

PRISE EN COMPTE DES DISCONTINUITES DANS L'ESTIMATION DU MOUVEMENT PAR LES CHAMPS DE MARKOV COMBINES A L'AMPLITUDE DE FOURIER

A.Dipanda*, M.kardouchi*, L.Legrand*

*Faculté des sciences Mirande,
Liesib-Département d'informatique
BP : 138, 21000 Dijon cedex

*Faculté de médecine,
laboratoire d'informatique médicale
7 bd Jeanne d'arc, 21033 Dijon cedex

ABSTRACT

The goal of this study is to estimate the motion between two successive images in a sequence. In this paper, we present an approach of the regularization problem which enables us to take into account motion discontinuities. This approach is based on the theory of Markov random fields and the Fourier analysis.

1. INTRODUCTION

L'étude de séquences d'images numériques introduit un champ spécifique d'investigation en analyse d'images, à savoir l'extraction d'informations sur le mouvement. Ces informations peuvent concerner : l'estimation du mouvement, la détection de mouvement, la segmentation au sens du mouvement, etc.

Dans cet article nous nous intéressons à l'estimation du champ de déplacement entre deux images d'une séquence donnée.

Pour estimer le mouvement, plusieurs méthodes existent parmi lesquelles :

- l'approche différentielle [1] : elle travaille dans le domaine direct. Elle est basée sur une ou plusieurs contraintes locales reliant les dérivées spatiale et temporelle du signal au mouvement apparent. L'hypothèse de base de cette méthode est l'invariance de la luminance d'un point lors d'un déplacement dans l'espace (x,y,t) constitué de deux images successives.
- l'approche de mise en correspondance [2] : ce type de méthode est basé sur la similarité entre blocs. Pour chacun des blocs, un vecteur de déplacement unique est alors estimé par la recherche du bloc le plus proche en terme de similarité ou de corrélation dans l'image suivante.
- l'approche par transformée : elle utilise le domaine de la transformée (principalement la transformée de Fourier [3] ou celle de Gabor [4]). Le mouvement se traduit par des modifications des coefficients de la transformée (déphasage de la transformée de Fourier pour un mouvement de translation par exemple).

- l'approche statistique par modélisation markovienne[5][6] : comme dans l'approche différentielle, on travaille dans le domaine direct ; elle est basée sur un traitement statistique des images et la théorie des champs de Markov.

Quelle que soit l'approche retenue, l'estimation d'un champ dense de déplacement introduit une prise en compte d'un certain voisinage. Il est alors supposé que le champ de déplacement possède une certaine propriété de continuité. Cependant, cette propriété est violée au voisinage des contours des objets en mouvement. Un des problèmes posés pour l'estimation du champ de déplacement est de savoir gérer les discontinuités susceptibles d'apparaître.

Dans cet article, nous proposons une méthode d'estimation du mouvement qui permet le traitement des discontinuités. Elle utilise un modèle markovien au sein duquel est intégré l'amplitude locale du mouvement, qui est une information obtenue par l'analyse de Fourier. Le modèle ainsi obtenu permet de faire un lissage adaptatif des vecteurs déplacement dans l'image.

Dans le prochain chapitre, nous présentons les principaux aspects de la modélisation markovienne en traitement d'images, en particulier pour l'estimation du mouvement. Dans le chapitre 3, nous décrivons les conditions à vérifier par une contrainte de régularisation pour la préservation des discontinuités du mouvement. Ensuite dans le chapitre 4, nous définissons l'amplitude locale du mouvement. Le chapitre 5 est consacré au modèle que nous avons développé. Enfin, dans les chapitres 6 et 7, nous présentons des résultats et nous donnons les conclusions sur la méthode proposée.

2. MODELISATION MARKOVIENNE ET ESTIMATION DU MOUVEMENT

2.1 Modélisation markovienne et estimation bayésienne en analyse d'images

Soit $\Omega = \{s = (i, j) \mid (1 \leq i \leq M \text{ et } 1 \leq j \leq M)\}$ une grille 2D de taille $M \times M = n$. On note par $V = \{v(s), s \in \Omega\}$ une famille arbitraire de voisinages locaux de Ω . Par exemple, le voisinage d'ordre un d'un pixel s , correspond à l'ensemble des pixels premier voisins de celui-ci (fig : 1) :

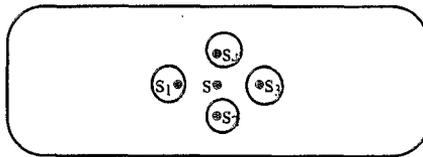


fig 1 : voisinage d'ordre 1.

○ : pixel voisin de s.

On note par C_s l'ensemble des cliques associées à un pixel s . Pour un voisinage d'ordre un, l'ensemble des cliques (fig : 2) est égal à :

$$C_s = \{ \{s_1\}, \{s_2\}, \{s_3\}, \{s_4\}, \{s, s_1\}, \{s, s_2\}, \{s, s_3\}, \{s, s_4\} \}$$

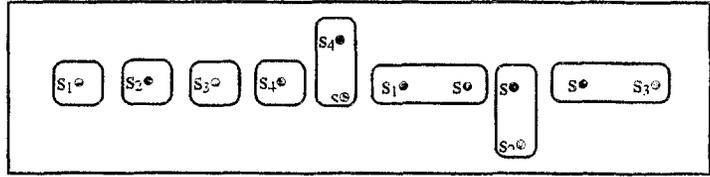


fig 2 : Cliques associées au voisinage d'ordre un

Un champ de variables aléatoires $I = (I_{s_1}, I_{s_2}, \dots, I_{s_n})$ défini sur une grille S est un champ de Markov relativement à un système de voisinage V si et seulement si :

$$\Pr(I_{s_k} = i_{s_k} / I_{s_l} = i_{s_l}, s_l \in \Omega - \{s_k\}) = \Pr(I_{s_k} = i_{s_k} / I_{s_l} = i_{s_l}, s_l \in \nu(s_k) - \{s_k\}) \quad (2.1.1)$$

Où $i = (i_{s_1}, i_{s_2}, \dots, i_{s_n})$ est une réalisation de I .

