut orientation dans le repère espace-temps. En d'autres termes, si une barre verticale se déplace horizontalement (direction des x) sur le plan x-y,

alors elle génère dans le volume x-y-t un solide dont l'orientation peut être facilement détectée par un filtre détecteur de contours 3-D.

Supposons un objet comme une texture en 2-D avec ω_x sa fréquence spatiale en x, et ω_y sa fréquence spatiale en y. Quand l'objet se déplace avec une vitesse (u, v) , sa fréquence temporelle sera donnée par :

$$u\omega_x + v\omega_y = -\omega_t \quad (7)$$

Pour une meilleure illustration, la translation d'une image $f(x, y)$ avec une vitesse (u, v) génère un volume spatio-temporel défini par :

$$f(x, y, t) = f(x, y) * \delta(x - ut, y - vt) \quad (8)$$

où * représente une convolution et δ la fonction delta 2-D. La transformée de Fourier de l'équation (8) donne :

$$\begin{aligned} \hat{f}(\omega_x, \omega_y, \omega_t) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y, t) e^{-j(x\omega_x + y\omega_y + t\omega_t)} dx dy dt \\ &= \int_{-\infty}^{+\infty} e^{-jt\omega_t} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) * \delta(x - ut, y - vt) e^{-j(x\omega_x + y\omega_y)} dx dy dt \\ &= \hat{f}(\omega_x, \omega_y) \int_{-\infty}^{+\infty} e^{-jt\omega_t} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(x - ut, y - vt) e^{-j(x\omega_x + y\omega_y)} dx dy dt \\ &= \hat{f}(\omega_x, \omega_y) \int_{-\infty}^{+\infty} e^{-jt\omega_t} \int_{-\infty}^{+\infty} e^{-j(x\omega_x + y\omega_y)} dx dy dt \\ &= \hat{f}(\omega_x, \omega_y) \delta(t\omega_t) \delta(u\omega_x + v\omega_y) \\ &= \hat{f}(\omega_x, \omega_y) \delta(t\omega_t + u\omega_x + v\omega_y) \end{aligned} \quad (9)$$

Ceci décrit le plan (7) dans l'espace $(\omega_x, \omega_y, \omega_t)$, car $\hat{f}(\omega_x, \omega_y, \omega_t)$ est différente de zéro seulement où $\delta(u\omega_x + v\omega_y + t\omega_t) \neq 0$

Donc, si une image contient plusieurs textures orientées, le triplet ordonné dans le domaine fréquentielle $(\omega_x, \omega_y, \omega_t)$ du solide appartenant à l'objet formera un plan. Par conséquent, pour estimer la vitesse d'un objet, il faut :

- 1- Trouver les textures orientées de fréquences $(\omega_x, \omega_y, \omega_t)$ dans le volume spatio-temporel. Ceci peut être établi avec trois filtres passe-bandes, par exemple :
- 2- Tracer les points correspondants dans le domaine de fréquence spatio-temporelle
- 3- Trouver le plan qui englobe les points correspondants
- 4- A partir de la relation (7), déterminer l'estimation du flot optique (u, v) de l'objet

La figure 2 montre le concept d'estimation du flot optique en utilisant le domaine fréquentiel espace-temps.

La motivation suscitée par l'étude des algorithmes de filtrage spatio-temporel réside dans le lien qu'ils ont en commun avec les systèmes de vision biologiques. Plusieurs chercheurs dans le domaine ont affirmé que les zones de vision du cerveau fonctionnent comme des filtres passe bandes spatio-temporels. En ce qui concerne la détection du mouvement, les systèmes de vision des animaux sont nettement supérieurs aux systèmes de vision artificiels.

