

Estimation paramétrique multirésolution de mouvements transparents

Vincent AUVRAY^{1,2}, Patrick BOUTHEMY¹, Jean LIÉNARD²

¹IRISA/INRIA, Campus de Beaulieu, 35042 Rennes Cedex, France

²General Electric Healthcare, 283 rue de la Minière, 78530 Buc, France

Vincent.Auvray@irisa.fr, patrick.bouthemy@irisa.fr, jean.lienard@med.ge.com

Résumé – Nous présentons une nouvelle méthode permettant d’estimer les mouvements dans des images avec transparence. Il s’agit d’une estimation par blocs basée sur une minimisation multirésolution. Elle est initialisée par la technique du downhill simplex. Un post-traitement original, effectuant des estimations séparées des mouvements sur des différences d’images recalées de manière appropriée, améliore la qualité des mouvements estimés. L’algorithme se conclut par une étape de régularisation. Nous démontrons sur un grand nombre de simulations réalistes que nous atteignons une précision d’un quart de pixel sur des images non bruitées. Le cas des images bruitées est étudié, avec des résultats satisfaisants sur des images cliniques peu contrastées. Nous présentons également une expérimentation sur des images cliniques réelles.

Abstract – A new framework dealing with motion estimation in transparent images is presented. It relies on a block-oriented estimation involving an efficient multiresolution minimization. A downhill simplex method provides an appropriate initialization to this scheme. The estimated velocity vectors are greatly improved by an original postprocessing stage which performs a single motion estimation on differences of warped images. Finally, a regularization step is carried out. It is demonstrated on a large set of simulations that a quarter-pixel accuracy can be attained on noise-free images. The case of noisy images is also addressed and provides satisfactory results, even in the case of low-contrasted medical images. An experiment on a real clinical image sequence is also reported with promising results.

1 Introduction

Les images médicales acquises par rayons X ont une structure très particulière puisque leur processus de formation est gouverné par le phénomène de transparence. A la différence des images vidéo, il n’y a pas d’occultation lorsqu’un objet en couvre un autre mais une addition des niveaux de gris. Nous ne pouvons donc pas appliquer directement les hypothèses et modèles de mouvements classiques à ce type d’image mais devons explicitement prendre en compte la transparence. Nous proposons d’estimer les mouvements contenus dans ces séquences transparentes par une approche exploitant des modèles paramétriques et une estimation multirésolution du mouvement.

Le calcul des mouvements transparents a déjà fait l’objet de nombreuses études, principalement dans le cadre d’images vidéo contenant des effets de transparence particuliers: reflets dans un miroir ou sur l’eau, objet en mouvement vu au travers d’un rideau ou d’une barrière...

Un premier ensemble de travaux adapte des techniques d’estimation classiques au cas de la transparence [1]. Une autre approche consiste à modéliser explicitement la transparence en utilisant l’équation fondamentale introduite par Shizawa et Mase [2]. Si l’on considère deux couches d’image I_1 et I_2 se déplaçant avec les vitesses relatives \vec{u} et \vec{v} , supposées constantes entre $t - 1$ et $t + 1$, l’image composite $I = I_1 + I_2$ vérifie en tout point \vec{p} (les mouvements étant en fait orientés vers le passé):

$$\begin{aligned} r(\vec{p}, \vec{u}, \vec{v}) \\ &= I(\vec{p} + \vec{u} + \vec{v}, t - 1) + I(\vec{p}, t + 1) \\ &\quad - I(\vec{p} + \vec{u}, t) - I(\vec{p} + \vec{v}, t) = 0 \end{aligned} \quad (1)$$

Cette équation peut être étendue au cas de n couches en faisant intervenir $n + 1$ images, toujours sous l’hypothèse d’invariance du mouvement. Nous nous limitons par la suite au cas de deux couches, mais notre algorithme peut s’adapter facilement à n couches transparentes.

De nombreuses méthodes permettant d’estimer les mouvements transparents ont été construites sur cette équation, que ce soit en minimisant une fonction d’énergie comprenant un terme de régularisation [3], en introduisant une décomposition des vitesses en ondelettes ou sur une base de B-splines [4], en utilisant des champs de Markov ou une mise en correspondance par blocs [5]. Une alternative consiste à formuler le problème dans l’espace des fréquences [6], mais il faut alors supposer le mouvement constant sur un intervalle de temps beaucoup plus long (quelques dizaines d’images).

