

Estimation parametrique multiresolution de mouvements transparents

Vincent AUVRAY^{1,2}, Patrick BOUTHEMY¹, Jean LIENARD²

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

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

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

Résumé – Nous presentons une nouvelle methode pour l’estimation de mouvements sur des images transparentes. Il s’agit d’une estimation par blocs basee sur une minimisation multiresolution de fonction. Elle est initialisee par la methode du downhill simplex. La qualite des vecteurs estimees est grandement amelioree par un posttraitement original qui effectue des estimations monomouvement sur des differences d’images recalees. Nous effectuons finalement une regularisation. Nous demontrons sur un grand nombre de simulations que nous atteignons une precision d’un quart de pixel sur des images non bruitees. Le cas des images bruitees est egalement etudie, avec des resultats satisfaisant pour une niveau de bruit raisonnable, y compris dans le cas d’images cliniques peu contrastees. Nous donnons egalement un exemple prometteur sur des images cliniques reelles.

Abstract – A new framework dealing with motion estimation in transparent images is presented. It relies on a block-oriented estimation with efficient multiresolution function 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 precision 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 example on real clinical images is also reported with promising results

1 INTRODUCTION

Les images medicales acquises par rayons X ont une structure tres particuliere puisque leur processus de formation est gouverne par le phenomene de transparence. A la difference des images video classiques, il n’y a pas d’occlusion lorsqu’un objet en couvre un autre mais un ajout de leur niveau de gris. Nous ne pouvons donc pas appliquer directement les hypotheses et modeles de mouvements classiques a ce type d’image mais devons explicitement prendre en compte la transparence. Nous proposons une nouvelle approche a base de modeles parametriques et d’estimation multiresolution du mouvement pour estimer les mouvements contenus dans des images composees de deux couches transparentes.

Le calcul de mouvement transparent a deja fait l’objet de nombreuses etudes, principalement dans le cadre d’images video 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 barriere... Un premier ensemble de methodes adapte des techniques d’estimation classiques au cas de la transparence [1]. Une autre approche consiste a modeliser explicitement la transparence en utilisant l’equation fondamentale due a Shizawa et Mase [2]. Si l’on considere deux images couches I_1 et I_2 se deplacant avec les vitesses relatives $u = (u_x, u_y)$ et $v = (v_x, v_y)$, constantes entre $t - 1$ et $t + 1$,

$$\begin{aligned} r(x, y, u, v) = & \\ & I(x + u_x + v_x, y + u_y + v_y, t - 1) + I(x, y, t + 1) \\ & - I(x + v_x, y + v_y, t) - I(x + u_x, y + u_y, t) = 0 \quad (1) \end{aligned}$$

De nombreuses methodes permettant d’estimer les mouvements transparents ont ete baties sur cette equation: minimisation d’une fonction d’energie comprenant un terme de regularisation [3], introduction d’une decomposition des vitesses en ondelettes ou sur une base de B-splines [4], utilisation des champs de Markov ou de block-matching [5]. Une autre possibilite est de formuler le probleme dans l’espace des frequences [6], mais il faut alors supposer le mouvement constant sur un intervalle de temps beaucoup plus long.

Cet article est structure comme suit. Le paragraphe 2 developpe la structure de l’estimateur multiresolution propose. La section 3 decrit une importante etape de post-traitement et des resultats experimentaux sont presentes dans la section 4.

2 ESTIMATION DE MOUVEMENT TRANSPARENT

2.1 Contrainte de mouvement transparent

Considerons l’equation (1). Elle n’est correcte que si nous supposons que I_1 se deplace avec la vitesse u non seulement au point $p = (x, y)$ a la date t , mais aussi au point $p + v$ a $t - 1$. Il en va de meme pour I_2 en p et $p + u$. Ceci implique en pratique que (1) ne sera rigoureusement respectee sur une region donnee que si les couches sont en translation.

C’est pourquoi nous choisissons de traiter le probleme de mouvements complexes en divisant l’image en petits blocs, sur lesquels le mouvement est considere translationnel.

