

Mosaïquage d’images cystoscopiques : étude comparative de méthodes de recalage

Achraf BEN HAMADOU, Charles SOUSSEN, Walter BLONDEL, Christian DAUL, Didier WOLF

Centre de Recherche en Automatique de Nancy (CRAN), UMR 7039 CNRS-Nancy-Université
2, Avenue de la forêt de Haye, F-54516 Vandœuvre-lès-Nancy, France

Achraf.Ben-Hamadou@ensem.inpl-nancy.fr, Charles.Soussen@cran.uhp-nancy.fr,
Walter.Blondel@ensem.inpl-nancy.fr, Christian.Daul@ensem.inpl-nancy.fr,
Didier.Wolf@ensem.inpl-nancy.fr

Résumé – Le cancer de la vessie est le onzième cancer le plus répandu dans le monde. L’examen cystoscopique est la procédure de référence pour l’exploration visuelle de la paroi interne de la vessie. Chaque image de la séquence d’images acquise lors de cet examen correspond à la visualisation d’une surface de tissu de l’ordre du cm^2 . Cependant, les lésions sont généralement étendues sur plusieurs images. Cette contribution présente une étude comparative de deux méthodes de construction d’images panoramiques de la paroi interne de la vessie en termes de robustesse, de précision et de vitesse de calcul. Cette étude est effectuée sur des données réelles extraites d’examen cystoscopiques cliniques et sur des données acquises pour un fantôme de vessie en contrôlant le déplacement 3D de l’endoscope.

Abstract – Bladder cancer is widely spread in the world. Many adequate diagnosis techniques exist. Video-endoscopy remains the standard clinical procedure for visual exploration of the bladder internal surface. However, video-endoscopy presents the limit that the imaged area for each image is about 1 cm^2 . However, lesions are typically spread over several images. The aim of this contribution is to assess the performance of two mosaicing algorithms leading to the construction of panoramic maps (one unique image) of bladder walls. The quantitative comparison study is performed, on the one hand, on images extracted from clinical examination data and, on the other hand, on data acquired from a bladder phantom under controlled 3D endoscope displacements.

1 Introduction

La cystoscopie est l’examen exploratoire de référence pour la détection clinique et le diagnostic du cancer vésical. Cet examen permet l’investigation visuelle de la paroi interne de la vessie à l’aide d’un cystoscope (endoscope) rigide ou fibre. Le seul support visuel disponible est la séquence vidéo cystoscopique et les impressions papier de certaines images d’intérêt pour le clinicien. Chaque image correspond, en général, à la visualisation d’une surface de tissu de l’ordre du cm^2 . Par ailleurs, retrouver une image d’intérêt pour le diagnostic au sein d’une séquence de plusieurs minutes est relativement fastidieux pour le clinicien. Pour résoudre ce problème, nous proposons la construction d’un nouveau support permettant au clinicien d’observer, en une seule fois, des surfaces de vessie plus étendues. Cette approche doit faciliter la localisation des lésions et le suivi des patients d’un examen à un autre. Ce type de support peut être obtenu par mosaïquage d’image, technique qui permet la construction d’images panoramiques en superposant des images successives par recalage. Dans la littérature, le mosaïquage d’images médicales reste un sujet peu abordé (voir [1, 2, 3, 4] pour les applications du mosaïquage en mammographie, angiographie X, ophtalmologie et microscopie). Les solutions proposées qui ne traitent qu’un nombre limité d’images, ne sont pas totalement automatiques et néces-

sitent en général des informations connues *a priori*; ces développements ne sont pas adaptés à la cystoscopie. En effet, le mosaïquage de la paroi interne de la vessie pose des problèmes spécifiques. Premièrement, les images ne contiennent pas de primitives détectables et présentent une grande variabilité inter- et intra-patient en termes de textures. Deuxièmement, il est impossible d’avoir des informations sur la position ou le mouvement du cystoscope dans la mesure où le clinicien le déplace librement. Finalement, la séquence d’images cystoscopiques comporte un nombre très important d’images (plusieurs milliers). Dans cette contribution, nous nous focalisons sur le processus de recalage d’images (qui est au cœur des méthodes de mosaïquage) afin de présenter une étude comparative en terme de robustesse, de précision et de rapidité de deux algorithmes que nous avons développés pour le mosaïquage de la paroi interne de la vessie humaine.

