

Débruitage spatio-temporel de séquences d'images fluoroscopiques

C AMIOT^{1,2}, C GIRARD², J PESCATORE², J CHANUSSOT¹; M DESVIGNES¹

¹Univ. Grenoble Alpes, GIPSA-Lab,
F-38000 Grenoble France

²Thales Electronic Device
Moirans, France

carole.amiot@gipsa-lab.grenoble-inp.fr, catherine.girard@gmail.com,
jeremie.pescatore@thalesgroup.com, jocelyn.chanussot@gipsa-lab.grenoble-inp.fr,
michel.desvignes@gipsa-lab.grenoble-inp.fr

Résumé – Nous proposons un filtre spatio temporel dans le domaine des ondelettes dual-tree pour débruiter des séquences fuloroscopiques à très faible dose. Les propriétés de l'espace multi-échelle assurent une détection rapide et fiable des objets en mouvement. Le filtre temporel utilise cette détection pour débruiter efficacement tant les objets fixes que les structures mobiles. Une analyse qualitative et quantitative des résultats sur des séquences synthétiques et réelles montre que cette méthode est un peu supérieure à l'état de l'art et autorise une réduction de dose lors de l'acquisition de 80%.

Abstract – We propose a temporal motion-compensated filter operating on dual-tree complex wavelet coefficients to denoise low-dose X-ray sequences. This filter allows for a great noise reduction while preserving moving objects and structures. We take advantage of the properties of multi-scale spaces to perform a fast and robust motion tracking. The result of this step is then used in the temporal filter for denoising purpose. Quantitative and qualitative analysis on real fluoroscopic sequences show that our method outperforms state-of-the-art VBM3D method and allows for a dose reduction as high as 80%.

1 Introduction

Les séquences fluoroscopiques sont utilisées lors d'examen médicaux dits minimalement invasifs. Les objets médicaux sont introduits dans une artère et guidés jusqu'au site à soigner. Le guidage s'effectue à l'aide de séquences d'images rayons X acquises à 15 ou 30 images/secondes. Pour éviter une trop forte exposition aux rayons X qui entraîne de nombreuses complications pour la santé du patient et du personnel médical, il faut réduire les doses de rayons X émises lors de ces examens tout en conservant la qualité visuelle des images. Diminuer la dose a pour conséquence d'augmenter le bruit présent et de dégrader le contraste des images, ce qui peut être corrigé par des outils performants de débruitage d'images. Les contraintes fortes à respecter sont les suivantes :

- les contours anatomiques et les objets médicaux doivent être préservés, tant en position, forme ou contraste.
- aucun artefact spatial ou temporel, en particulier de retard temporel, ne doit être introduit
- l'implémentation en temps réel doit être possible

Pour améliorer la qualité des séquences fluoroscopiques, la plupart des travaux antérieurs travaillent dans un espace multi-échelle [1]-[5]. [1] offre de bonnes performances, mais n'effectue pas de compensation de mouvements, ce qui la rend peu adaptée aux séquences présentant de forts mouvements comme les séquences cardiaques ou abdominales. [2] et [3] basent

le débruitage sur des moyennes non locales. VBM3D [2] est la méthode de référence donnant les meilleurs résultats, mais elle a tendance à supprimer les objets de faible contraste en mouvement. [5] utilise une transformation 3D multi-échelle, qui n'est cependant adaptée qu'au mouvement uniforme. Dans l'approche SURELET [6], le débruitage est exprimé comme une combinaison linéaire de fonctions de seuillage. Pour des vidéos, elle utilise une compensation de mouvement par block-matching et présente des résultats aussi bons que VBM3D.

La plupart des méthodes de débruitage travaillent dans un espace multi-échelle car la représentation parcimonieuse permet souvent de mieux séparer le bruit et le signal [7]. Nous avons développé un filtre spatio-temporel utilisant un espace multi-échelle à la fois pour la détection de mouvement et pour le débruitage tant spatial que temporel. L'approche que nous proposons, outre un débruitage spatial, effectue une détection des objets en mouvement sur les coefficients multi-échelles des images successives, puis apparie les coefficients des objets détectés entre 2 images. Le filtrage temporel dépend alors de cette mise en correspondance, afin de lisser différemment les objets fixes, les objets en mouvement, les objets disparaissant ou apparaissant. Nous décrivons dans un premier temps la transformation multi-échelle choisie, puis la méthode de détection des objets en mouvement suivie de la mise en correspondance et du filtrage temporel. Nous comparons enfin les résultats sur des séquences réelles à VBM3D.