L'expression a minimiser sur chaque bloc B_i est donnee par:

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

ou $r(x, y, u^i, v^i)$ est donne par l'equation (1)

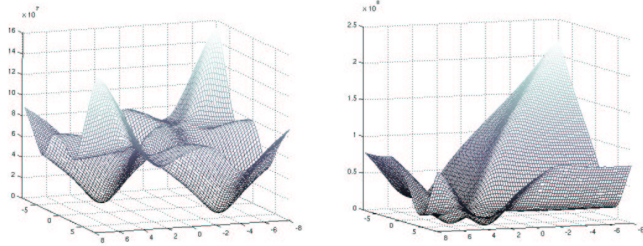


FIG. 1: Deux traces de la fonction $J(u, v)$ reduite au sous-espace (u_x, v_x) . Vraies vitesses: gauche: $u = (3, 0)$, $v = (-4, 0)$, droite: $u = (3, 0)$ et $v = (5, 0)$.

Deux traces de la fonction $J(u, v)$ sont representes sur la Fig.1 afin d'illustrer ses proprietes. Nous avons construit trois images en deplacant deux couches par des vitesses connues orientees dans la direction x . Ainsi, le graphique peut etre limite a l'espace (u_x, v_x) . Notons que ces fonctions presentent deux minima symetriques puisque u et v jouent des roles similaires dans l'equation (1). Cette ambiguïte ne pose pas de probleme pour l'heure car nous ne cherchons pas encore a identifier u et v avec des couches specifiques.

2.2 Estimation multiresolution de mouvements transparents

Considerant trois images successives contenant deux mouvements transparents, nous estimons leur vitesse en minimisant $J(u, v)$. Si les amplitudes des deplacements etaient petites, nous pourrions travailler sur une version linearisee de J et la minimiser avec un algorithme iteratif de minimisation classique. Seulement, puisque nous pouvons rencontrer des mouvements importants, nous utilisons un schema multiresolution base sur les decompositions pyramidales des trois images consecutives. Au niveau de resolution le plus grossier, les deplacements sont assez petits pour permettre une minimisation par les gradients conjuges de la fonction $J(u, v)$ linearisee autour de $(0, 0)$. Nous obtenons ainsi une premiere approximation (u^L, v^L) des deux mouvements. Nous procedons ensuite de la meme maniere au niveau inferieur $L - 1$, mais cette fois en linearisant J autour de $(2.u^L, 2.v^L)$. Cette methode est ensuite iteree le long de la decomposition pyramidale jusqu'a son niveau le plus fin.

Nous avons en fait etabli que les estimates sont plus precis si ce schema de linearisation et minimisation etait repete quelques fois a un meme niveau de resolution.

2.3 Initialisation par l'algorithme du simplex

Cette strategie de minimisation est efficace et rapide, mais est aussi sensible a son initialisation, et ce d'autant plus que nous travaillons avec des images medicales obtenues par rayons X qui sont peu contrastees et tres bruitees.

Nous nous appuyons sur la methode du downhill simplex pour

initialiser correctement notre algorithme d'estimation iterative. Cette technique minimise des fonctions non lineaires dans un espace de dimension quelconque n en deplacant iterativement $n + 1$ points de cet espace. Ainsi, pour minimiser J dans l'espace de dimension quatre (u, v) , elle manipule cinq echantillons de vitesses (u, v) . A chaque iteration, elle remplace le point de l'echantillon correspondant a la plus haute valeur de J par un nouvel echantillon recherche sur une ligne perpendiculaire a l'hyperplan contenant la n autres points [7]. Cet algorithme fournit une initialisation appropriee a notre probleme de minimisation. Les couts de calculs restent faibles puisque nous utilisons l'algorithme du simplex seulement au niveau le plus grossier de la pyramide de resolution.

3 ETAPE DE POSTTRAITEMENT

3.1 Recalage et estimation monomouvement

Remarquons sur les diagrammes de la fonction J sur la Fig.1 que le minimum global peut etre difficile a atteindre quand les vitesses des deux couches sont proches (exemple sur la droite de Fig.1). (D'ailleurs, si les vitesses sont quasiment identiques, le probleme de la transparence disparaît car nous sommes ramenes a une situation avec un mouvement unique!) Pour surmonter ce probleme, nous avons mis au point l'etape de posttraitement suivante.