2 Recalage d’images cystoscopiques

Soient deux images successives I_k (image cible) et I_{k+1} (image source), où k est l’indice de l’image dans la séquence vidéo. Le recalage de I_k et I_{k+1} consiste à trouver la transformation perspective 2D/2D $(x', y') = T(x, y; \theta_k)$ qui les superpose au mieux, avec (x, y) les coordonnées d’un point de

l'image I_{k+1} , (x', y') les coordonnées du point (x, y) transformé dans le système de coordonnées de l'image cible I_k et θ_k l'ensemble des paramètres de transformation perspective a_{ij} tels que :

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{1}{w} \begin{bmatrix} u \\ v \end{bmatrix}, \text{ avec } \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} f \cos(\phi) & -S_x \sin(\phi) & t_x \\ a_{11} & a_{12} & a_{13} \\ S_y \sin(\phi) & f \cos(\phi) & t_y \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1)$$

Dans l'équation (1), (t_x, t_y) désigne la translation dans le plan image, f est un facteur d'échelle, ϕ est l'angle de rotation dans le plan, (S_x, S_y) sont les paramètres de cisaillement, (a_{31}, a_{32}) sont les paramètres de la perspective et $a_{33} = 1$.

Le recalage des images I_k et I_{k+1} est effectué via la maximisation/minimisation d'un critère \mathcal{S} de similarité/dissimilarité :

$$\theta_k = \operatorname{argopt}_{\theta \in \mathbb{R}^8} \mathcal{S}(I_k, T(I_{k+1}; \theta)). \quad (2)$$

La mesure de similarité/dissimilarité et l'algorithme d'optimisation doivent être choisis en fonction de scène. Les images cystoscopiques ne contiennent pas de primitives d'images (contours, coins, *etc*) segmentables d'une manière systématique [5]. Pour cette raison, aucune primitive ne peut être exploitée pour guider le recalage. Par conséquent, le calcul de la mesure de similarité/dissimilarité doit se baser sur l'intensité des pixels. Deux algorithmes ont été développés en tenant compte de cette observation.

2.1 A_{DQ} : Algorithme basé sur la distance quadratique

L'algorithme A_{DQ} [5] utilise la mesure de dissimilarité \mathcal{S}_{DQ} définie comme la somme des distances quadratiques entre les niveaux de gris des pixels de l'image I_k et ceux des pixels homologues de l'image transformée $T(I_{k+1}; \theta_k)$:

$$\mathcal{S}_{DQ} = \sum_{(x,y) \in I_k \cap T(I_{k+1})} [I_k(x, y) - I_{k+1}(T(x, y; \theta))]^2 \quad (3)$$

où (x, y) sont les coordonnées des pixels commun entre les images I_k et $T(I_{k+1}; \theta)$. La minimisation \mathcal{S}_{DQ} est réalisée par une méthode itérative basée sur l'algorithme de Baker et Matthews [6] qui estime le flot optique entre les images à recalcer.

2.2 A_{IM} : Algorithme basé sur l'information mutuelle

L'algorithme A_{IM} est basé sur l'approche EMMA [8] (*Empirical entropy Manipulation and Analysis*) de Viola et Wells. A_{IM} aligne les images I_k et I_{k+1} en maximisant l'information mutuelle entre I_k et $T(I_{k+1}; \theta)$. L'information mutuelle est une mesure statistique qui combine les entropies des niveaux de gris $H(I_k)$ et $H(T(I_{k+1}; \theta))$ des parties communes

des images I_k , $T(I_{k+1}; \theta)$ et l'entropie jointe $H(I_k, T(I_{k+1}; \theta))$:

$$\mathcal{S}_{IM} = H(I_k) + H(T(I_{k+1}; \theta)) - H(I_k, T(I_{k+1}; \theta)) \quad (4)$$

La maximisation de \mathcal{S}_{IM} est réalisée à l'aide d'un gradient descendant stochastique (les pixels dont les niveaux de gris sont utilisés pour estimer l'information mutuelle sont tirés au sort à chaque itération).

Les méthodes d'optimisation des algorithmes A_{DQ} et A_{IM} étant itératives, elles sont supposées avoir convergé quand la variation de la transformation T est inférieure d'une itération à la suivante à un certain seuil conduisant à un déplacement inférieur à 1 pixel.