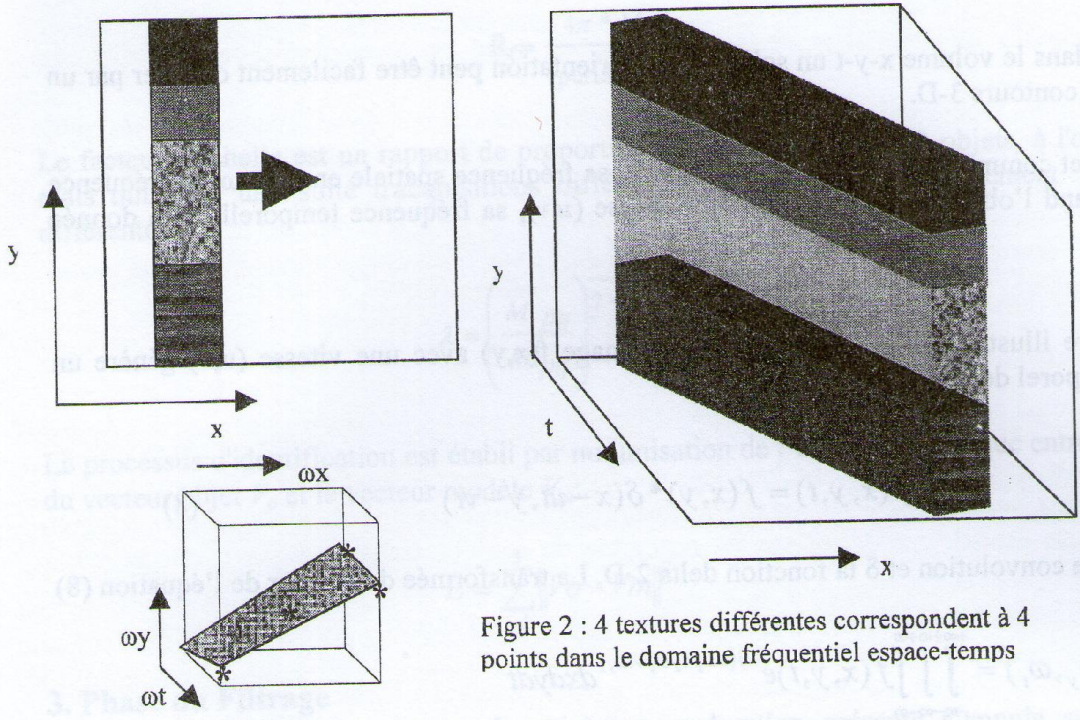


Figure 2 : 4 textures différentes correspondent à 4 points dans le domaine fréquentiel espace-temps

4. Résultats expérimentaux

Pour effectuer la tâche de préhension des objets on a utilisé un petit robot manipulateur comme le montre la figure 3.

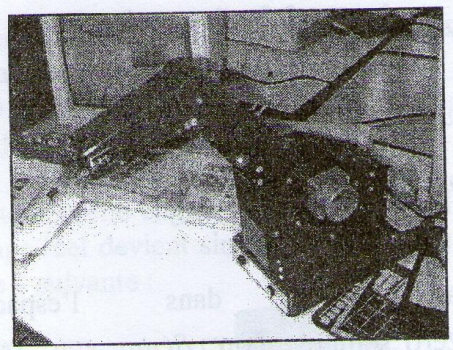
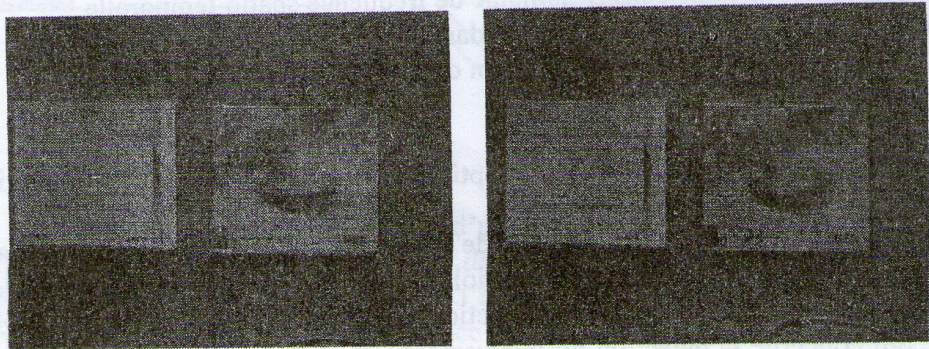


Figure 3 : robot manipulateur

Pour illustrer en général le processus, on considère deux images de tailles 100x100 pixels, contenant les mêmes objets prises à des instants successifs sous différentes conditions comme le montre les images illustrées par la figure 4. Le mouvement qui a généré l'image b est une pure translation dans le sens positif des X. Après détection de contours et modélisation, l'algorithme de reconnaissance des formes identifie dans l'image b les objets contenus dans l'image a de la figure 4.