Ceci signifie que la connaissance d'un voisinage local d'un site s donné est suffisante pour calculer la probabilité en celui-ci. Cette propriété de dépendance locale est rendue exploitable grâce au théorème d'Hammersley-Clifford [7] qui démontre qu'un champ aléatoire défini sur un réseau est un champ de Markov si et seulement si sa distribution de probabilité est une distribution de Gibbs définie par :

$$\Pr(I=i) = (1/Z) \exp(-U(I=i)) \quad (2.1.2)$$

Où Z est une constante de normalisation appelée fonction de partition.
 $U(I=i)$ est la somme de fonctions locales appelées fonctions énergies.

La modélisation markovienne en analyse d'images s'appuie sur deux champs de variables aléatoires distincts :

- le champ des observations : il est formé par l'ensemble des données sur lesquels s'appuie l'analyse.
- le champ des primitives : qui est l'information non observable que l'on cherche à extraire des observations.

Ils sont supposés former deux champs de variables aléatoires discrets que nous noterons respectivement O et P . Ils sont définis sur deux grilles rectangulaires R et S (identiques ou non) :

$$O = \{O_{r_i}, r_i \in R\} \quad P = \{P_{s_j}, s_j \in S\} \quad (2.1.3)$$

où r_i et s_j désignent des sites des grilles R et S .

On notera par $o = \{o_{r_i}, r_i \in R\}$ et $p = \{p_{s_j}, s_j \in S\}$ une réalisation quelconque des champs aléatoires O et P .

Le critère d'estimation du champ de primitives le plus souvent retenu est le Maximum A Posteriori (M.A.P) : étant donné le champ d'observation o , on cherche le champ de primitives $p(\text{estimé})$ le plus probable au sens de la distribution a posteriori :

$$p(\text{estimé}) = \underset{p}{\text{Arg Max}} \Pr (P = p / O = o). \quad (2.1.4)$$

En utilisant la règle de Bayes, à un facteur constant près, on obtient :

$$p(\text{estimé}) = \underset{p}{\text{Arg Max}} \Pr (O = o / P = p) \Pr (P = p) \quad (2.1.5)$$

Le théorème d'Hammersley-Clifford permet de ramener le problème d'estimation de primitives à la minimisation d'une fonction énergie U :

$$p(\text{estimé}) = \min_p U \quad (2.1.6)$$

La recherche d'un tel minimum est un problème d'optimisation. Une solution approchée peut être obtenue soit, par des algorithmes déterministes tel que l'I.C.M. (Itérative Conditional Mode) [7], soit, par des algorithmes stochastiques de type 'recuit simulé' [5]. Les algorithmes stochastiques convergent vers le minimum global, mais restent coûteux en temps de calcul. Par contre, les algorithmes déterministes convergent beaucoup plus rapidement, mais présentent le défaut de pouvoir se faire piéger par un minimum local. Une bonne initialisation de départ du champ de primitives est donc nécessaire.

2.2 Estimation du mouvement

Dans le cadre de l'estimation du mouvement, les observations sont constituées par deux images successives de la séquence entre lesquelles le mouvement sera estimé, et les primitives sont représentées par les déplacements calculés sur chaque pixel. L'hypothèse émise sur la fonction énergie suppose que la luminance d'un pixel est invariante entre deux images successives. Ainsi si on note par $f(s,t)$ le niveau de gris à l'instant t correspondant au pixel s et ds le déplacement de s entre t et $t+dt$, l'hypothèse décrite ci-dessus se traduit par :

$$\begin{aligned} DFD(s) &= (f(s+ds, t+dt) - f(s, t)) = 0 \\ DFD &\text{ signifiant 'displaced frame difference'}. \end{aligned} \quad (2.2.1)$$

Ainsi, l'estimation du champ de déplacement en un pixel s revient à trouver le déplacement ds tel que $DFD(s) = 0$, ce qui équivaut à minimiser $DFD(s)^2$. La fonction énergie s'écrit alors :

$$U_{DFD}(ds, f_t, f_{t+dt}) = \sum_s U_{DFD}^s(ds, f_t, f_{t+dt}) \quad (2.2.2)$$

$$\text{avec } U_{DFD}^s(ds, f_t, f_{t+dt}) = (f(s+ds, t+dt) - f(s, t))^2 \quad (2.2.3)$$

Le critère d'estimation du champ de déplacement est obtenu pour chaque pixel s en minimisant l'équation (2.2.3) :

$$\begin{aligned} ds(\text{estimé}) &= \text{Min}_{ds \in D} U_{DFD}^s(ds, f_t, f_{t+dt}) \\ &\text{où } D \subset \mathbb{Z}^2 \end{aligned} \quad (2.2.4)$$

Dans la majorité des cas, les phénomènes qui président à la formation des observations s'accompagnent d'une dramatique perte d'informations intervenant aux différentes étapes du processus. On est en présence d'un problème *mal posé*, la seule connaissance des données ne suffit pas pour assurer l'existence, l'unicité et la stabilité de la solution. Pour contourner cette difficulté, il est nécessaire d'émettre des hypothèses a priori sur les primitives. On est ainsi amené à introduire une contrainte qualifiée de contrainte de régularisation (ou de lissage).