3 Étude comparative

Dans cette section, nous comparons les résultats obtenus avec les deux algorithmes A_{DQ} et A_{IM} appliqués à des images d'examen cystoscopiques réels et des images acquises pour un fantôme réaliste en contrôlant des déplacements 3D d'un endoscope (voir figure 1 (a)).

3.1 Robustesse

Pour évaluer la robustesse du recalage, nous avons utilisé trois images extraites de trois séquences cystoscopiques cliniques. Les images choisies présentent des niveaux d'illumination et de texture très différents. Ces images correspondent aux images références (cibles) I_0 . Les images sources I_1 sont obtenues en simulant un mouvement 3D du cystoscope. Pour ce faire, ces images I_1 sont calculées en appliquant aux images références I_0 une transformation T^{-1} connue. La transformation T à trouver inclue trois translations (deux translations t_x et t_y et une troisième translation t_z correspondant au facteur d'échelle f dans l'équation (1)) et trois rotations (une rotation ϕ dans le plan image et deux autres hors plan, ψ et α , étant liées aux paramètres a_{31} et a_{32} dans l'équation (1)). La simulation des couples d'images (I_0, I_1) pour différentes transformations nous a permis de délimiter les intervalles de variation des valeurs des paramètres aboutissant à un recalage réussi pour chacune des deux méthodes. Pour l'algorithme A_{IM} , ces intervalles sont : $t_x = t_y = \pm 30$ pixels, $f = \pm 25\%$, $\phi = \pm 20^\circ$ et $\alpha = \psi = \pm 20^\circ$. Les mêmes limites sont plus réduites pour l'algorithme A_{DQ} avec : $t_x = t_y = \pm 25$ pixels, $f = \pm 15\%$, $\phi = \pm 10^\circ$ et $\alpha = \psi = \pm 10^\circ$. Même si pour les deux méthodes, les limites des translations sont approximativement les mêmes (avec un léger avantage en faveur de l'information mutuelle), la méthode basée l'information mutuelle témoigne de la plus grande robustesse vis-à-vis des changements du facteur d'échelle et des variations des rotations dans et hors plan.

3.2 Précision

Pour évaluer la précision de recalage, nous avons construit un fantôme à partir d'une vessie de porc dont la texture de la

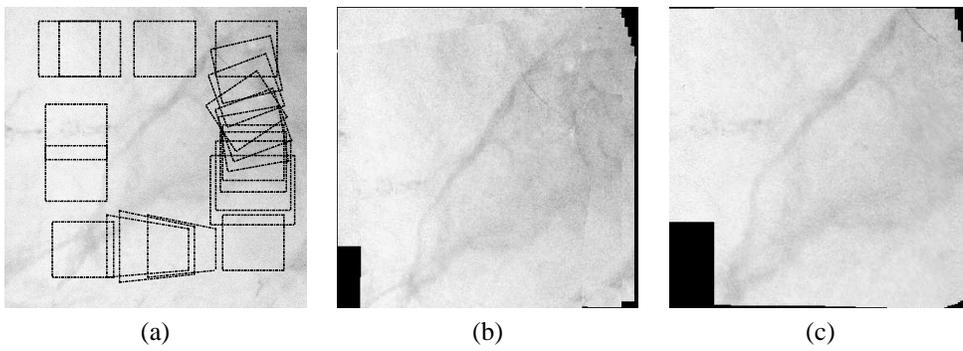


FIGURE 1 – (a) Photographie d’une vessie de porc : les carrés indiquent la séquence d’images simulées. (b) Image panoramique générée par l’algorithme A_{IM} à partir de la séquence d’images simulées. (c) Image panoramique obtenue avec l’algorithme A_{DQ} .