Cet article est structuré comme suit. Le section 2 développe la structure de l’estimateur multirésolution proposé. La section 3 décrit une importante étape de post-traitement et des résultats expérimentaux sont présentés dans la section 4.

2 Estimation de mouvements en transparence

2.1 Contrainte de mouvement transparent

L’équation (1) n’est en fait exacte que si nous supposons que I_1 se déplace avec la vitesse \vec{u} non seulement au point \vec{p} à la date t , mais aussi au point $\vec{p} + \vec{v}$ à la date $t - 1$. Il en va de même pour I_2 en \vec{p} et $\vec{p} + \vec{u}$. Ceci implique en pratique que (1) n’est rigoureusement respectée sur une région donnée que

si les couches y sont en translation.

C'est pourquoi nous choisissons de traiter le problème de mouvements complexes en divisant l'image en petits blocs, sur lesquels le mouvement est considéré comme translationnel. L'expression J à minimiser sur chaque bloc B_i est donnée par:

$$J(\vec{u}^i, \vec{v}^i) = \sum_{\vec{p} \in B_i} r(\vec{p}, \vec{u}^i, \vec{v}^i)^2 \quad (2)$$

où $r(\vec{p}, \vec{u}^i, \vec{v}^i)$ est donné par l'équation (1)

Dans la pratique, nos expériences ont montré que des blocs de taille 32×32 offraient un bon compromis entre le respect de l'hypothèse de translation d'une part, et le nombre de pixels nécessaires à une estimation fiable du mouvement d'autre part.

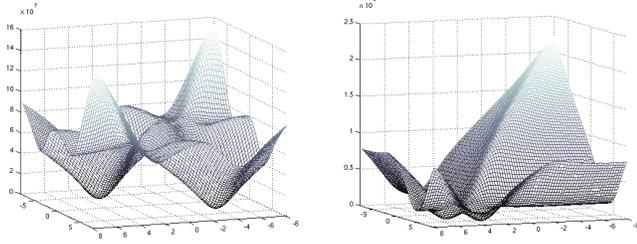


FIG. 1: Deux tracés de la fonction $J(u, v)$ réduite au sous-espace (u_x, v_x) . Vraies vitesses: gauche: $u = (3, 0)$, $v = (-4, 0)$, droite: $u = (3, 0)$ et $v = (5, 0)$.

Deux tracés de la fonction $J(\vec{u}, \vec{v})$ sont représentés sur la figure 1 afin d'illustrer ses propriétés. Nous avons construit une séquence de trois images en déplaçant deux couches selon la direction x avec des vitesses connues. Ainsi, le graphique peut être limité à l'espace (u_x, v_x) . Notons que ces fonctions présentent deux minima symétriques puisque \vec{u} et \vec{v} jouent des rôles similaires dans l'équation (1). Cette ambiguïté ne pose pas de problème pour l'heure car nous ne cherchons pas à associer \vec{u} et \vec{v} à des couches spécifiques.

2.2 Estimation multirésolution de mouvements transparents

Considérant trois images successives contenant deux mouvements transparents, nous estimons leurs vitesses en minimisant $J(\vec{u}, \vec{v})$. Si les amplitudes des déplacements étaient petites, nous pourrions travailler sur une version linéarisée de J et la minimiser avec un algorithme itératif de minimisation classique. Cependant, comme nous pouvons rencontrer des mouvements importants, nous utilisons un schéma multirésolution basé sur une décomposition pyramidale gaussienne des trois images consécutives. Au niveau de résolution le plus grossier, les déplacements sont assez petits pour permettre une minimisation par un algorithme de gradient conjugué de la fonction $J(\vec{u}, \vec{v})$ linéarisée autour de $(\vec{0}, \vec{0})$. Nous obtenons ainsi une première approximation (\vec{u}^L, \vec{v}^L) des deux mouvements. Nous procédons de la même manière au niveau inférieur $L - 1$, mais cette fois en linéarisant J autour de $(2 \cdot \vec{u}^L, 2 \cdot \vec{v}^L)$. Cette méthode est ensuite itérée à travers la décomposition pyramidale jusqu'à son niveau le plus fin.