Ainsi, l'estimation du vecteur déplacement en un pixel s est obtenu par l'équation suivante :

$$ds(\text{estimé}) = \text{Arg Min}_{ds \in D} (U_{DFD}^s(ds, f_t, f_{t+dt}) + U_R^s(ds)) \quad (2.2.5)$$

Un modèle de régularisation classique est donné par Tikhonov [8]. Il est défini par :

$$U_R^s(ds) = \sum_{\{s,t_i\} \in C_s} \beta^2 \|ds - dt_i\|^2 = \sum_{\{s,n\} \in C_s} \phi_s(\|ds - dt_i\|^2) \quad (2.2.6)$$

avec $\phi_s(\|ds - dt_i\|^2) = \beta^2 \|ds - dt_i\|^2$
 et $\| \cdot \|$ désigne la norme euclidienne.

β^2 est un paramètre de pondération entre $U_{D.F.D}^s(ds, f_i, f_{i+d_i})$ et $U_R^s(ds)$.

$t_i : i = 1, 2, 3, 4$ étant l'ensemble des premiers voisins de s .

La convexité du modèle de régularisation donné par Tikhonov permet aux algorithmes déterministes de converger vers le minimum global. Cependant, la nature isotropique de ce modèle ne permet pas de gérer correctement les discontinuités du mouvement susceptibles d'apparaître au niveau des contours des objets en mouvement.

3. MODELISATION DES DISCONTINUITES

Pour qu'une contrainte de régularisation préserve les discontinuités, elle doit appartenir à la classe des "phi-fonction" qui vérifient [9] :

$$\begin{aligned} \lim_{x \rightarrow 0^+} (\phi'(x) / 2x) &= M \quad \text{avec } M \text{ finie} \\ \lim_{x \rightarrow +\infty} (\phi'(x) / 2x) &= 0 \end{aligned} \quad (3.1)$$

$(\phi'(x) / 2x)$ est continue et strictement décroissante dans $]0, +\infty[$.

Plusieurs "phi-fonction" sont proposées dans la littérature [9][10]. Cependant la plupart d'entre elles ne sont pas convexes, et par conséquent, sont difficiles à minimiser. Nous proposons dans cet article de construire une approximation localement convexe de la phi-fonction définie par $\phi(x) = x^2 / (p^2 + p^2 x^2)$ en utilisant l'analyse de Fourier.

4. ANALYSE DE FOURIER ET AMPLITUDE LOCALE DU MOUVEMENT

Soit $f(x)$ une fonction périodique de période T_0 ($T_0 = 1/v_0$) où x est une variable temporelle ou spatiale. Si $f(x)$ satisfait les conditions de Dirichlet [11], alors $f(x)$ peut être représentée par une série réelle infinie de fonctions trigonométriques. En se limitant à une somme finie et en discrétisant $f(x)$, on obtient :

$$f(n) \approx \sum_{k=0}^{N-1} (a_k \cos(2\pi n k v_0) + b_k \sin(2\pi n k v_0)) \quad (4.1)$$

où a_k et b_k sont les coefficients de Fourier :

$$a_k = \sum_{n=0}^{N-1} f(n) \cos(2\pi n k v_0) \quad \text{et} \quad b_k = \sum_{n=0}^{N-1} f(n) \sin(2\pi n k v_0) \quad (4.2)$$

cette fonction peut encore s'écrire :

$$f(n) = \sum_{k=0}^{N-1} C_k \cos(2\pi n k v_0 - \Phi_k) \quad (4.3)$$

$$\text{avec} \quad C_k = \sqrt{a_k^2 + b_k^2} \quad \text{et} \quad \Phi_k = \text{Arctg}(-b_k / a_k) \quad (4.4)$$

C_k et Φ_k sont respectivement l'amplitude et la phase de la $k^{\text{ième}}$ harmonique.

Considérons maintenant une séquence de N images. Soit $f(s,n)$ le niveau de gris du pixel s de la $n^{\text{ième}}$ image. Le signal $f(s,n)$ ($n=1,2,\dots,N$) est de période égale à N et de fréquence égale à $(1/N)=v_0$. On connaît l'évolution en niveau de gris $f(s,n)$ pour chacun des pixels s le long de la séquence. En se limitant à une somme finie (nombre d'images dans la séquence) de toutes les fréquences possibles, on obtient :

$$f(s,n) \approx \sum_{k=0}^{N-1} (a_k(s) \cos(2\pi n k v_0) + b_k(s) \sin(2\pi n k v_0)) \quad (4.5)$$

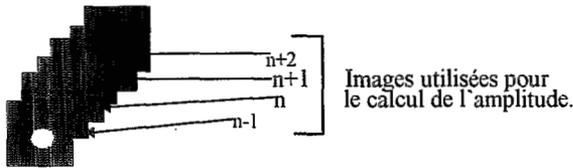
$$\text{avec} \quad a_k(s) = \sum_{n=0}^{N-1} f(s,n) \cos(2\pi n k v_0) \quad (4.6)$$

$$b_k(s) = \sum_{n=0}^{N-1} f(s,n) \sin(2\pi n k v_0) \quad (4.7)$$

$$c_k(s) = \sqrt{a_k(s)^2 + b_k(s)^2} \quad (4.8)$$

Pour un pixel s donné, le coefficient de Fourier $c_k(s)$ le plus important correspond à la fréquence égale à $1/N$ qui est la fréquence de la première harmonique, et par conséquent, seuls les coefficients $a_1(s)$ et $b_1(s)$ de la première harmonique sont calculés. L'intérêt de l'amplitude de Fourier ainsi calculée, est qu'elle nous donne une information sur l'amplitude du mouvement en chaque pixel le long de la séquence.