2 Méthode proposée

Nous proposons un filtre temporel compensé en mouvement travaillant sur les coefficients de l'espace multi-échelle. Le principal avantage provient de l'étape de compensation du mouvement. Les objets sont en effet mieux détectés dans cet espace à des forts niveaux de bruit et avec des contrastes faibles.

2.1 Les ondelettes Dual-Tree

La transformation en ondelettes Dual-Tree (DT-CW) a été introduite par [8] pour suppléer les problèmes d'orientation, d'invariance et d'oscillations introduites par les ondelettes. Elle utilise des ondelettes complexes composées de 2 ondelettes réelles formant une paire de Hilbert. Cette transformation représente bien les formes curvilinéaires telles que contours anatomiques et objets médicaux. L'approche proposée est cependant indépendante de la transformation et a aussi été utilisée avec les curvelets. La suite du papier utilise les notations suivantes : J indique les échelles dans le domaine transformé DT-CW, J variant entre 1 et $N + 1$, N est le nombre d'échelles (6 par la suite). Les hautes fréquences correspondent aux faibles valeurs de J .

2.2 Détection dans l'espace DT-CW

La détection est réalisée pour chaque orientation, mais sur une carte regroupant les informations des différentes échelles pour améliorer le rapport signal à bruit et minimiser les temps de calcul [9]. Cette carte est une somme pondérée des coefficients de même position aux différentes échelles. La première échelle, ne comportant presque que du bruit, n'est pas utilisée car elle dégrade les performances de la méthode. Nous utilisons ensuite des filtres orientés tournants [10] pour détecter les structures linéaires présentes dans ces cartes. Comme chaque carte correspond à une orientation de la décomposition, seuls 2 filtres d'orientations différentes sont utilisés par cartes, la plus forte réponse étant conservée. Les structures linéaires détectées par un seuillage par hystérésis sont segmentées en clusters [11].

2.3 Mise en correspondance temporelle

Les coefficients détectés précédemment correspondant à des objets en mouvement sur 2 images successives sont regroupés en clusters. Le suivi temporel de ces clusters est assuré par la mise en correspondance des clusters. Elle est réalisée par la recherche du couplage maximum d'un graphe biparti. Les sommets du graphe sont les clusters détectés à t et $t - 1$, les arêtes du graphe relient les clusters détectés sur 2 images consécutives. Les poids de ces arêtes représentent la similarité entre 2 clusters t et $t - 1$. Le couplage minimise le cout total des arêtes retenues et les arêtes retenues indiquent la mise en correspondance entre les clusters à t et $t - 1$.

La similarité est basée sur les variations de position, longueur et énergie des clusters. La caractéristique la plus impor-

tante est la préservation de l'énergie des clusters, due à la relation de Parseval que respecte la transformée DT-CW. Comme elle n'est pas complètement invariante, chaque coefficient n'est pas préservé et seule l'énergie moyenne des clusters est préservée.

Soit u un cluster détecté au temps t et v celui détecté au temps $t - 1$, la fonction de cout à minimiser est la somme de :

$$\Delta_L(u, v) = 1 - \frac{\min(L_u, L_v)}{\max(L_u, L_v)}$$

$$\Delta_d(u, v) = \frac{1}{2}(\alpha_x dist_x(u, v) + \alpha_y dist_y(u, v))$$

$$\Delta_E(u, v) = 1 - \frac{\min(E_u, E_v)}{\max(E_u, E_v)}$$

avec L_u est la longueur du cluster u , E_u est son énergie. $dist_x(u, v)$ and $dist_y(u, v)$ sont les distances selon les axes x and y entre les clusters u et v . Les paramètres α_x et α_y normalisent les distances entre 0 and 1. Typiquement, leur valeur est égale au déplacement maximal autorisé entre 2 images, avec $\alpha_x = \frac{8}{m}$ et $\alpha_y = \frac{8}{n}$, où m et n sont les dimensions de l'image. Un déplacement d'un huitième de l'image est un déplacement correspondant à des mouvements rapides assez peu fréquents.

L'orientation intervient de manière binaire dans ce cout. Le cout n'est calculé que pour les clusters dont la variation d'orientation est inférieure à un seuil correspondant à 60 degrés.. Un cout maximal autorisé évite finalement la mise en correspondance de clusters trop distincts.

La fonction de cout globale est minimisée par l'algorithme itératif de Kunh-Munkres [13] qui trouve le cout minimum global et la meilleure mise en correspondance globale. Chaque détection au temps t est alors associée au plus à une détection au temps $t - 1$ et réciproquement.