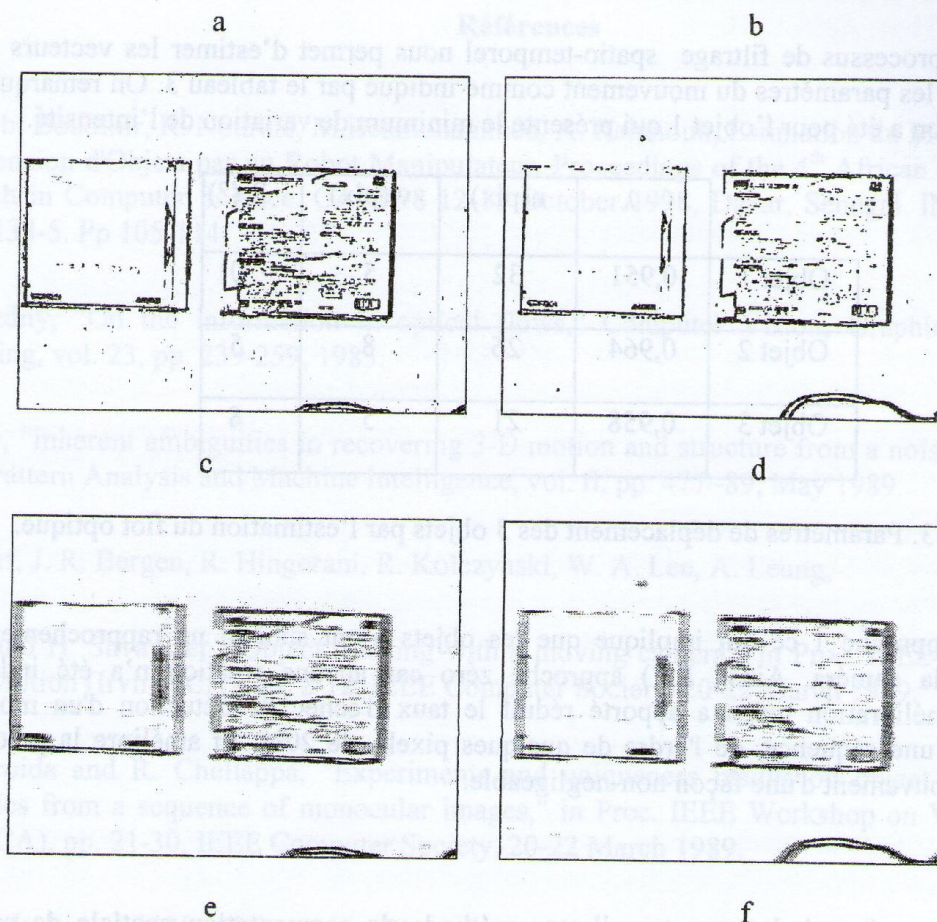


Figure 4 : Séquence de deux images consécutives. a,b : images originales, c et d : images respectives après la phase d'identification. e et f images respectives après la phase de filtrage

Par conséquent, durant la phase d'identification les paramètres de chaque objet sont ainsi déterminés comme le montre le tableau suivant :

	Image N° 1			image N° 2		
	obj 1	obj 2	obj 3	obj 1	obj 2	obj 3
Aire (pix)	1603	1501	46	1665	1572	155
Périmètre(pix)	165	127	96	123	132	103
Rayon minimum	22	18	3	26	21	5

Tableau 1. Paramètres calculés des objets

Les paramètres mesurés du déplacement sont finalement donnés par le tableau 2.

	Tx (pix)	Ty (pix)	F	R(°)
Objet 1	32	7	0,775	2
Objet 2	29	8	1,057	1
Objet 3	23	3	0,951	2

Tableau 2. Paramètres mesurés du déplacement de chaque objet.

Enfin, le processus de filtrage spatio-temporel nous permet d'estimer les vecteurs du flot optique, ainsi que les paramètres du mouvement comme indiqué par le tableau 3. On remarque que la meilleure estimation a été pour l'objet 1 qui présente le minimum de variation de l'intensité.

	λ	u(pix)	v (pix)	R (°)
Objet 1	0,951	32	5	0
Objet 2	0,964	26	8	5
Objet 3	0,958	21	5	8

Tableau 3. Paramètres de déplacement des 3 objets par l'estimation du flot optique.

En outre, λ approche 1 ce qui implique que les objets n'ont subi ni un rapprochement ni un éloignement de la caméra. Ainsi, $R(^{\circ})$ approche zéro car aucune rotation n'a été induite. En conséquence, l'amélioration qu'on a apporté réduit le taux d'échec de détection d'un mouvement quelconque dans une séquence, de l'ordre de quelques pixels, de 20%. Et améliore la précision de l'estimation du mouvement d'une façon non-négligeable.

5. Conclusion

Nous avons présenté dans ce travail une méthode de segmentation spatiale de séquences d'images qui s'appuie sur une phase de reconnaissance de formes et de prédiction des déplacements par filtrage spatio-temporel des points d'intérêts caractéristiques et descriptifs des scènes successives. Ceux-ci sont déterminés par une segmentation suivie d'un filtrage passe-bande. L'association de ces traits avec les contours des objets établis durant la phase de reconnaissance de formes constitue un ensemble suffisamment varié et pertinent pour que le filtrage soit robuste et efficace.

En outre, on a amélioré de façon importante les performances du processus de filtrage au niveau des points d'intérêts. Ceci concerne l'utilisation d'une méthode basée sur la relaxation et en même temps conditionnée par la prédiction qui focalise la recherche. Des raffinements sont sans doute possibles au niveau de la modélisation du déplacement (surtout si ce dernier n'est plus uniforme et constant).