L'estimation du mouvement porte sur deux images successives, n et $n+1$. Pour calculer l'amplitude locale du mouvement entre ces deux images, nous nous limitons à quatre images $n-1$, n , $n+1$, et $n+2$ de la séquence. Ceci permet de limiter les "perturbations" induites par les images éloignées de n et $n+1$. On obtient ainsi :



On obtient ainsi :

$$a_1(s) = \sum_{n'=n-1}^{n'+n+2} f(s,n') \cos(2\pi n' v_0) \quad (4.9)$$

$$b_1(s) = \sum_{n'=n-1}^{n'+n+2} f(s,n') \sin(2\pi n' v_0) \quad (4.10)$$

L'amplitude locale du mouvement du mouvement en chaque pixel s s'écrit alors :

$$C_1(s) = \sqrt{a_1(s)^2 + b_1(s)^2} \quad (4.11)$$

5. MODELE DE REGULARISATION AVEC L'AMPLITUDE LOCALE DU MOUVEMENT

Considérons la "phi-fonction" définie par :

$$\phi^s(ds, p^2, q^2) = \sum_{\{s, ti\} \in C_s} \|ds - dt_i\|^2 / (p^2 + q^2 \|ds - dt_i\|^2) \quad (5.1)$$

et où p et q sont deux paramètres constants.

Nous proposons d'introduire l'amplitude locale du mouvement dans la contrainte de régularisation (5.1) à fin de la rendre convexe. Nous avons émis l'hypothèse qu'en chaque pixel s , cette amplitude est proportionnelle au module du vecteur déplacement. Ainsi si on note par A_s et A_{ti} les amplitudes locales du mouvement des pixels s et $ti_{i=1,2,3,4}$, (les ti étant les quatre premiers voisins de s) et $ds(s_x, s_y)$ et $dt_i(ti_x, ti_y)$ leurs déplacements respectifs entre n et $n+1$, on peut alors écrire :

$$A_s \approx \gamma \|ds(s_x, s_y)\| \quad \text{et} \quad A_{ti} \approx \gamma \|dt_i(ti_x, ti_y)\| \quad (5.2)$$

où $\| \cdot \|$ désigne la norme euclidienne.

d'où :

$$\begin{aligned} \|ds - dt_i\| &= \sqrt{(s_x - ti_x)^2 + (s_y - ti_y)^2 - 2s_x ti_x + 2s_y ti_y} \\ &\approx \sqrt{[(A_s)^2 / \gamma^2 + (A_{ti})^2 / \gamma^2 - 2(s_x ti_x + s_y ti_y)]} \\ &= \sqrt{[(A_s)^2 / \gamma^2 + (A_{ti})^2 / \gamma^2 - 2(ds \cdot dt_i)]} \\ &= \sqrt{[(A_s)^2 / \gamma^2 + (A_{ti})^2 / \gamma^2 - 2 \|d_s\| \|d_{ti}\| \cos(ds, dt_i)]} \end{aligned} \quad (5.3)$$

En Considérant la relation (5.3), on peut remarquer que celle-ci peut s'écrire différemment, suivant que les pixels voisins s et ti appartiennent à l'objet ou pas :

- si les pixels s et ti appartiennent à l'objet en mouvement, les directions de leurs vecteurs déplacement sont très voisines, et par conséquent $\cos(ds, dt_i) \approx 1$. L'égalité (5.3) devient : $\|ds - dt_i\|^2 = [(A_s - A_{ti})^2 / \gamma^2]$.
- si s et ti appartiennent au fond, les amplitudes de mouvement A_s et A_{ti} sont nulles, on a alors $\|ds\| = 0$ et $\|dt_i\| = 0$. L'égalité (5.3) devient : $\|ds - dt_i\|^2 = 0$.
- si s appartient à l'objet et ti au fond, alors on a $A_s \neq 0$ et $A_{ti} = 0$. L'égalité (5.3) devient : $\|ds - dt_i\|^2 = (A_s)^2 / \gamma^2 = [(A_s - A_{ti})^2 / \gamma^2]$.

A partir de l'étude ci-dessus, on montre que : $\|ds - dt_i\|^2 = [(A_s - A_{ti})^2 / \gamma^2]$. En remplaçant $\|ds - dt_i\|^2$ dans le dénominateur de l'équation (5.1), on obtient :

$$\phi^s(ds, p^2, q^2) \cong \sum_{\{s, ti\} \in C_s} \|ds - dt_i\|^2 / (p^2 + q^2 ((A_s - A_{ti})^2 / \gamma^2)) \quad (5.4)$$

Cette nouvelle contrainte de lissage présente les avantages suivants :

- d'une part, elle a un caractère localement convexe : la quantité $1 / (p^2 + q^2 ((A_s - A_{ti})^2 / \gamma^2))$ représente le paramètre de pondération des interactions entre les pixels s et ti . Cette quantité reste constante lors de la minimisation par l'I.C.M. par rapport à ds au point s . De part l'aspect local du traitement effectué par l'I.C.M., $\phi^s(ds, p^2, q^2)$ sera donc perçue par celui-ci comme une fonction convexe.

◦ d'autre part, elle permet de faire un lissage adaptatif dans l'estimation du champ de déplacement :

- ◊ pour deux pixels voisins appartenant à l'objet, A_s et A_n sont très voisins, et par conséquent, $p^2 \rightarrow q^2 ((A_s - A_n)^2 \cdot \gamma^2)$, et $\|ds - dt\|^2 : (p^2 + q^2 ((A_s - A_n)^2 \cdot \gamma^2))$ tend vers $\|ds - dt\|^2 / p^2$. On retrouve la contrainte de lissage introduite par Tikhonov, le paramètre $1/p^2$ correspond au paramètre β^2 dans l'équation (2.2.5).
- ◊ au niveau des contours, deux pixels voisins, n'appartenant pas tous les deux à l'objet, ont des amplitudes de mouvement très différentes, ce qui donne $p^2 \rightarrow q^2 ((A_s - A_n)^2 \cdot \gamma^2)$. La contrainte de lissage devient : $\|ds - dt\|^2 / q^2 ((A_s - A_n)^2 \cdot \gamma^2)$. Ceci va donc atténuer fortement le lissage entre ces deux pixels.

6. RESULTATS

Nous avons appliqué la méthode sur une séquence d'images synthétiques puis sur deux séquences d'images réelles. La première séquence concerne une ellipse gaussienne qui se dilate dans un fond noir, et la deuxième est une séquence d'images d'angiographie d'un cycle cardiaque. Les résultats sont comparés aux résultats obtenus par une méthode de régularisation classique.

