TREIZIÈME COLLOQUE GRETSI - JUAN-LES-PINS DU 16 AU 20 SEPTEMBRE 1991

Estimation de descripteurs de mouvement : Application au codage de séquences monoculaires et stéréoscopiques

H.NICOLAS, A. TAMTAOUI et C. LABIT
IRISA/INRIA, Campus de Beaulieu 35042 Rennes Cedex, France
Tel: 99 36 20 00 Telex: UNIRISA 950 473 F
FAX: 99 38 38 32

RÉSUMÉ

L'étude proposée ici tente de fournir une estimation compacte d'un champ de mouvement par descripteurs globaux sur des régions homogènes au sens d'un modèle de mouvement et qui soit efficace en reconstruction par compensation de mouvement. Le cadre méthodologique réside en les méthodes de minimisation par gradient (cas monoculaire) et son extension aux méthodes de minimisation sous contraintes (cas stéréoscopique). L'évaluation des résultats se fait en termes d'erreur de reconstruction et de qualité d'interprétation des champs de vecteurs mouvements apparents reconstruits.

ABSTRACT

Efficient motion estimators for image sequence coding by motion compensation are needed both for monocular and stereovision applications. The new proposed method presents an adaptive model for 2D global motion descriptors estimation. Then, we estimate motion parameters (using steepest descent method) without constraints for monocular application or with stereoscopic constraints for stereovision. Results are evaluated according to reconstruction errors and qualitative interpretation of apparent motion fields.

1. Introduction

L'acquisition des séquences se fait par l'intermédaire d'une ou deux caméras caractérisées par leurs centres optiques et leurs plan image (selon un modèle de projection perspective). L'estimation du mouvement apparent des objets (par l'intermédaire d'un modèle parametré de mouvement) permet de réduire l'information redondante existante entre les images et donc de diminuer le coût de transmission. Cette estimation diffère selon que l'on est dans un cadre de monovision ou de stéréovision :

- en monovision : estimation de paramètres de mouvement par la minimisation de l'erreur quadratique moyenne de reconstruction.
- en stéréovision : estimation conjointe de paramètres de mouvement gauche et droit sous contraintes stéréoscopique (géométrique et de cohérence) en minimisant l'erreur quadratique moyenne de reconstruction issues des deux vues.

Le cadre général dans ces deux cas est l'optimisation d'une fonctionnelle basée sur l'erreur de reconstruction, par des méthodes itératives du gradient. Dans le cas de la stéréovision on se restreint à un espace convexe de cohérence pour estimer les paramètres, et cela par des projections du gradient sur cet espace (méthodes itératives du gradient projeté [LUE84]).

Cet article se décompose en quatre parties :

1. description du modèle de mouvement utilisé dans les deux cas d'applications (monovision et stéréovision).
2. développement d'une méthode d'estimation des paramètres résolvant le problème des minima locaux.
3. spécifications de contraintes de cohérence sur les mouvements apparents en stéréovision, mise en correspondance des régions et estimation conjointe des paramètres de mouvement associés.
4. conclusion et résultats obtenus sur séquences réelles. Les résultats de cette étude sont comparés avec les methodes classiques de codage avec compensation de mouvement en termes de reconstruction et de coût [WAL84].

2. Modèles de mouvement 2D

De nombreux travaux ont permis d'estimer un champ dense de déplacement permettant la reconstruction de l'image $t+1$ à partir de l'image t, par compensation de mouvement [WAL84] . S'il est possible d'obtenir par ces méthodes une erreur de reconstruction faible, le volume d'information (deux paramètres par pixel correspondant aux deux composantes de mouvement translationnel) est considérable. Pour permettre une utilisation plus efficace de l'information de mouvement dans un schéma de transmission, il est nécessaire de définir un modèle permettant une description globale (ou par région) du mouvement [ADI85]. Les modèles généralement utilisés (en se limitant au 1er ordre) sont (avec : $\vec{T}$ vecteur de paramètres, $\vec{\omega}(dx, dy)$ vecteur déplacement au point (x, y) et $G(x_g, y_g)$ centre de gravité 2D d'un "objet" ou région) :

- modèle de mouvement constant (2 paramètres : (t_x, t_y))

$$dx = t_x$$
$$dy = t_y$$

- modèle linéaire simplifié (4 paramètres : (t_x, t_y, k, θ))

$$dx = t_x + k(x - x_g) - \theta(y - y_g)$$
$$dy = t_y + k(y - y_g) + \theta(x - x_g)$$