L'etape d'estimation multiresolution donne les vitesses estimees $(\hat{u}_{me}, \hat{v}_{me})$. Si par exemple \hat{u}_{me} est une bonne estimee, nous pouvons faire disparaître la couche correspondante (disons I_1) implicitement sans l'identifier. Pour cela, nous compensons le mouvement de cette couche avec \hat{u}_{me} afin qu'elle apparaisse statique, et nous formons la difference des images recalees:

$$\begin{aligned} \Delta I_{\hat{u}_{me}, t+1}(x, y, t) &= \tilde{I}^{\hat{u}_{me}}(x, y, t) - \tilde{I}^{2 \cdot \hat{u}_{me}}(x, y, t - 1) \\ &= \tilde{I}_2^{\hat{u}_{me}}(x, y, t) - \tilde{I}_2^{2 \cdot \hat{u}_{me}}(x, y, t - 1) \end{aligned} \quad (3)$$

$$\begin{aligned} \Delta I_{\hat{u}_{me}, t}(x, y, t + 1) &= I(x, y, t + 1) - \tilde{I}^{\hat{u}_{me}}(x, y, t) \\ &= I_2(x, y, t + 1) - \tilde{I}_2^{\hat{u}_{me}}(x, y, t) \end{aligned} \quad (4)$$

avec $\tilde{I}^w(x, y, t) = I(x + w_x, y + w_y, t)$ l'image recallee par w . Alors, $\Delta I_{\hat{u}_{me}, t+1}(x, y, t)$ forme une sequence monocouche se deplacant de $s = \hat{u}_{me} + v$. Sa vitesse peut etre estimee avec une methode d'estimation monomouvement classique, ce qui permet de reestimer v comme $\hat{v}_{pp} = \hat{s} - \hat{u}_{me}$. De meme, nous reestimons \hat{u}_{pp} a partir de $t = u + \hat{v}_{me}$. Ainsi, si une vitesse seulement est correctement estimee, cette methode permet de calculer la seconde.

En resume, nous construisons les images \tilde{I} en recalant par \hat{u}_{me} et \hat{v}_{me} respectivement, nous calculons les images ΔI grace aux equations (3) et (4), nous estimons les vitesses s et t en utilisant la methode d'estimation robuste multiresolution decrite dans [8] (avec un modele de mouvement constant), et nous en deduisons \hat{u}_{pp} et \hat{v}_{pp} .

Il reste alors a choisir les meilleurs estimates parmi les quatre paires possibles $(\hat{u}_{me}, \hat{v}_{me})$, $(\hat{u}_{me}, \hat{v}_{pp})$, $(\hat{u}_{pp}, \hat{v}_{me})$ et $(\hat{u}_{pp}, \hat{v}_{pp})$.

3.2 Strategie de choix

Une strategie intuitive consisterait a selectionner les estimees u et v donnant la plus petite valeur par $J(u, v)$. De fait, cela donne de bons resultats sur des images non bruitees, mais ne fonctionne plus du tout sur des images bruitees.

En effet, l'expression de J est formee de la somme algebrique de quatre images. Si les images originales sont entachees d'un bruit de variance σ^2 , J est affecte par un bruit de variance $4.\sigma^2$, ce qui rend le choix tres peu fiable.

Nous proposons d'appliquer alors un schema de regularisation. Nous souhaitons privilegier au niveau des blocs les vitesses qui minimisent raisonnablement (2), tout en etant coherentes avec les vitesses des blocs voisins.

Pour cela, nous avons utilise un processus d'estimation alterne. Dans un premier temps, nous calculons les deux champs affines de mouvement (un pour chaque couche) compatibles avec mouvements estimes sur chaque bloc, puis nous choisissons sur chaque bloc celui des quatre couples de vitesses candidates qui est le plus proche du champ de mouvement estime.

Sur chaque bloc, les deux vecteurs deplacements correspondant aux champs de mouvements estimes sont calcules et compares (en utilisant la norme L_2) aux quatre paires de candidats, et l'on choisit la paire qui en est la plus proche.

La aussi, cet algorithme iteratif fonctionne bien pour peu qu'il soit correctement initialise. Les experiences que nous avons menees nous ont conduit a initialiser les champs de vecteurs sur les blocs par $(\hat{u}_{me}, \hat{v}_{me})$. Ces vitesses sont avant tout regroupees en deux groupes $\{\hat{u}_{me}\}$ et $\{\hat{v}_{me}\}$ supposes correspondre aux deux vecteurs vitesses a l'aide d'une technique de clustering. Cette operation est indispensable pour s'assurer que u et v correspondent bien a la meme couche de bloc a bloc. La structure globale proposee est resumee sur la Fig.2.

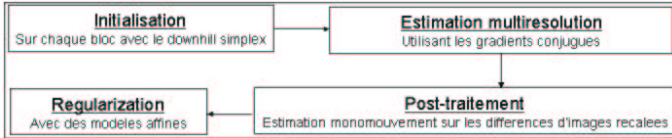


FIG. 2: Structure globale de la methode d'estimation de mouvements transparents.