2.4 Filtrage temporel

Le filtre temporel est un filtre à réponse impulsionnelle infinie du premier ordre. Il est donc rapide à calculer et ne requiert que 2 images en mémoire. Ce filtre réduit le bruit gaussien en réalisant une moyenne pondérée des coefficients dans l'espace multi-échelle de ces 2 images. La pondération dépend de la nature des coefficients multi-échelle (objets détectés, fixes, en mouvement, etc...).

Après une transformation d'Anscombe généralisée [14], les parties réelles et imaginaires des coefficients multi-échelles des images ont des distributions gaussiennes. En présence de mouvement, seule la norme des coefficients est préservée et le filtrage doit être réalisé sur le module dans ce cas. Pour éviter les artefacts entre les zones fixes et mobiles, le filtrage de tous les coefficients, avec ou sans mouvement se base sur leur module.

Dans ce cas, les parties réelles et imaginaires des coefficients de bruit suivant une distribution gaussienne de moyenne nulle, la norme suit une distribution de Rayleigh, de moyenne non nulle. Un moyennage pondéré conduira ainsi à une distribution de moyenne non nulle, égale à $\sqrt{\frac{\pi}{2}}$, alors que nous souhaitons que les coefficients de bruit doivent tendre vers 0 lorsque l'on

débruite l'image. Nous modifions donc les coefficients pour imposer une moyenne nulle à la distribution résultante par :

$$|\hat{c}_{i,t}| = (1 - \alpha)|\hat{c}_{i,t-1}| + \alpha(|c_{i,t}| - \sigma_{sb}\sqrt{\frac{\pi}{2}})_+ \quad (1)$$

où $\hat{c}_{i,t-1}$ est le coefficient filtré temporellement à $t - 1$, $c_{i,t}$ est le coefficient non filtré à t .

Filtrage Après détection et mise en correspondance, nous distinguons 4 cas de filtrage selon que les coefficients :

1. n'appartiennent pas à un cluster détecté (fixes)
2. appartiennent à un cluster détecté mais ne sont pas mis en correspondance (apparition d'un objet)
3. appartiennent à un cluster détecté et sont mis en correspondance (objet en mouvement)
4. sont détectés et mis en correspondance avec un cluster précédent (disparition d'un objet à cet endroit)

Le premier cas est celui des coefficients fixes, débruités simplement à l'aide de l'équation (1).

Le deuxième cas est celui des fausses détections de mouvement ou des objets apparaissant dans la séquence. Le débruitage est alors une moyenne pondérée des valeurs filtrées spatialement et temporellement par :

$$|\hat{c}_{i,t}| = \alpha_{ts}|\tilde{c}_{i,t}| + (1 - \alpha_{ts})|c_{i,t}^{temp}| \quad (2)$$

$|c_{i,t}^{temp}|$ est le coefficient filtré temporellement (eq 1), $|\tilde{c}_{i,t}|$ est le coefficient filtré spatialement par [11]. Le poids α_{ts} est fonction inverse de l'énergie du coefficient, car un nouvel objet présente à priori une forte énergie au contraire d'un coefficient de bruit.

Le troisième cas correspond à des clusters en mouvement. Comme il y a conservation de l'énergie du cluster et non de chaque coefficient, l'énergie moyenne du cluster débruité est évaluée et les coefficients modifiés par :

$$\alpha_{DP} = \alpha E_t + (1 - \alpha)E_{t-1} \quad (3)$$

$$|\hat{c}_{i,t}| = \alpha_{DP}(|c_{i,t}| - \sigma_{sb}\sqrt{\frac{\pi}{2}})_+ \quad (4)$$

α_{DP} est l'énergie moyenne du cluster débruité, E_t est l'énergie au temps t et E_{t-1} est l'énergie du cluster associé débruité au temps $t - 1$.

Le quatrième cas est celui de l'ancienne position d'un objet en mouvement. Pour ne pas introduire d'artefacts de retard, le filtrage est uniquement spatial par [11].