2.3 Initialisation par l'algorithme du simplex

Cette stratégie de minimisation est efficace et rapide, mais elle est sensible à son initialisation, et ce d'autant plus que nous travaillons avec des images médicales obtenues par rayons X qui sont peu contrastées et très bruitées.

Nous nous appuyons sur la méthode du downhill simplex pour initialiser correctement notre estimation itérative. Cet algorithme minimise des fonctions non linéaires dans un espace de dimension n quelconque en déplaçant itérativement $n + 1$ points de cet espace vers un minimum. Ainsi, pour minimiser J dans l'espace de dimension quatre (\vec{u}, \vec{v}) , nous considérons cinq échantillons de couples de vitesses (\vec{u}, \vec{v}) . A chaque itération, nous remplaçons l'échantillon correspondant à la plus grande valeur de J par un nouvel échantillon recherché sur une droite perpendiculaire à l'hyperplan contenant les n autres points [7]. Cet algorithme fournit une initialisation appropriée de notre problème de minimisation.

Les coûts de calcul restent faibles puisque nous utilisons l'algorithme du simplex seulement au niveau le plus grossier de la pyramide de résolution.

3 Étape de post-traitement

3.1 Recalage et estimation mono-mouvement

Nous remarquons sur les diagrammes de la fonction J sur la Fig.1 que le minimum global peut être difficile à atteindre quand les vitesses des deux couches sont proches (exemple de droite). (Au passage, si les vitesses sont quasiment identiques, le problème de la transparence disparaît car nous sommes ramenés à l'étude d'un mouvement unique!) Pour surmonter ce problème, nous avons mis au point l'étape de post-traitement suivante.

L'étape d'estimation multirésolution fournit les vitesses estimées $(\hat{u}_{me}, \hat{v}_{me})$. Supposons que \hat{u}_{me} soit une estimée correcte, nous pouvons faire disparaître implicitement la couche correspondante (disons I_1) sans l'identifier.

Recalons $I(\vec{p}, t)$ et $I(\vec{p}, t - 1)$ vers l'instant $t + 1$ selon les déplacements \hat{u}_{me} et $2 \cdot \hat{u}_{me}$ respectivement. La position de I_1 est censée être la même dans les deux images recalées, ainsi que dans $I(\vec{p}, t + 1)$. Nous formons les deux cartes de différence suivantes, qui permettent de supprimer la couche I_1 :

$$\begin{aligned} K_\alpha(\vec{p}, t + 1) &= I(\vec{p} + \hat{u}_{me}, t) - I(\vec{p} + 2 \cdot \hat{u}_{me}, t - 1) \\ &= I_2(\vec{p} + \hat{u}_{me}, t) - I_2(\vec{p} + 2 \cdot \hat{u}_{me}, t - 1) \end{aligned} \quad (3)$$

$$\begin{aligned} K_\beta(\vec{p}, t + 1) &= I(\vec{p}, t + 1) - I(\vec{p} + \hat{u}_{me}, t) \\ &= I_2(\vec{p}, t + 1) - I_2(\vec{p} + \hat{u}_{me}, t) \end{aligned} \quad (4)$$

On remarque alors que:

$$K_\alpha(\vec{p}, t + 1) = K_\beta(\vec{p} + \hat{u}_{me}, t) \quad (5)$$

Comme les cartes K_β ne font intervenir que I_2 , l'équation du mouvement classique peut s'écrire pour ces cartes selon:

$$K_\beta(\vec{p}, t + 1) = K_\beta(\vec{p} + \vec{v}, t) \quad (6)$$

D'où:

$$K_\alpha(\vec{p}, t + 1) = K_\beta(\vec{p} + \hat{u}_{me} - \vec{v}, t + 1) \quad (7)$$