Références

- 1) M. Elarbi Boudihir, R. Nourine, M. Benmohammed, A. Rahmoun, Estimation du Mouvement pour la Préhension d'Objets par un Robot Manipulateur. Proceedings of the 4th African Conference on Research in Computer Science. CARI'98 12-15 October 1998, Dakar, Senegal. INRIA ISBN 2-7261-1133-5. Pp 105-114.
- 2) K. Prazdny, "On the information in optical flows," Computer Vision, Graphics, and Image Processing, vol. 23, pp. 239-259, 1983.
- 3) G. Adiv, "Inherent ambiguities in recovering 3-D motion and structure from a noisy field," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. II, pp. 477-89, May 1989.
- 4) P. J. Burt, J. R. Bergen, R. Hingorani, R. Kolczynski, W. A. Lee, A. Leung,
- 5) Lubin, and H. Shvaytser, "Object tracking with a moving camera," in Proc. IEEE Workshop on Visual Motion (Irvine, CA), pp. 2-12, IEEE Computer Society, 20-22 March 1989.
- 6) T. J. Brodia and R. Chellappa, "Experiments and uniqueness results on object structure and kinematics from a sequence of monocular images," in Proc. IEEE Workshop on Visual Motion (Irvine, CA), pp. 21-30, IEEE Computer Society, 20-22 March 1989.
- 7) R. C. Nelson and J. Y. Aloimonos, "Obstacle avoidance using field divergence," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. II, pp. 1102-1106, October 1989.
- 8) M. Subbarao, "Bounds on time-to-collision and rotation component from first-order derivatives of image flow," Computer Vision, Graphics, and Image Processing, vol. 50, pp. 329-341, 1990.
- 9) S. Uliman, The Interpretation Of Visual Motion. Cambridge, MA and London, England: The MIT Press, 1979.
- 10) R. Y Tsai and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of a rigid object with curved surfaces," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 6, no. 3, 1984.
- 11) H. Shariat and K. E. Price, "Motion estimation with more than two frames," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, pp. 417-34, May 1990.
- 12) K. Jain. Fundamentals of Digital Image Processing. Englewood Cliffs, NJ: Prentice Hall, 1989.
- 13) Verri and T. Poggio, "Against quantitative optical flow," in Proc. First Int. Conf Computer Vision, Washington, D.C.: IEEE Computer Society Press, pp. 171-200, April 1988.
- 14) Singh. Optic Flow Computation: A Unified Perspective. Los Alamitos, CA: IEEE Computer Society Press, 1991.
- 15) Girod, D. J. Le Gail, M. I. Sezan, M. Vetterli, and H. Yasuda, "Guest editorial: Introduction to the special issue on image sequence compression," IEEE Trans Image Processing, vol. 3, no. 5, pp. 465-68, September 1994.

- 16) G. K. Allache, "The JPEG still picture compression standard," *Communications of the ACM*, vol. 34, no. 4, pp. 31-4, April 1991.
- 17) Le Gaïl, "MPEG: a video compression standard for multimedia applications," *Communications of the ACM*, vol. 34, no. 4, pp. 46-58, April 1991.
- 18) *Signal Processing: Image Communication*, vol. 5, no. 1, February 1993.
- 19) H. Li, A. Lundinark, and R. Forchheimer, "Image sequence coding at very low bitrates: a review," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 599-609, September 1994.
- 20) L. Fennema and W. B. Thompson, "Velocity determination in scenes containing several moving objects," *Computer Vision, Graphics, and Image Processing*, vol. 9, pp. 301-315, 1979.
- 21) K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.
- 22) M. Elarbi Boudihir, B. Boudjra, L. Kaddouri. "Poursuite d'un objet par un robot mobile en utilisant la vision 3-D". *Rob' 95. Actes des Journées sur la robotique et son environnement. 16-18 septembre 95. (Algérie) pp103-110*
- 23) H.-H. Nagel, "Displacement vectors derived from second order intensity variations in image sequences," *Computer Vision, Graphics, and Image Processing*, vol. 21, pp. 85-117, 1983.
- 24) M. Yachida, "Determining velocity maps by spatio-temporal neighborhoods from image sequences," *Computer Vision, Graphics, and Image Processing*, vol. 21, pp. 262-279, February 1983.
- 25) H.-H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 5, pp. 565-593, September 1986.
- 26) G. Schunck, "Image flow: fundamentals and future research," *Proc. IEEE Conf Computer Vision and Pattern Recognition*, pp. 560-571, June 1985.
- 27) H.-H. Nagel, "On a constraint equation for the estimation of displacement rates in image sequences," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 1, pp. 13-30, January 1989.
- 28) Furht, "Multimedia systems: an overview," *IEEE Multimedia*, pp. 47-59, Spring 1994.
- 29) M. Subbarao, "Interpretation of image flow: a spatio-temporal approach," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 3, pp. 266-278, March 1989.