4 RESULTATS EXPERIMENTAUX

Nous avons synthetise des sequences de trois images contenant deux couches en mouvement transparent a partir d'images reelles afin d'evaluer la performance de la methode. Pour ce faire, nous avons additionne deux images (correspondant aux deux couches) deplacees de maniere translationnelle ou affine. Les mouvements ont ete tires aleatoirement pour assurer une amplitude maximale de deplacement de 8 pixels entre deux images. Nous avons simule deux types de situations, en utilisant d'abord des images video classiques, puis des images issues d'examen par rayons X, plus precisement des images fluoroscopiques de la region du coeur.

Pour simuler le bruit temporel present dans des sequences reelles, nous avons egalement ajoute a nos images un bruit gaussien independant d'ecart type compris entre 0 et 20.

TAB. 1: Erreur d'estimation moyenne en pixel pour differents types d'images, niveaux de bruit et versions de l'esimateur (voir texte principal).

Type	Video			Rayons X		
Bruit	0	10	20	0	10	20
ME	0.12	0.60	1.05	0.23	2.48	6.09
PP	0.14	0.79	1.61	0.20	3.63	7.45
RE	0.11	0.48	0.89	0.19	1.83	4.34
PR	0.12	0.19	0.33	0.27	0.93	2.42

Dans les experiences presentees ici, une des deux couches avait un deplacement affine et l'autre subissait une translation. La duree totale de l'estimation etait d'environ 10 secondes pour des images 288*288 sur un Pentium IV a 2.4 GHz avec 1Go de memoire. Tab.1 presente les erreurs d'estimation moyennes sur tous les blocs effectuees sur les vecteurs vitesses pour 250 generations aleatoires d'images.

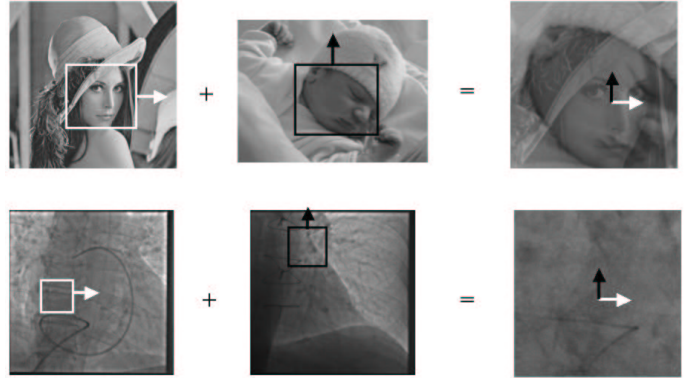


FIG. 3: Generation d'images transparentes, a la fois dans les cas video et cliniques.

Nous avons compare quatre versions de notre algorithme pour evaluer les contributions des differents modules. *ME* n'accomplit que les deux premieres etapes: initialisation et estimation multiresolution. *PP* passe aussi par l'etape de post-processing mais retient la meilleure paire de vitesses selon les valeurs de la fonction $J(u, v)$ seulement, sans accomplir de regularisation. *RE* effectue la regularisation et garde les couples selectionnes sur chaque bloc, tandis que *PR* va lui aussi au terme de l'algorithme mais il ne retient que les deux champs de vecteurs affines estimes (ou plus precisement leur vecteur moyen sur chaque bloc).

Nous remarquons sur les images video que l'etape de post-traitement ameliore tres sensiblement la qualite des estimes mais qu'il ne suffit pas de se baser sur J pour selectionner les couples de candidats des lors que les images sont bruitees. La methode proposee dans son ensemble donne par contre des resultats tres satisfaisant dans tous les cas, mais la strategie la meilleure consiste a choisir les vecteurs vitesses correspondant aux champs affines estimes. Soulignons cependant que nos experiences constituent un cas particulier dans la mesure ou ces modeles de mouvements correspondent exactement a la realite des images, ce qui ne serait pas le cas avec des images reelles. Ces experiences ont ete egalement menees sur des images obtenues par rayon X. Les resultats se trouvent dans Tab.1

egalement.

Nous pouvons effectuer les meme commentaires en ce qui concerne l'apport de chaque etape de notre methode, mais la precision obtenue ici diminue avec le niveau de bruit. Cependant, les mouvements deviennent peu perceptibles dans ce cas.