Dans le cas de l'ellipse, avec le modèle classique (fig : 3.1), les vecteurs déplacement débordent de l'objet. Par contre notre modèle permet de rompre la continuité du mouvement au niveau des contours de l'ellipse (fig : 3.2). Sur le coeur, on peut noter un meilleur suivi des contours et de plus, une meilleure détection des zones en mouvement (fig : 4.1 et fig : 4.2). Pour la séquence de la main, on peut observer qu'on a une meilleure détection du mouvement des contours aux niveaux des doigts (fig : 5.1 et 5.2).

fig 3 : Résultats avec les images synthétiques

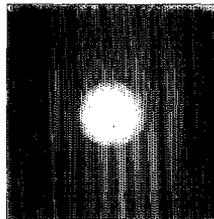


Image de la séquence.

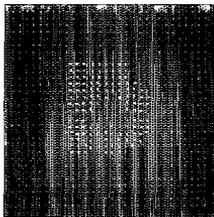


fig 3.1 : Image des directions des vecteurs déplacements (Modèle de Tikhonov)

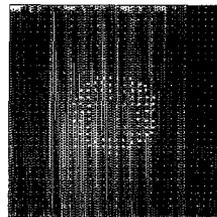


fig 3.2 : Image des directions des vecteurs déplacements (modèle proposé)

fig 4 : Résultats avec une séquence d'images angiographiques

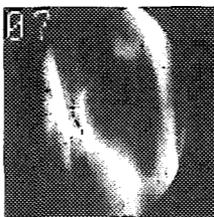


Image de la séquence.



**fig 4.1 : Image des directions
des vecteurs déplacements
(Modèle de Tikhonov)**



**fig 4.2 : Image des directions
des vecteurs déplacements
(modèle proposé)**

fig 5 : Résultats avec la séquence de la main.

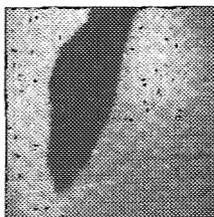
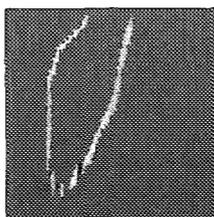
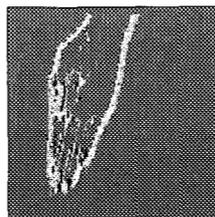


Image de la séquence.



**fig 5.1 : Image des directions
des vecteurs déplacements
(Modèle de Tikhonov)**



**fig 5.2.: Image des directions
des vecteurs déplacements
(modèle proposé)**

7. CONCLUSION

Nous avons présenté un modèle d'estimation du mouvement qui permet de prendre en compte les discontinuités. C'est une extension d'un modèle de régularisation non convexe introduit par Geman et McClure. Nous intégrons, dans ce modèle, l'amplitude locale du mouvement, calculée par l'analyse de Fourier. Cette amplitude est une information directement liée au mouvement. Par conséquent, elle fournit une nouvelle connaissance qui vient renforcer le terme de régularisation. De plus, nous montrons que le modèle obtenu présente un caractère localement convexe, ce qui est un avantage important pour l'application des algorithmes déterministes tel que l'I.C.M. Les résultats expérimentaux montrent que le modèle permet alors de traiter de manière très efficace les zones où apparaissent les discontinuités du mouvement, tels que les contours des objets.

8. REMERCIEMENTS : Ce travail a été financé par le département Français de la défense, contrat N° 94-089 / DRET.

9. REFERENCES

- [1] B.K.P.Horn and B.G.Schunck. "Determining Optical Flow". *Artificial intelligence*, 17 :185-203, 1981.
- [2] W.K.Pratt. "Digital Image Processing : Image Detection and Registration". Wiley-Inerscience Publication. 1978.
- [3] D.J.Fleet and Jepson. "Velocity Extraction without Form Interpretation". Proc. 3rd Workshop on Computer Vision : Representation and Control, pages 179-185, October 1985.
- [4] D.J.Heeger. "Optical Flow using Spatio-Temporal filters". *Int.Journal of Computer Vision*. Vol.1.N°4, pages 279-302. January 1988.
- [5] S. Geman and D. Geman "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images". *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 6, N°6: pages 721-741, November 1984.
- [6] J.Konrad and E.Dubois. "Multigrid Bayesian estimation of image motion fields using stochastic relaxation". In Proc.2nd Int. Conf. Computer Vision, pages 354-662, Tarpon Springs, Florida, dec 1988.
- [7] J.Besag. "On the Statistical Analysis of dirty pictures". *J. Royal Statist. Sos.*, Vol. 48. Serie B, N°3: pages 259-302, 1986.
- [8] A.N.Tikhonov, "Regularization of incorrectly posed problem," *Sov. Math. Dokl* 4, 1624-1627 1963.
- [9] G. Aubert, M. Barlaud and P. Charbonnier, "Theorems for Edge-Preserving Regularization in Computed Imaging". Submitted to *IEEE Transactions on Image Processing*, Tech. Rep., Research Report N° 94-01, I3S, University of Nice-Sophia Antipolis, 1994.
- [10] S.Geman and D.E.McClure. "Bayesian image analysis: an application to single photon emission tomography". in Proc. Statist. Comput. Sect. Amer. Statist. Assoc., Washington, DC, pages12-18, 1985.
- [11] Hohné, K. H., Obermoller, U., Riemer, M. And Witte, G., "Fourier Domain Techniques for Digital Angiography of the Heart," *E Trans Med. Imaging Vol MI-3* (1984) n°2 pp 62-67.