- modèle linéaire (6 paramètres : (t_x, t_y, a, b, c, d)) .

$$dx = t_x + a(x - x_g) + b(y - y_g)$$
$$dy = t_y + c(x - x_g) + d(y - y_g)$$

En codage, il faut néanmoins garder présent à l'esprit que l'on cherche à minimiser le volume d'information à transmettre et donc à minimiser le nombre de paramètres nécessaire à la reconstruction des images. Le problème est donc de

trouver le meilleur compromis possible entre la qualité de la reconstruction et le nombre de paramètres pour une région donnée. Dans le cadre de cette étude, nous avons utilisé le modèle linéaire simplifié qui semble représenter le meilleur compromis dans de nombreux cas (le modèle constant donne de médiocres résultats tandis que le modèle linéaire complet n'apporte que peu d'amélioration par rapport au modèle linéaire simplifié).

3. Estimation en monovision

La méthode itérative proposée ici se décompose en trois phases [NIC91] (cf figure 2) :

- *La phase d'estimation : méthode d'optimisation par gradient*

Cette méthode permet de converger - à partir d'une valeur initiale $\vec{X^o}$- vers le minimum local le plus proche.
Le processus itératif (en utilisant le modèle linéaire simplifié) est défini par:

$$\vec{X_R}^{(j+1)} = \vec{X_R}^{(j)} - \frac{1}{2.N} \cdot \sum_{\vec{p} \epsilon R} \epsilon . \begin{pmatrix} \frac{\partial}{\partial t_x}(DFD^2(\vec{p},\vec{d}) \\ \frac{\partial}{\partial t_y}(DFD^2(\vec{p},\vec{d}) \\ \frac{\partial}{\partial k}(DFD^2(\vec{p},\vec{d}) \\ \frac{\partial}{\partial \theta}(DFD^2(\vec{p},\vec{d}) \end{pmatrix} \quad (1)$$

ϵ étant une matrice diagonale de gains, N la taille de la région, $DFD = f(\vec{p},t+1) - f(\vec{p}-\vec{d},t)$ la différence inter-image déplacée, ∇_x et ∇_y les gradients spatiaux au point (x,y).
Après développement:

$$\vec{X_R}^{(j+1)} = \vec{X_R}^{(j)} - \epsilon . \vec{grad}(DFD^2) \quad (2)$$

où

$$\vec{grad}(DFD^2) = \frac{1}{N} \sum_{\vec{p} \epsilon R} DFD \begin{pmatrix} \nabla_x \\ \nabla_y \\ (x-x_g)\nabla_x + (y-y_g)\nabla_y \\ -(y-y_g)\nabla_x + (x-x_g)\nabla_y \end{pmatrix}$$

- *La phase de relaxation*

La méthode du gradient ne permettant que de converger vers des minima locaux, il est nécessaire de contourner ce problème en établissant une coopération entre régions voisines appartenant au même objet physique . Cette coopération revient à réinitialiser le processus itératif par le vecteur de paramètre d'une des régions voisines minimisant la DFD^2 (ce qui permet dans de nombreux cas de ressortir des minima locaux successifs). Chaque vecteur est comparé à ceux des régions voisines et cela jusqu'à stabilisation du processus (cf figure 1).

- *La phase de division*

Après stabilisation des deux phases précédentes une segmentation quadtree est mise en oeuvre. Une région est divisée en quatre sous-régions si l'erreur de reconstruction reste supérieure à un seuil.

4. Application à la stéréoscopie télévisuelle (TV3D : Télévision en relief)

Quelques travaux ont étudié des schémas de compensation par disparité pour interpoler une vue connaissant l'autre. Ces schémas nécessitent une acquisition des images dans une condition stable d'eclairement [JEN91].