En appliquant une méthode d'estimation robuste mono-mouvement (telle que [8]) entre les deux cartes K_α et K_β des instants $t + 1$, nous estimons en fait le mouvement $\vec{s}^t = \vec{u}_{me}^t - \vec{v}^t$, ce qui permet de réestimer \vec{v}^t comme $\vec{v}_{pp}^t = \vec{u}_{me}^t - \vec{s}^t$. De même, nous réestimons \vec{u}_{pp}^t à partir de $\vec{t}^t = \vec{v}_{me}^t - \vec{u}^t$. Ainsi, si une vitesse seulement est correctement estimée, cette méthode permet de recalculer la seconde. Il reste alors à choisir les meilleurs estimées parmi les quatre paires possibles $(\vec{u}_{me}^t, \vec{v}_{me}^t)$, $(\vec{u}_{pp}^t, \vec{v}_{pp}^t)$, $(\vec{u}_{pp}^t, \vec{v}_{me}^t)$ et $(\vec{u}_{me}^t, \vec{v}_{pp}^t)$.

3.2 Stratégie de choix

Une stratégie intuitive consisterait à sélectionner les estimées \vec{u} et \vec{v} correspondant à la plus petite valeur du critère J de la relation (2). De fait, cela donne de bons résultats sur des images non bruitées, mais ne fonctionne plus sur des images bruitées. En effet, l'expression de J est formée de la somme algébrique de quatre images. Si les images originales sont entachées d'un bruit de variance σ^2 , J est affecté par un bruit de variance $4.\sigma^2$, ce qui rend le choix très peu fiable.

Nous proposons d'appliquer un schéma de régularisation. Nous souhaitons privilégier les couples de vitesse qui minimisent raisonnablement (2), tout en étant cohérents avec les vitesses des blocs voisins.

Pour cela, nous avons utilisé un procédé d'estimation alternée. Dans un premier temps, nous calculons deux modèles affines de mouvement (un pour chaque couche) compatibles avec les mouvements estimés sur les blocs, puis nous choisissons sur chaque bloc celui des quatre couples de vitesses candidates qui est le plus proche des modèles de mouvement estimés. Sur chaque bloc, les deux déplacements moyens correspondant aux modèles affines de mouvements estimés sont évalués et comparés (en utilisant la norme L_2) aux quatre paires de candidats, et l'on retient la paire qui en est la plus proche.

Cet algorithme itératif fonctionne bien pourvu qu'il soit correctement initialisé. Les expériences que nous avons menées nous ont conduits à initialiser les couples de vecteurs de vitesse sur les blocs par $(\vec{u}_{me}, \vec{v}_{me})$. Ces vecteurs de vitesse sont regroupés au préalable en deux groupes $\{\vec{u}_{me}\}$ et $\{\vec{v}_{me}\}$, correspondant aux deux couches impliquées, par une technique de clustering. Cette opération est indispensable pour s'assurer que u et v correspondent bien à la même couche de bloc en bloc.

La Fig.2 résume la structure de l'algorithme complet:

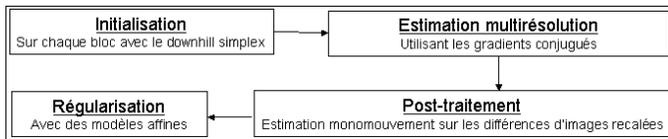


FIG. 2: Structure globale de la méthode d'estimation de mouvements transparents.

4 Résultats expérimentaux

4.1 Synthèse d'images cliniques réalistes

Nous avons synthétisé des séquences d'images par rayons X réalistes pour évaluer la performance de notre méthode en con-

naissant la réalité terrain. Pour ce faire, nous avons utilisé deux images cliniques (correspondant aux deux couches) enregistrées sur un système d'imagerie par rayons X, que nous avons déplacées par des champs de mouvements translationnels et affines connus. L'amplitude des mouvements est tirée aléatoirement tout en assurant des déplacements maximaux de 8 pixels en chaque point. L'intensité des images est multipliée par un facteur simulant le rayonnement utilisé lors de l'examen: plus la dose de rayons X simulée est haute, plus l'intensité et le contraste des images sont importants.

Puisque la physique des rayons X implique une transparence multiplicative, les deux couches déplacées sont multipliées entre elles, puis nous considérons le logarithme de l'image composite pour obtenir une situation de transparence additive.