PRISE EN COMPTE DES DISCONTINUITES DANS L'ESTIMATION DU MOUVEMENT PAR LES CHAMPS DE MARKOV COMBINES A L'AMPLITUDE DE FOURIER

A.Dipanda*, M.kardouchi*, L.Legrand*

*Faculté des sciences Mirande,
Liesib-Département d'informatique
BP : 138, 21000 Dijon cedex

*Faculté de médecine,
laboratoire d'informatique médicale
7 bd Jeanne d'arc, 21033 Dijon cedex

ABSTRACT

The goal of this study is to estimate the motion between two successive images in a sequence. In this paper, we present an approach of the regularization problem which enables us to take into account motion discontinuities. This approach is based on the theory of Markov random fields and the Fourier analysis.

1. INTRODUCTION

L'étude de séquences d'images numériques introduit un champ spécifique d'investigation en analyse d'images, à savoir l'extraction d'informations sur le mouvement. Ces informations peuvent concerner : l'estimation du mouvement, la détection de mouvement, la segmentation au sens du mouvement, etc.

Dans cet article nous nous intéressons à l'estimation du champ de déplacement entre deux images d'une séquence donnée.

Pour estimer le mouvement, plusieurs méthodes existent parmi lesquelles :

- l'approche différentielle [1] : elle travaille dans le domaine direct. Elle est basée sur une ou plusieurs contraintes locales reliant les dérivées spatiale et temporelle du signal au mouvement apparent. L'hypothèse de base de cette méthode est l'invariance de la luminance d'un point lors d'un déplacement dans l'espace (x,y,t) constitué de deux images successives.
- l'approche de mise en correspondance [2] : ce type de méthode est basé sur la similarité entre blocs. Pour chacun des blocs, un vecteur de déplacement unique est alors estimé par la recherche du bloc le plus proche en terme de similarité ou de corrélation dans l'image suivante.
- l'approche par transformée : elle utilise le domaine de la transformée (principalement la transformée de Fourier [3] ou celle de Gabor [4]). Le mouvement se traduit par des modifications des coefficients de la transformée (déphasage de la transformée de Fourier pour un mouvement de translation par exemple).

- o l'approche statistique par modélisation markovienne[5][6] , comme dans l'approche différentielle, on travaille dans le domaine direct : elle est basée sur un traitement statistique des images et la théorie des champs de Markov.

Quelle que soit l'approche retenue, l'estimation d'un champ dense de déplacement introduit une prise en compte d'un certain voisinage. Il est alors supposé que le champ de déplacement possède une certaine propriété de continuité. Cependant, cette propriété est violée au voisinage des contours des objets en mouvement. Un des problèmes posés pour l'estimation du champ de déplacement est de savoir gérer les discontinuités susceptibles d'apparaître.

Dans cet article, nous proposons une méthode d'estimation du mouvement qui permet le traitement des discontinuités. Elle utilise un modèle markovien au sein duquel est intégré l'amplitude locale du mouvement, qui est une information obtenue par l'analyse de Fourier. Le modèle ainsi obtenu permet de faire un lissage adaptatif des vecteurs déplacement dans l'image.

Dans le prochain chapitre, nous présentons les principaux aspects de la modélisation markovienne en traitement d'images, en particulier pour l'estimation du mouvement. Dans le chapitre 3, nous décrivons les conditions à vérifier par une contrainte de régularisation pour la préservation des discontinuités du mouvement. Ensuite dans le chapitre 4, nous définissons l'amplitude locale du mouvement. Le chapitre 5 est consacré au modèle que nous avons développé. Enfin, dans les chapitres 6 et 7, nous présentons des résultats et nous donnons les conclusions sur la méthode proposée.

2. MODELISATION MARKOVIENNE ET ESTIMATION DU MOUVEMENT

2.1 Modélisation markovienne et estimation bayésienne en analyse d'images

Soit $\Omega = \{s = (i, j) : (1 \leq i \leq M \text{ et } 1 \leq j \leq M)\}$ une grille 2D de taille $M \times M = n$. On note par $V = \{v(s), s \in \Omega\}$ une famille arbitraire de voisinages locaux de Ω . Par exemple, le voisinage d'ordre un d'un pixel s , correspond à l'ensemble des pixels premier voisins de celui-ci (fig : 1) :

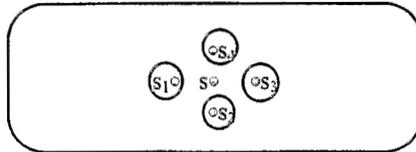


fig 1 : voisinage d'ordre 1.
○ : pixel voisin de s.

On note par C_s l'ensemble des cliques associées à un pixel s . Pour un voisinage d'ordre un, l'ensemble des cliques (fig : 2) est égal à :

$$C_s = \{ \{s_1\}, \{s_2\}, \{s_3\}, \{s_4\}, \{s, s_1\}, \{s, s_2\}, \{s, s_3\}, \{s, s_4\} \}$$

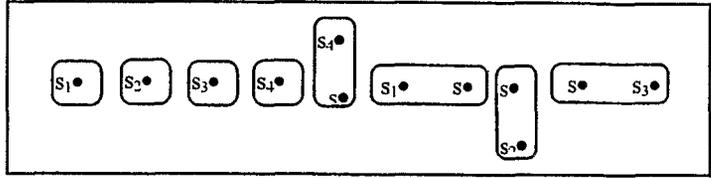


fig 2 : Cliques associées au voisinage d'ordre un

Un champ de variables aléatoires $I = (I_{s_1}, I_{s_2}, \dots, I_{s_n})$ défini sur une grille S est un champ de Markov relativement à un système de voisinage V si et seulement si :

$$\Pr(I_{s_k} = i_{s_k} / I_{s_j} = i_{s_j}, s_j \in \Omega - \{s_k\}) = \Pr(I_{s_k} = i_{s_k} / I_{s_j} = i_{s_j}, s_j \in v(s_k) - \{s_k\}) \quad (2.1.1)$$

Où $i = (i_{s_1}, i_{s_2}, \dots, i_{s_n})$ est une réalisation de I .