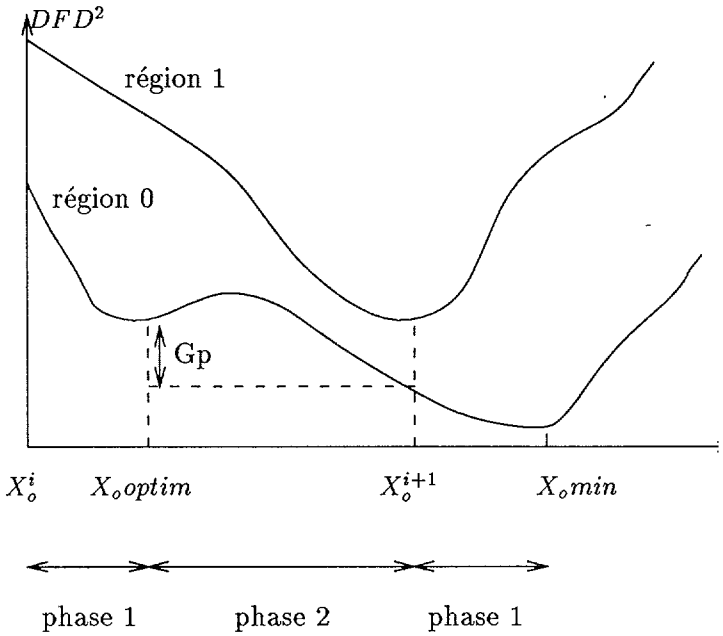

Figure 1 :**Principe de la phase de relaxation**

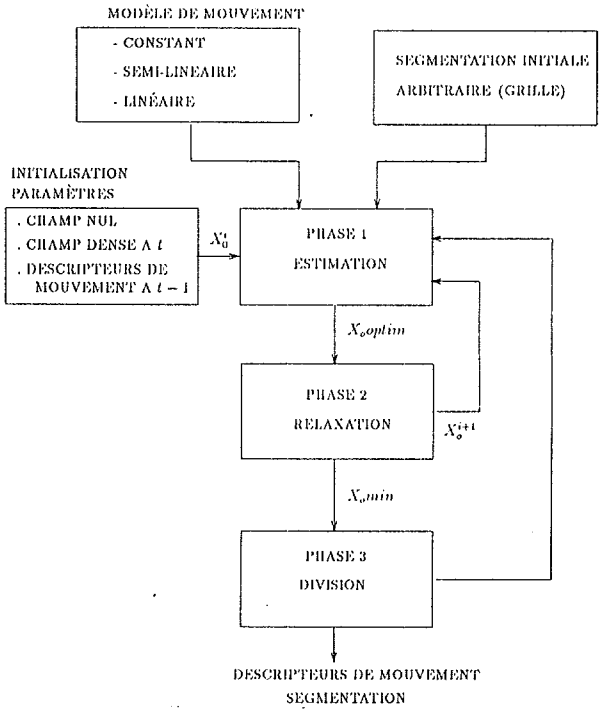

Figure 2 : Schéma de l'étude

Dans le cadre d'une coopération stéréo-mouvement [SNY91], nous nous sommes intéressés à des schémas coopératifs d'estimation de mouvement entre les deux séquences d'images issues de la paire stéréoscopiques. Les algorithmes de minimisation utilisés sont basés sur des techniques de gradient projeté sur un espace convexe qui verifie les contraintes stéréoscopiques. Il s'agit d'estimer conjointement deux vecteurs de descripteurs de mouvements apparents gauche et droit, liés par la fonction disparité, sous contraintes stéréoscopiques. Ces derniers schémas sont avérés bien adaptés pour résoudre efficacement les phases

d'estimations. Deux sortes de contraintes ont été étudiées : contrainte géométrique (correspondance des lignes épipolaires) et contraintes de cohérences (les composantes du mouvement apparent de la séquence gauche peuvent être reliées à celles de la séquence droite par une équation lineaire). Les attributs s'ils sont choisis trés locaux (au niveau pixel) dans ce schéma d'estimation sous contraintes, rend le calcul du champ de disparité trop coûteux. Il est préférable de travailler avec une interprétation plus globale du mouvement apparent utilisant un espace de paramètres (descripteurs) plus globaux.

• Cohérence de mouvement (Equations de cohérences)

Les projections images issues des deux caméras sont supposés être contenus dans un même plan (deux caméras coplanaires, i.e $z_l = z_r$ pour tout objet). En plus de la contrainte géométrique des lignes épipolaires, des contraintes de cohérence sur les composantes de mouvement ont été développées utilisant deux hypothèses sur la profondeur des objets : profondeur peu variable par rapport à la fréquence d'echantillonnage des caméras ($\dot{z}_r = \dot{z}_l = 0$)et profondeur variable ($z_l = z_r$). La deuxième hypothèse est plus réaliste vis à vis des scènes naturelles télévisuelles.