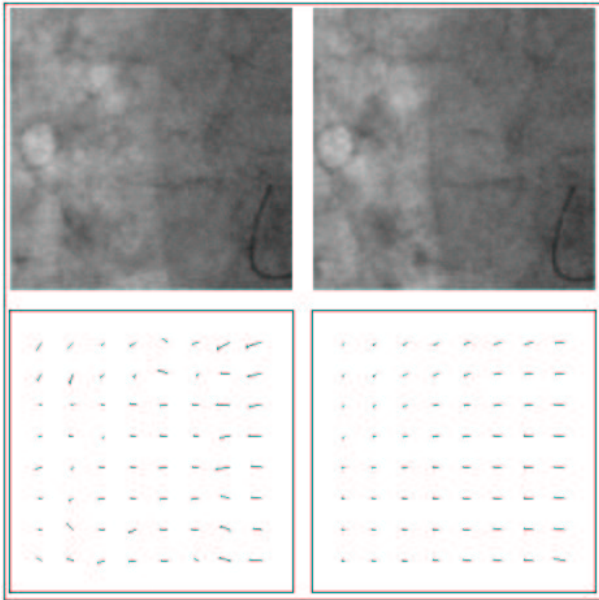


FIG. 4: Haut: Images numero 1 et 8 de la sequence. Bas: Vitesses estimees sur les trois premieres images. Gauche: Estimation finale sur chaque bloc utilisant l'algorithme *RE* (deux vecteurs sont representes par bloc), droite: les deux modeles affines estimees (algorithme *PR*).

Nous avons egalement applique nos resultats a des images medicales reelles. Nous presentons nos resultats sur une sequence cardiaque fluoroscopique acquise a 30 Hz. Elle represente un zone d'environ 5cm*5cm a droite du coeur. Deux images de cette sequence sont reproduites sur Fig.4, ainsi que les mouvements transparents calcules.

Le coeur (qui apparait sombre ici) bat sur la droite de l'image sur un fond fixe correspondant aux cotes et a la colonne vertebrale. Les tissus clairs des poumons a 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 affectees par un bruit important ($\sigma \simeq 20$).

Nos estimations sont coherentes avec les mouvements observes sur la sequence. Deux mouvements ont ete estimees, et ils correspondent a la verite anatomique: le fond, qui est statique (et n'apparait donc pas sur les champs de fleches de Fig.4) et le mouvement (affine) du coeur. L'amplitude du mouvement affine estime decroit bien dans les poumons avec la distance au coeur. Ces images sont bruitees, peu contrastees, et contiennent des mouvements complexes. En outre, les mouvements ne sont pas parfaitement constants sur trois images successives. Meme dans ce cas difficile, la methode d'estimation proposee donne des resultats interessants.

5 CONCLUSION

Nous avons mis au point une nouvelle methode complete d'estimation multiresolution de mouvements transparents pour

trois couches. Il contient en particulier une initialisation efficace et un posttraitement original qui ameliorent grandement la qualite de l'estimation multiresolution seule. Nous avons montre sur un grand nombre de simulations que nous atteignons une precision de l'ordre du quart de pixel sur des images non bruitees. Nous avons egalement etudie l'estimation sur images bruitees, et nous obtenons des resultats satisfaisants pour des niveaux de bruit temporel raisonnable, y compris pour des images cliniques peu contrastees. Nous avons egalement produit un exemple sur donnees reelles qui a donne des resultats encourageants.

Nous allons mener plus d'experiences sur des images reelles acquises par rayons X, et egalement etudier les cas contenant plus de deux couches.

Références

- [1] M. Irani, B. Rousso, and S. Peleg, "Computing occluding and transparent motions," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 5–16, 1994.
- [2] M. Shizawa and K. Mase, "Simultaneous multiple optical flow," in *Int. Conf. on Pattern Recognition*, Atlantic City, New Jersey, 1990, vol. 1, pp. 274–278.
- [3] J. Toro, F. Owens, and R. Medina, "Multiple motion estimation and segmentation in transparency," in *Proc. of the IEEE Int. Conference on Acoustics, Speech and Signal Processing*, Istanbul, June 2000, pp. 2087–2090.
- [4] M. Pingault, E. Bruno, and 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, and 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, California, Jan 2004, pp. 274–278.
- [6] M. Pingault and D. Pellerin, "Motion estimation of transparent objects in the frequency domain," *Signal Processing*, vol. 84, no. 4, pp. 709–719, April 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 and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *Journal of Visual Communication and Image Representation*, vol. 6, no. 4, pp. 348–365, Dec. 1995.