3 Résultats et discussion

3.1 Evaluation de la mise en correspondance

La précision de la mise en correspondance est mesurée par l'indicateur Multiple Object Tracking Accuracy (MOTA) [15]. Les fausses détections et mauvaises mises en correspondance font décroître l'indicateur vers 0. La référence est une séquence

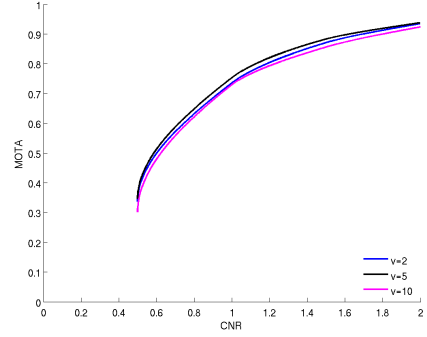


FIGURE 1 – Précision de la mise en correspondance pour 3 types de mouvements.

d'images de synthèse fluoroscopiques à 15nGy contenant un cercle de CNR donné et de rayon croissant dans le temps (2,5,10 pixels/images) correspondant à des vitesses faible, moyenne et forte. Le mouvement simule les déformations et le déplacement rapide des objets.

Pour des $CNR < 1$ (contraste très faible), la faiblesse de l'indicateur est due à des mauvaises détections. Dès que le CNR augmente un peu, le cercle est bien détecté et mis en correspondance. Globalement, très peu de détections ou mises en correspondance sont manquées. Le résultat important est qu'il n'y a pas de dégradations des performances lorsque l'amplitude du mouvement augmente.

3.2 Séquences fluoroscopiques

La séquence abdominale acquise sur un panneau Thales Pixium RF 4343 présente un mouvement global de translation de la table d'opération et des mouvements "libres" des organes abdominaux. Un guide manipulé par l'équipe médicale présente un fort mouvement. Acquisée à 30nGy, des séquences à 15 et 5nGy ont été simulées en ajoutant un bruit poissonien. Les résultats sont comparés à VBM3D (Figure 2).

A la dose habituelle de 30nGy, les 2 méthodes donnent des résultats identiques. La figure 3 montre la différence entre la version bruitée et débruitée des séquences. Pour un débruitage efficace, ces images ne doivent pas comporter de structures visibles. A 30nGy, ces images sont similaires, alors qu'à 5nGy, VBM3D ne peut préserver les structures en mouvement rapide telles que le guide. L'étape de block-matching de VBM3D ne peut sans doute pas réaliser une mise en correspondance à faible dose. Sur les autres images de la séquence, les structures en mouvement sont bien préservées par notre méthode, qui détecte et met en correspondance efficacement les objets. Le filtrage réalisé à 5nGy répond bien aux exigences fixées.

4 Conclusion

Nous avons proposé un filtrage spatio-temporel compensé en mouvement qui permet une réduction de dose de l'ordre de 80%. Les expériences sur des données réelles et simulées

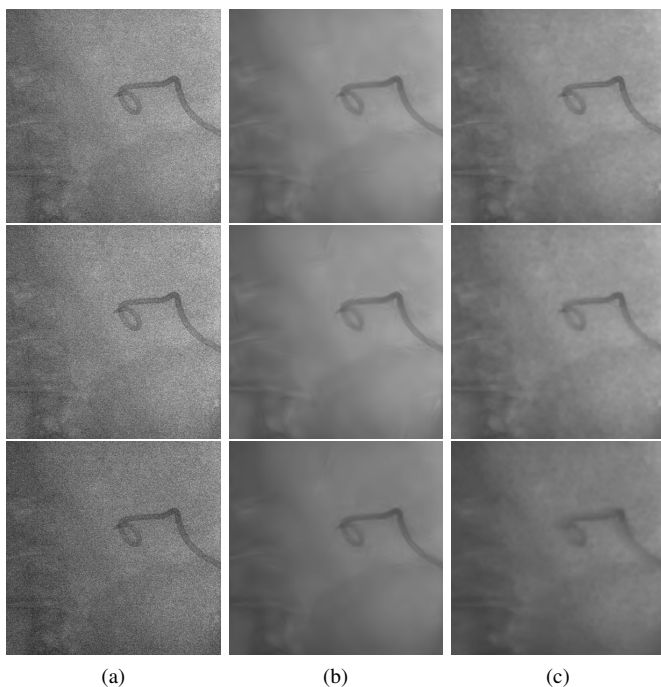


FIGURE 2 – Dernière image à 30nGy (haut), 15nGy (milieu), 5nGy (bas), séquence bruitée (a) débruitée par notre méthode (b) et VBM3D (c).

montrent qu'il est meilleur que VBM3D, considérée comme une des meilleures méthodes actuelles. Le temps actuel de traitement pour une image de taille 1024×1024 pixels est de 3.2s, mesuré sur un processeur à 2.67GHz avec une implémentation Matlab. Une implémentation parallèle sur GPU devrait permettre d'atteindre les cadences souhaitées.