Ceci signifie que la connaissance d'un voisinage local d'un site s donné est suffisante pour calculer la probabilité en celui-ci. Cette propriété de dépendance locale est rendue exploitable grâce au théorème d'Hammersley-Clifford [7] qui démontre qu'un champ aléatoire défini sur un réseau est un champ de Markov si et seulement si sa distribution de probabilité est une distribution de Gibbs définie par :

$$\Pr(I=i) = (1/Z) \exp(-U(I=i)) \quad (2.1.2)$$

Où Z est une constante de normalisation appelée fonction de partition.
 $U(I=i)$ est la somme de fonctions locales appelées fonctions énergies.

La modélisation markovienne en analyse d'images s'appuie sur deux champs de variables aléatoires distincts :

- le champ des observations : il est formé par l'ensemble des données sur lesquels s'appuie l'analyse.
- le champ des primitives : qui est l'information non observable que l'on cherche à extraire des observations.

Ils sont supposés former deux champs de variables aléatoires discrets que nous noterons respectivement O et P . Ils sont définis sur deux grilles rectangulaires R et S (identiques ou non) :

$$O = \{O_{r_i}, r_i \in R\} \quad P = \{P_{s_j}, s_j \in S\} \quad (2.1.3)$$

où r_i et s_j désignent des sites des grilles R et S .

On notera par $o = \{o_{r_i}, r_i \in R\}$ et $p = \{p_{s_j}, s_j \in S\}$ une réalisation quelconque des champs aléatoires O et P .

Le critère d'estimation du champ de primitives le plus souvent retenu est le Maximum A Posteriori (M.A.P) : étant donné le champ d'observation o , on cherche le champ de primitives $p(\text{estimé})$ le plus probable au sens de la distribution a posteriori :

$$p(\text{estimé}) = \underset{p}{\text{Arg Max}} \Pr (P = p / O = o). \quad (2.1.4)$$

En utilisant la règle de Bayes, à un facteur constant près, on obtient :

$$p(\text{estimé}) = \underset{p}{\text{Arg Max}} \Pr (O = o / P = p) \Pr (P = p) \quad (2.1.5)$$

Le théorème d'Hammerley-Cliford permet de ramener le problème d'estimation de primitives à la minimisation d'une fonction énergie U :

$$p(\text{estimé}) = \min_p U \quad (2.1.6)$$

La recherche d'un tel minimum est un problème d'optimisation. Une solution approchée peut être obtenue soit, par des algorithmes déterministes tel que l'I.C.M. (Itérative Conditional Mode) [7], soit, par des algorithmes stochastiques de type 'recuit simulé' [5]. Les algorithmes stochastiques convergent vers le minimum global, mais restent coûteux en temps de calcul. Par contre, les algorithmes déterministes convergent beaucoup plus rapidement, mais présentent le défaut de pouvoir se faire piéger par un minimum local. Une bonne initialisation de départ du champ de primitives est donc nécessaire.

2.2 Estimation du mouvement

Dans le cadre de l'estimation du mouvement, les observations sont constituées par deux images successives de la séquence entre lesquelles le mouvement sera estimé, et les primitives sont représentées par les déplacements calculés sur chaque pixel. L'hypothèse émise sur la fonction énergie suppose que la luminance d'un pixel est invariante entre deux images successives. Ainsi si on note par $f(s, t)$ le niveau de gris à l'instant t correspondant au pixel s et ds le déplacement de s entre t et $t+dt$, l'hypothèse décrite ci-dessus se traduit par :

$$DFD(s) = (f(s+ds, t+dt) - f(s, t)) = 0 \quad (2.2.1)$$

DFD signifiant 'displaced frame difference'.

Ainsi, l'estimation du champ de déplacement en un pixel s revient à trouver le déplacement ds tel que $DFD(s) = 0$, ce qui équivaut à minimiser $DFD(s)^2$. La fonction énergie s'écrit alors :

$$U_{D,F,D}(ds, f_t, f_{t+dt}) = \sum_s U_{D,F,D}^s(ds, f_t, f_{t+dt}) \quad (2.2.2)$$

$$\text{avec } U_{D,F,D}^s(ds, f_t, f_{t+dt}) = (f(s+ds, t+dt) - f(s, t))^2 \quad (2.2.3)$$

Le critère d'estimation du champ de déplacement est obtenu pour chaque pixel s en minimisant l'équation (2.2.3) :

$$ds(\text{estimé}) = \underset{ds \in D}{\text{Min}} U_{D,F,D}^s(ds, f_t, f_{t+dt}) \quad (2.2.4)$$

où $D \subset \mathbb{Z}^2$

Dans la majorité des cas, les phénomènes qui président à la formation des observations s'accompagnent d'une dramatique perte d'informations intervenant aux différentes étapes du processus. On est en présence d'un problème *mal posé*, la seule connaissance des données ne suffit pas pour assurer l'existence, l'unicité et la stabilité de la solution. Pour contourner cette difficulté, il est nécessaire d'émettre des hypothèses a priori sur les primitives. On est ainsi amené à introduire une contrainte qualifiée de contrainte de régularisation (ou de lissage).

Ainsi, l'estimation du vecteur déplacement en un pixel s est obtenu par l'équation suivante :

$$ds(\text{estimé}) = \underset{ds \in D}{\text{Arg Min}} (U_{D,F,D}^s(ds, f_t, f_{t+dt}) + U_R^s(ds)) \quad (2.2.5)$$

Un modèle de régularisation classique est donné par Tikhonov [3]. Il est défini par :

$$U_R^s(ds) = \sum_{\{s,t_i\} \in C_T} \beta^2 \|ds - dt_i\|^2 = \sum_{\{s,t_i\} \in C_T} \phi_s(\|ds - dt_i\|^2) \quad (2.2.6)$$

$$\text{avec } \phi_s(\|ds - dt_i\|^2) = \beta^2 \|ds - dt_i\|^2$$

et $\| \cdot \|$ désigne la norme euclidienne.