Les images 12 bits finales sont corrompues par un bruit indépendant gaussien dont l'écart type dépend de la dose simulée. Nous présentons les résultats obtenus sur des images non bruitées (à titre de référence), sur des images correspondant à de fortes doses ($\sigma = 10$, typique des examens diagnostiques) et sur des images correspondant à de faibles doses ($\sigma = 20$, typique des examens interventionnels). Le niveau moyen des intensités des images simulées est de l'ordre de 500.

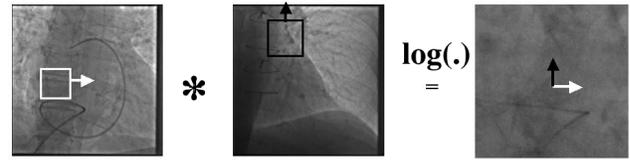


FIG. 3: Génération d'images transparentes

4.2 Résultats

Nous avons comparé quatre versions de notre algorithme pour évaluer les contributions des différents modules. *ME* n'accomplit que les deux premières étapes: initialisation et estimation multirésolution. *PP* inclut l'étape de post-traitement mais retient la meilleure paire de vitesses selon les valeurs de la fonction J seulement, sans accomplir de régularisation. *RE* effectue la régularisation et garde les couples sélectionnés sur chaque bloc selon le critère présenté au paragraphe 3.2, alors que *PR* retient au final les vecteurs de mouvement sur chaque bloc correspondant effectivement aux modèles affines estimés. Dans les expériences présentées ci-dessous, une des deux couches a subi un mouvement affine et l'autre une translation. Le temps de calcul pour estimer les vecteurs de vitesse est d'environ 10 secondes pour des images 288×288 sur un Pentium IV à 2.4 GHz avec 1Go de mémoire. Le tableau 1 contient les erreurs moyennes d'estimation sur les vecteurs de vitesse pour 250 générations aléatoires d'exemples traités.

Nous observons que l'étape de post-traitement améliore la qualité des estimées sur les images bruitées, dès lors que la sélection de la paire candidate est effectuée comme expliqué en 3.2 (méthode *RE*).

Il se trouve que pour cette évaluation, le meilleur choix serait de conserver les vecteurs de vitesse correspondant aux mouvements affines estimés (méthode *PR*). Elle assure une précision subpixelique sur les images bruitées caractéristiques d'examen fluoroscopiques. Cela s'explique par le fait que les modèles de

TAB. 1: Erreur d'estimation moyenne en pixels pour différents niveaux de bruit et différentes versions de l'estimateur (voir texte principal).

Bruit	0	10	20
ME	0.33	0.72	1.59
PP	0.3	0.77	1.69
RE	0.25	0.62	1.43
PR	0.20	0.34	0.87

mouvements utilisés sont exacts pour ces séquences d'images simulées, ce qui ne serait pas le cas pour des séquences d'images réelles.

Nous avons également appliqué notre algorithme à des images médicales réelles. Nous présentons des premiers résultats sur une séquence cardiaque fluoroscopique acquise à 30 images/s. Elle représente une zone d'environ $5\text{cm} \times 5\text{cm}$ à droite du coeur. Deux images de cette séquence sont reproduites à la figure 4, ainsi que les mouvements transparents calculés.

Le coeur (qui apparaît sombre) bat sur la droite de l'image sur un fond fixe correspondant aux côtes et à la colonne vertébrale. Les tissus clairs des poumons à gauche suivent les mouvements du coeur. L'amplitude du mouvement sur un cycle est de 25 pixels. Les images ont un faible contraste et sont affectées par un bruit important ($\sigma \simeq 20$).