Soit (dx^l, dy^l) et (dx^r, dy^r) deux vecteurs vitesses apparents gauche et droit liés par la fonction disparité, une relation de cohérence suivante dans le cas ($z_l = z_r$) peut être établie :

$$Adx^l + Bdy^l + Cdx^r + Ddy^r = 0 \qquad (3)$$

où A,B,C et D sont des paramètres qui dépendent des paramètres de calibration des caméras. En utilisant les descripteurs de mouvement decrit ci-dessus (modèle à 4 paramètres), on peut ecrire pour les images gauche et droite :

$$\begin{cases} dx &= t_x + k(x - x_g) - \theta(y - y_g) \\ dy &= t_y + k(y - y_g) - \theta(x - x_g) \end{cases} \qquad (4)$$

l'equation (4) conduit aux équations de cohérences suivantes en posant par hypothèse que le centre de gravité x_g^l est le correspondant de x_g^r (ce qui est toujours assuré dans le cas continu) :

$$\begin{cases} At_x^l + Bt_y^l + Ct_x^r + Dt_y^r &= 0 \\ Ak^l + B\theta^l + \alpha_1 k^r + \beta_1 k^r &= 0 \\ Bk^l + A\theta^l + \alpha_2 k^r + \beta_2 k^r &= 0 \end{cases} \qquad (5)$$

que l'on peut ecrire de manière matricielle sous la forme : $CX = 0$ où $X = (t_x^l, t_y^l, k^l, \theta^l, t_x^r, t_y^r, k^r, \theta^r)$ et $\mathcal{C}$ est la matrice de cohérence. $\alpha_1, \alpha_2, \beta_1$ et β_2 dépendent aussi des paramètres de calibration. On appelera dans la suite Ω l'espace convexe qui vérifie ces équations (5). Sous l'hypothèse de la correspondance des centres de gravités, deux régions mises en correspondance doivent vérifier les équations de cohérences ci-dessus. Ceci peut être défini en chaque pixel mais reste valable dans le cas des régions à profondeur constante dans l'espace (Les objets sont décomposés en facettes parallèles au plan image).

• Mise en correspondance de regions

Supposons la connaissance à priori de descripteurs initiaux sur les deux vues de la paire stéréoscopique (par exemple des estimations issues d'une analyse monoculaires de chaque vue). Soit une région R_r de l'image droite que l'on cherche à mettre en correspondance avec une région R_l de l'image gauche. Deux régions gauche et droite se correspondent si leurs mouvements d'ensemble minimisent les équations de cohérences (5). Le problème de mise en correspondence de régions revient à minimiser $||CX||$ le long de la ligne épipolaire gauche qui corresponde à la ligne épipolaire droite contenant le centre de gravité de la région droite. A l'issue de la mise en correspondance de toutes les régions de l'image droite, une erreur moyenne de cohérence non nulle sur toute la paire d'images est calculée par

$$\mathcal{E}_{moy} = \frac{1}{N_r} \sum_{i=1}^{N_r} \mathcal{E}_i$$

où $\mathcal{E}_i = ||CX_i||$ est l'erreur correspondante à la région droite i et N_r le nombre de régions dans l'image droite.

• Estimation conjointe des paramètres

Le but de cette étude est d'estimer conjointement des descripteurs (gauche/droit) cohérents, au sens des équations de cohérences (5), en minimisant l'erreur quadratique moyenne aprés reconstruction par compensation de mouvement. Ainsi, pour deux régions gauche et droite mises en correspondance, il est possible d'estimer simultanément leurs descripteurs de mouvement en se restreignant à l'espace Ω de cohérence. Il s'agit de trouver le $arg\,\min_{X \in \Omega} E(f(X))$ par des estimations itératives.

La méthode d'estimation utilisée s'apparente à celle introduite au paragraphe 3 mais en prenant comme entrée à minimiser, la fonctionnelle $E(f(T)) = E(DFD_l^2 + DFD_r^2)$ en chaque région. La méthode de minimisation choisie est une méthode itérative de descente selon le gradient projeté sur l'espace de cohérence Ω [LUE84]. L'équation itérative du gradient projeté s'ecrit :

$$X_k = X_{k-1} - \epsilon P \vec{grad} E(f(X_{k-1}))$$

où $P = I - \mathcal{C}^T(\mathcal{C}\mathcal{C}^T)^{-1}\mathcal{C}$ est la matrice de projection sur l'espace de cohérence Ω et ϵ un gain adaptatif [TAM91].