Références

- [1] Tomic, M. and Loncaric, S. and Sersic, D. *Adaptive spatio-temporal denoising of fluoroscopic X-ray sequences*. Biomedical Signal Process. and Control, 7,(2), pp 173-179, 2012.
- [2] Dabov, K. and Foi, A. and Egiazarian, K., *Video denoising by sparse 3D transform-domain collaborative filtering*, EUSIPCO, 2007.
- [3] Gooßen, André and Pralow, Thomas and Grigat, Rolf-Rainer, *Medical X-ray image enhancement by intra-image and inter-image similarity*, SPIE Medical Imaging, pp 72590G–72590G, 2009.
- [4] Bismuth, V. and Vaillant, R., *A device enhancing and denoising algorithm for X-ray cardiac fluoroscopy*, Int. Conf. on Pattern Recognition, pp1-4, 2008,
- [5] Selesnick, I. and Li, K., *Video denoising using 2D and 3D dual-tree complex wavelet transforms*, SPIE Optical Science and Technology, pp 607–618, 2003.
- [6] F. Luisier, T. Blu, and M. Unser, *SURE-LET for Orthogonal Wavelet-Domain Video Denoising*, IEEE Transac-

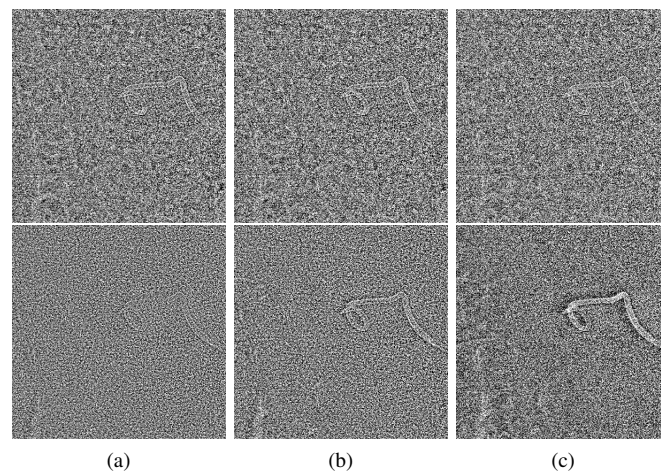


FIGURE 3 – Différence entre la versions bruitée et débruitée par notre méthode (haut) et par VBM3D (bas) à 30nGy (a), 15nGy (b), 5nGy (c).

- tions on Circuits and Systems for Video Technology, vol. 20, no. 6, pp. 913-919, 2010
- [7] Starck, J.-L. and Fadili, J., *An overview of inverse problem regularization using sparsity*, 16th IEEE Int. Conf. on Image Processing, pp 1453-1456, 2009.
- [8] Selesnick, I. and Baraniuk, R. and Kingsbury, N. *The dual-tree complex wavelet transform*, IEEE Sig. Process. Mag., Vol 22, n 6 ; pp 123-151, 2005.
- [9] Amiot, C. and Pescatore, J. and Chanussot, J. and Desvignes, M *Contextual filtering in curvelet domain for fluoroscopic sequences*, SPIE Med. Imaging, pp 86690N–86690N, 2013,
- [10] Bismuth, V. and Vancamberg, L. and Gorges, S. *A comparison of line enhancement techniques : applications to guide-wire detection and respiratory motion tracking*, SPIE Medical Imaging, pp 72591M–72591M,2009
- [11] Amiot, C. and Girard, C. and Chanussot, J. and Pescatore, J. and Desvignes, M, *Curvelet based Contrast Enhancement in Fluoroscopic Sequences*, IEEE Trans. on Med. Im., vol 65, n 2, pp, 2015.
- [12] Starck, J.-L. and Murtagh, F. and Fadili, J., *Sparse image and signal processing : wavelets, curvelets, morphological diversity*, Cambridge University Press, 2010.
- [13] Munkres, J. *Algorithms for the assignment and transportation problems*, Journal of the Society for Industrial & Applied Mathematics, vol 5, n 1, pp 32-38, 1957.
- [14] Makitalo, M. and Foi, A., *Optimal Inversion of the Generalized Anscombe Transformation for Poisson-Gaussian Noise*, IEEE Transactions on Image Processing, 22, 1, pp 91-103, 2013.
- [15] Keni, B. and Rainer, S., *Evaluating multiple object tracking performance : the CLEAR MOT metrics*, EURASIP Journal on Image and Video Process.,vol 1, pp 1-10, 2008.