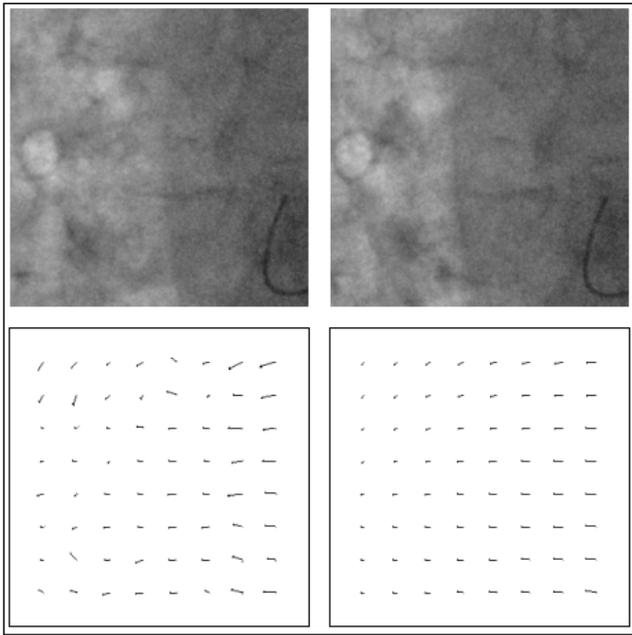


FIG. 4: Haut: Images 1 et 8 de la séquence. Bas: Vitesses estimées sur les trois premières images. Gauche: Estimation finale sur chaque bloc utilisant l'algorithme RE (par souci de visibilité, seul un vecteur est représenté pour chaque bloc car la deuxième couche est statique), droite: modèle affine estimé.

Nos estimations sont cohérentes avec les mouvements observés sur la séquence. Deux mouvements ont été estimés, et ils correspondent à la vérité anatomique: le fond, qui est statique (et n'apparaît donc pas sur les champs de vecteurs tracés à la figure 4), et le mouvement du coeur proche d'un modèle affine. L'amplitude du mouvement estimé décroît bien dans la zone des poumons selon la distance au coeur. Ces

images sont bruitées, peu contrastées, et contiennent des mouvements complexes. En outre, les mouvements ne sont pas parfaitement constants sur trois images successives. Même dans ce cas difficile, la méthode d'estimation proposée donne des résultats intéressants.

5 Conclusion

Nous avons décrit une nouvelle méthode complète d'estimation multirésolution de mouvements transparents pour deux couches. Elle contient en particulier une initialisation efficace et un post-traitement original qui améliorent la qualité de l'estimation multirésolution seule. Nous avons montré sur un grand nombre de simulations que nous atteignons une précision de l'ordre du quart de pixel sur des images non bruitées. Nous avons étudié l'estimation sur images bruitées avec des résultats satisfaisants pour des niveaux de bruit temporel réaliste, et ce pour des images cliniques peu contrastées. Nous avons également traité des données réelles avec des résultats fort encourageants.

Nous allons mener plus d'expériences sur des images réelles acquises par rayons X, et étudier des configurations de transparence plus complexes. Nous pensons en particulier étudier le cas, fréquent cliniquement, d'images transparentes contenant trois couches, mais avec au plus deux couches en chaque point.

Références

- [1] M. Irani, B. Rousso, et S. Peleg, "Computing occluding and transparent motions," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 5–16, 1994.
- [2] M. Shizawa et K. Mase, "Simultaneous multiple optical flow," in *Int. Conf. on Pattern Recognition*, Atlantic City, 1990, vol. 1, pp. 274–278.
- [3] J. Toro, F. Owens, et R. Medina, "Multiple motion estimation and segmentation in transparency," in *Proc. of the IEEE Int. Conference on Acoustics, Speech and Signal Processing*, Istanbul, juin 2000, pp. 2087–2090.
- [4] M. Pingault, E. Bruno, et D. Pellerin, "A robust multi-scale B-spline function decomposition for estimating motion transparency," *IEEE Trans. on Image Processing*, vol. 12, no. 11, pp. 1416–1426, nov. 2003.
- [5] I. Stuke, T. Aach, C. Mota, et E. Barth, "Estimation of multiple motions using block-matching and Markov random fields," in *SPIE Conf. Visual Communications and Image Processing VCIP'2004*, San Jose, janvier 2004, pp. 274–278.
- [6] M. Pingault et D. Pellerin, "Motion estimation of transparent objects in the frequency domain," *Signal Processing*, vol. 84, no. 4, pp. 709–719, avril 2004.
- [7] W.H.Press et al, *Numerical recipes in C : The Art of Scientific Computing*, chapter Downhill Simplex Method in Multidimensions, pp. 408–412, Cambridge University Press, 1988-1992.
- [8] J.-M. Odobez et P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *Journal of Visual Communication and Image Representation*, vol. 6, no. 4, pp. 348–365, décembre 1995.