L'erreur de cohérence moyenne est entièrement dépendante des descripteurs initiaux, et reste constante tout au long de l'estimation conjointe. Pour faire decroitre cette erreur, une succession de mise en correspondance des régions et projections est opérée jusqu'a stabilisation relative de celle-ci.

5. Conclusion

La méthode d'estimation du mouvement décrit dans cet article permet d'obtenir une bonne qualité de reconstruction par compensation de mouvement tant en monoculaire qu'en stéréovision (voir resultats) et cela en utilisant un faible volume d'information comparativement, les méthodes classique du type pel-récursif nécéssitent deux paramètres par pixel.

Pour réduire l'erreur résiduelle, plusieurs approches sont envisagées :

• prise en compte du voisinage temporel dans la phase de relaxation.

• modification du modèle de mouvement pour permettre la correction des erreurs dues à la variation d'illumination (images successives en monoculaire et images gauches et droites en stéréovision) [MOL91].

• amélioration de la mise en correspondance (diminuer l'erreur de cohérence $\mathcal{E}_{moy}$) par une méthode de division-fusion des régions (dans le cas stéréoscopique).

Bibliographie

[ADI85] ADIV G. Determining three-dimensional motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 7:pp 384–401, Jul. 1985.

[JEN91] JENKIN M.R. Techniques for disparity measurement. *CVGIP Image Understanding*, 53(1):14–30, January 1991.

[LUE84] LUENBERG. *Linear and nonlinear programming.* Adison-Wesley, 1984.

[MOL91] MOLONEY C.R. , DUBOIS E., Estimation of motion fields from image sequences with illumination variation. *Proc. of ICASSP'91, TORONTO*, Vol. No.4:pp 2425–2428, 1991.

[NIC91] NICOLAS H. , LABIT C. Global motion identification for image sequence analysis and coding. *Proc. PCS'91 TOKYO*, sept. 1991.

[SNY91] SNYDER M.A. The P-field : a computational model for binocular motion processing. *CVPR'91 USA*, June 1991.

[TAM91] TAMTAOUI A. LABIT C. Coherent disparity and motion compensation in 3 DTV image sequence coding. *Proc. ICASSP'91 Canada*, May 1991.

[WAL84] WALKER D.R. , RAO K.R. Improved pel-recursive motion compensation. *IEEE Transactions on Communications*, Vol.32, No.10:pp 1128–1134, Oct. 1984.

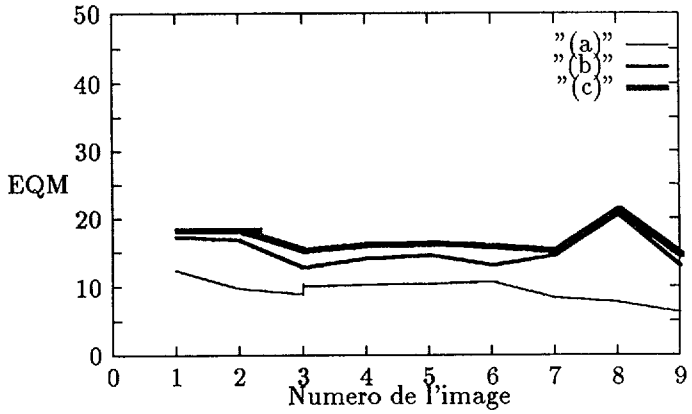

Figure 3 : Erreur quadratique moyenne entre les sequences reconstruite et originale. (a) : Walker-Rao, (b) : descripteurs en monoculaire, (c) : descripteurs en stéréoscopie