paroi interne est visuellement très proche de celle de l’homme (voir figure 1 (a)). La vessie est incisée, ouverte à plat et photographiée dans sa globalité avec une caméra. Ensuite, nous avons simulé l’acquisition d’une séquence d’images à partir de cette photographie. La surface couverte dans cette simulation est de l’ordre de 16 cm^2 . La première image de la séquence se situe dans le coin supérieur gauche de la photographie. La séquence des images est simulée selon un scénario de transformations perspectives comportant, notamment, des translations horizontales de 10 pixels (10 images en haut), des combinaisons de translations verticales de 10 pixels avec des rotations dans le plan de 2° (10 premières images sur la verticale à partir du coin supérieur droit), des combinaisons de translations verticales de 10 pixels avec des changements de facteur d’échelle de 5% (10 dernières images sur la verticale à droite) et des translations horizontales de 10 pixels (10 images sur l’horizontale en bas). Chaque couple d’images (I_k, I_{k+1}) de la séquence est recalé en utilisant chacun des deux algorithmes. La mesure d’erreur $\bar{\epsilon}_{k,k+1}$ utilisée pour quantifier la précision du recalage correspond à la moyenne des distances entre les positions des pixels homologues des images cibles I_k et recalées $T(I_{k+1}; \theta_k)$. Idéalement $\bar{\epsilon}_{k,k+1} = 0$ (en d’autres termes, θ_k superpose parfaitement I_{k+1} sur I_k). Dans le cas de translations pures, puis de la combinaison des translations avec les rotations hors plan puis dans le plan, les deux algorithmes donnent, approximativement, des résultats comparables avec des erreurs de recalage, respectivement, de 0.2, 0.6 et 3.5 pixels. Dans le cas de la combinaison des translations avec un changement du facteur d’échelle, l’algorithme A_{DQ} donne de bien meilleurs résultats ($\bar{\epsilon}_{k,k+1} \approx 1.5$ pixels) que l’algorithme A_{IM} ($\bar{\epsilon}_{k,k+1} \approx 4.5$ pixels). Ces erreurs de recalage sont, cependant, suffisamment petites pour ne pas affecter la cohérence visuelle de la carte générée (ainsi montrée sur la figure 1 (b) pour l’algorithme basé sur l’information mutuelle, la carte reste cohérente en terme de textures et est identique à l’image originale dans la figure 1 (a)). Il est important de souligner que les valeurs des transformations choisies pour la simulation des images tests sont bien plus importantes que les valeurs réelles des transformations lors d’un examen cystoscopique. En effet, les images sont acquises avec une fréquence de 25 images/seconde et la

vitesse de déplacement du cystoscope est de quelques millimètres/seconde. Dans ces conditions, les variations des valeurs des paramètres de transformation reliant deux images successives sont relativement petites ; moins de 5 pixels en translations avec des rotations ne dépassant pas 1° et un facteur d’échelle ne dépassant pas 5%. Dans le cas d’un examen cystoscopique réel, les deux méthodes donnent des erreurs de recalage inférieures à un pixel. Les figures 2 (a) et 2 (b) montrent deux images panoramiques construites à partir d’images cystoscopiques cliniques. La figure 2 (a) présente une image panoramique de taille 650×182 pixels construite à partir de 500 images en utilisant l’algorithme A_{IM} . Sur cette image panoramique, il n’y a pas de discontinuités visibles de la texture témoignant ainsi une bonne cohérence visuelle. L’image panoramique dans la figure 2 (b) est de taille 1479×1049 pixels. Elle est construite à partir de 450 images en utilisant l’algorithme A_{DQ} . Sur cette image panoramique, deux polypes sont visibles et facilement localisables.

3.3 Vitesse de calcul

Les algorithmes A_{IM} et A_{DQ} sont implémentés en C en utilisant la bibliothèque libre de vision par ordinateur OpenCV. L’évaluation de ces deux algorithmes est faite dans les mêmes conditions d’expérimentation avec un ordinateur de 2 Go de RAM et muni d’un processeur Intel Dual core(TM) à 2.40 GHz. La méthode d’optimisation de l’algorithme A_{IM} nécessite en moyenne 250 itérations et entre 50 et 60 secondes pour le recalage de deux images consécutives. Dans la figure 2 (a), la construction de l’image panoramique a pris 8 heures 27 minutes. Cependant, dans les mêmes conditions d’expérimentation A_{DQ} est 100 fois plus rapide que A_{IM} . En moyenne, A_{DQ} nécessite 12 itérations pour recalé deux images consécutives et le temps de recalage varie entre 0.3 et 0.6 seconde. L’image panoramique dans la figure 2 (b) est construite en 3 minutes 20 secondes. L’algorithme A_{DQ} permet ainsi la construction d’images panoramiques partielles de la surface interne de la vessie durant la procédure d’un examen cystoscopique clinique.