β^2 est un paramètre de pondération entre $U_{D.F.D}^s(ds, f_i, f_{i+dt})$ et $U_R^s(ds)$.

$t_i : i = 1, 2, 3, 4$ étant l'ensemble des premiers voisins de s .

La convexité du modèle de régularisation donné par Tikhonov permet aux algorithmes déterministes de converger vers le minimum global. Cependant, la nature isotropique de ce modèle ne permet pas de gérer correctement les discontinuités du mouvement susceptibles d'apparaître au niveau des contours des objets en mouvement.

3. MODELISATION DES DISCONTINUITES

Pour qu'une contrainte de régularisation préserve les discontinuités, elle doit appartenir à la classe des "phi-fonction" qui vérifient [9] :

$$\begin{aligned} \lim_{x \rightarrow 0^+} (\phi'(x) / 2x) &= M && \text{avec } M \text{ finie} \\ \lim_{x \rightarrow +\infty} (\phi'(x) / 2x) &= 0 \end{aligned} \quad (3.1)$$

$(\phi'(x) / 2x)$ est continue et strictement décroissante dans $[0, +\infty[$.

Plusieurs "phi-fonction" sont proposées dans la littérature [9][10]. Cependant la plupart d'entre elles ne sont pas convexes, et par conséquent, sont difficiles à minimiser. Nous proposons dans cet article de construire une approximation localement convexe de la phi-fonction définie par $\phi(x) = x^2 / (p^2 + p^2 x^2)$ en utilisant l'analyse de Fourier.

4. ANALYSE DE FOURIER ET AMPLITUDE LOCALE DU MOUVEMENT

Soit $f(x)$ une fonction périodique de période T_0 ($T_0 = 1/v_0$) où x est une variable temporelle ou spatiale. Si $f(x)$ satisfait les conditions de Dirichlet [11], alors $f(x)$ peut être représentée par une série réelle infinie de fonctions trigonométriques. En se limitant à une somme finie et en discrétisant $f(x)$, on obtient :

$$f(n) \approx \sum_{k=0}^{N-1} (a_k \cos(2\pi n k v_0) + b_k \sin(2\pi n k v_0)) \quad (4.1)$$

où a_k et b_k sont les coefficients de Fourier :

$$a_k = \sum_{n=0}^{N-1} f(n) \cos(2\pi n k v_0) \quad \text{et} \quad b_k = \sum_{n=0}^{N-1} f(n) \sin(2\pi n k v_0) \quad (4.2)$$

cette fonction peut encore s'écrire :

$$f(n) = \sum_{k=0}^{N-1} C_k \cos(2\pi n k v_0 - \Phi_k) \quad (4.3)$$

$$\text{avec } C_k = \sqrt{a_k^2 + b_k^2} \quad \text{et} \quad \Phi_k = \text{Arctg}(-b_k / a_k) \quad (4.4)$$

C_k et Φ_k sont respectivement l'amplitude et la phase de la $k^{\text{ième}}$ harmonique.

Considérons maintenant une séquence de N images. Soit $f(s,n)$ le niveau de gris du pixel s de la $n^{\text{ième}}$ image. Le signal $f(s,n)$ ($n=1,2,\dots,N$) est de période égale à N et de fréquence égale à $(1/N)=v_0$. On connaît l'évolution en niveau de gris $f(s,n)$ pour chacun des pixels s le long de la séquence. En se limitant à une somme finie (nombre d'images dans la séquence) de toute les fréquences possibles, on obtient :

$$f(s,n) \approx \sum_{k=0}^{N-1} (a_k(s) \cos(2\pi n k v_0) + b_k(s) \sin(2\pi n k v_0)) \quad (4.5)$$

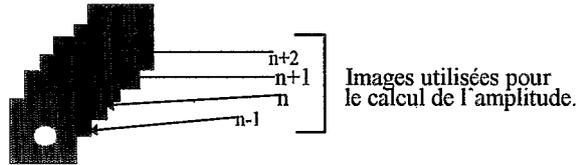
avec
$$a_k(s) = \sum_{n=0}^{N-1} f(s,n) \cos(2\pi n k v_0) \quad (4.6)$$

$$b_k(s) = \sum_{n=0}^{N-1} f(s,n) \sin(2\pi n k v_0) \quad (4.7)$$

$$c_k(s) = \sqrt{a_k(s)^2 + b_k(s)^2} \quad (4.8)$$

Pour un pixel s donné, le coefficient de Fourier $c_k(s)$ le plus important correspond à la fréquence égale à $1/N$ qui est la fréquence de la première harmonique, et par conséquent, seuls les coefficients $a_1(s)$ et $b_1(s)$ de la première harmonique sont calculés. L'intérêt de l'amplitude de Fourier ainsi calculée, est qu'elle nous donne une information sur l'amplitude du mouvement en chaque pixel le long de la séquence.

L'estimation du mouvement porte sur deux images successives, n et $n+1$. Pour calculer l'amplitude locale du mouvement entre ces deux images, nous nous limitons à quatre images $n-1$, n , $n+1$, et $n+2$ de la séquence. Ceci permet de limiter les "perturbations" induites par les images éloignées de n et $n+1$. On obtient ainsi :



On obtient ainsi :

$$a_1(s) = \sum_{n'=n-1}^{n'+n+2} f(s,n') \cos(2\pi n' v_0) \quad (4.9)$$

$$b_1(s) = \sum_{n'=n-1}^{n'+n+2} f(s,n') \sin(2\pi n' v_0) \quad (4.10)$$

L'amplitude locale du mouvement du mouvement en chaque pixel s s'écrit alors :

$$C_1(s) = \sqrt{a_1(s)^2 + b_1(s)^2} \quad (4.11)$$