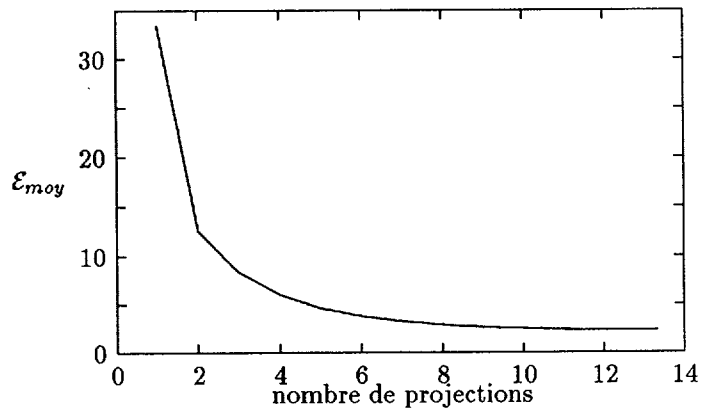

Figure 4 : Erreur moyenne de cohérence sur les descripteurs projetés

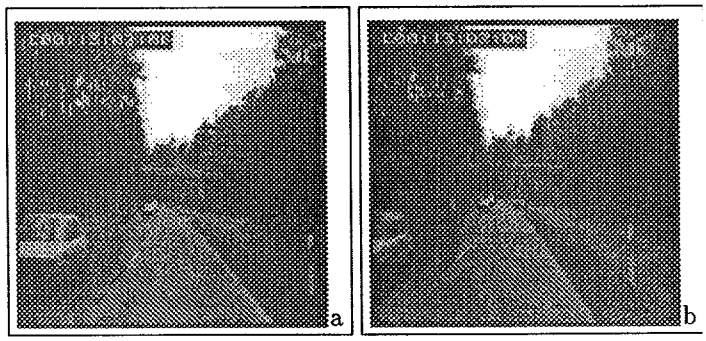

Figure 5 : Images originales : (a) gauche (b) droite

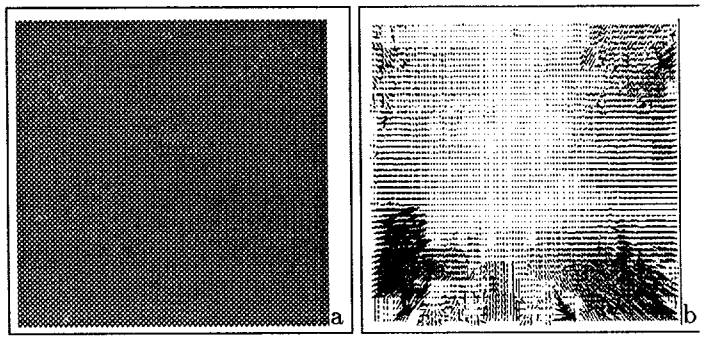

Figure 6 : cas monovision : (a) Image d'erreur (x2) (b) champ de mouvement estimé

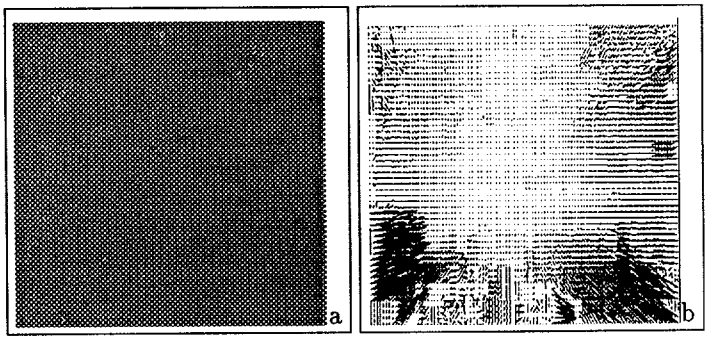

Figure 7 : cas stéréréovision : (a) Image d'erreur (x2) (b) champ de mouvement estimé

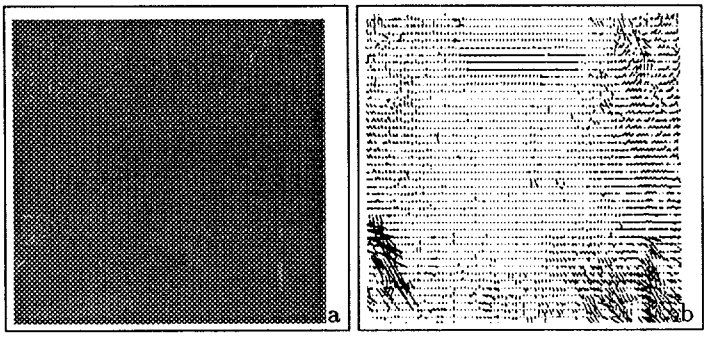

Figure 8 : cas Walker-Rao : (a) Image d'erreur (x2) (b) champ de mouvement estimé