(a)



(b)

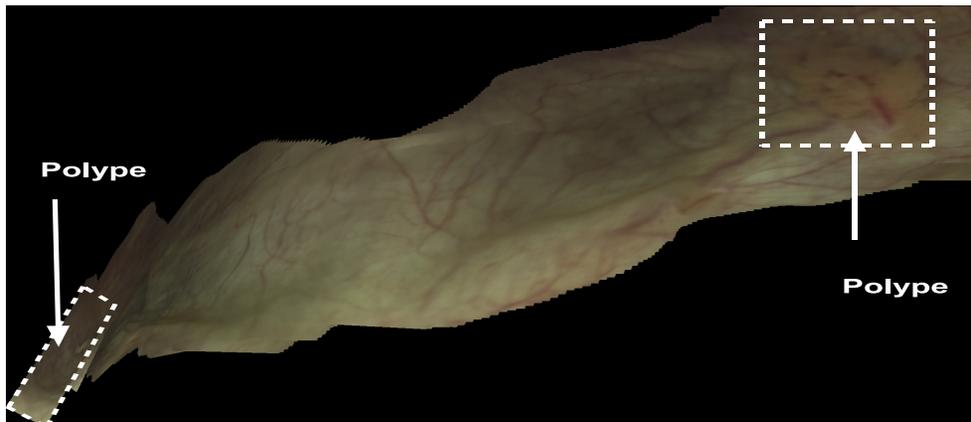


FIGURE 2 – (a) Image panoramique de taille 650×182 pixels construite à partir de 500 images en utilisant l’algorithme A_{IM} . Sur cette image panoramique, il n’y a pas de discontinuités visibles de textures et elle témoigne une bonne cohérence visuelle. (b) Image panoramique de taille 1479×1049 pixels construite à partir de 450 images en utilisant l’algorithme A_{DQ} . Deux polypes sont visibles et facilement localisables sur l’image panoramique.

4 Conclusion

Cette étude nous a permis de cerner les limites et les avantages de chacun des deux algorithmes que nous avons étudiés. En terme de précision de recalage, les deux algorithmes donnent des résultats comparables avec un avantage en faveur de l’algorithme A_{DQ} lors des changements de facteur d’échelle. Cependant, l’algorithme A_{IM} est globalement plus robuste que A_{DQ} surtout lorsque plusieurs paramètres subissent simultanément des changements significatifs. L’algorithme A_{DQ} est environ 100 fois plus rapide (quelques dixièmes de secondes pour recalculer deux images). Dans nos travaux futurs, nous allons essayer de combiner les deux algorithmes pour tendre vers une solution robuste, précise et rapide.

Références

- [1] A. Jalink, J. McAdoo, G. Halama, and H. Liu, “CCD mosaic technique for large-field digital mammography”, *IEEE Trans. Medical Imaging*, vol. 15, no. 3, pp. 260-267, juin 1996.
- [2] J.-S. Chou, J. Qian, Z. Wu, and H. Schramm, “Automatic mosaic and display from a sequence of peripheral angiographic images”, *Medical Imaging : Image Processing, Proc. SPIE*, vol. 3034, pp. 1077-1087, avril 1997.
- [3] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 347-364, mars 2002.
- [4] T. Vercauteren, A. Perchant, Pennec X., et N. Ayache. “Mosaicing of Confocal Microscopic In Vivo Soft Tissue Video Sequences”, *Lecture Notes in Computer Science, Proceedings of Medical Image Computing and Computer Assisted Intervention (MICCAI’05)*, Palm Springer, CA., pp. 804-810, Octobre 26-29-2005.
- [5] Y. Hernández-Mier, W. C. P. M. Blondel, C. Daul, D. Wolf, and G. Bourg-Heckly, “2-D panoramas from cystoscopic image sequences and potential application to fluorescence imaging”, in *Proc. 6th IFAC Symposium on Modelling and Control in Biomedical Systems (MCBMS)*, Reims, pp. 291-296, 2006.
- [6] Baker, M. and Matthews, I. Lucas-Kanade 20 years on : a unifying framework, *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221-255, 2004.
- [7] R. Miranda-Luna, C. Daul, W. C. P. M. Blondel, Y. Hernandez-Mier, D. Wolf, and F. Guillemin, “Mosaicing of bladder endoscopic image sequences : distortion calibration and registration algorithm”, *IEEE Trans. Biomed. Eng.*, vol. 55, no. 2, pp. 541-553, février 2008.
- [8] Viola et Wells III, “Alignment by maximization of mutual information”, *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137-154, 1997.