

# Segmentation d'images pour la reconnaissance de postures de mains

Jean-François COLLUMEAU<sup>1</sup>, Hélène LAURENT<sup>1</sup>, Bruno EMILE<sup>2</sup>, Rémy LECONGE<sup>2</sup>

<sup>1</sup>ENSI de Bourges, Laboratoire PRISME  
88 bd Lahitolle, 18020 Bourges Cedex, France

<sup>2</sup>Laboratoire PRISME, Université d'Orléans  
2 av. Mitterrand, 36000 Châteauroux  
12 rue de Blois, 45067 Orléans cedex2, France

jeanfrancois.collumeau@ensi-bourges.fr, helene.laurent@ensi-bourges.fr  
bruno.emile@univ-orleans.fr, remy.leconge@univ-orleans.fr

**Résumé** – La préservation de l'asepsie dans le bloc opératoire est vitale pour limiter la propagation des infections nosocomiales. De nos jours, elle est effectuée en imposant des mesures d'hygiène drastiques régulant l'utilisation d'équipements stériles par du personnel non-stérile et inversement. La commande sans contact par l'intermédiaire d'interfaces basées sur des dispositifs de vision informatique pourrait permettre de surmonter de telles limitations. Nous développons une chaîne de traitement logicielle visant à commander certains équipements du bloc opératoire de telle manière. L'étude présentée dans cet article concerne l'étape de segmentation de cette chaîne. Trois algorithmes de la littérature, *K-Means*, *Watershed* et *GrabCut*, sont comparés dans ce cadre. Les résultats de cette étude montrent l'algorithme K-Means comme le plus adapté à la segmentation de mains tant en qualité des résultats obtenus qu'en temps de calcul.

**Abstract** – Asepsis preservation is compulsory in operating rooms for avoiding the spread of hospital-acquired diseases. It is enforced nowadays by drastic measures regulating the use of sterile devices by non-sterile staff and vice versa. Contactless computer vision-based interfaces may overcome such limitations. We are developing an image processing chain aiming at allowing remote control of OR equipments. In this study, we focus on the segmentation step and investigate the capacities of three state-of-the-art segmentation algorithms, namely *K-Means classification*, *Watershed cuts* and *GrabCut*. This study's results give K-Means as the segmentation method the most adapted to the hand segmentation issue with respect to both segmentation accuracy and real-time processing aspects.

## 1 Introduction

Le maintien de l'asepsie dans les salles d'opérations est vital pour la limitation de la transmission d'infections nosocomiales au patient. Ces infections, qui touchent entre 7 et 12 % des personnes hospitalisées dans le monde, constituent une question brûlante dans les services de chirurgie [1, 2]. La transmission par le contact ayant été identifiée comme le principal mode de propagation de ces infections, la salle d'opération est actuellement divisée en deux zones distinctes, l'une stérile et l'autre non, de façon à limiter l'importance de ce facteur de transmission. Des mesures d'asepsie drastiques ont pour but de préserver la zone stérile de tout agent infectieux. Les contraintes associées à ces mesures empêchent les chirurgiens d'interagir directement avec des appareils non-stériles et les obligent à avoir recours à l'assistance d'une tierce personne. Bien que ce type d'interaction convienne parfaitement au chirurgien opérant pour le positionnement grossier des équipements et l'ajustement global des paramètres, il souhaite disposer d'un contrôle direct sur certains équipements spécifiques du bloc opératoire pour des réglages précis.

La partie commande sans contact du projet de recherche et de développement collaboratif *CORTECS* a pour objectif de

répondre à cette dernière exigence par le choix et la mise en oeuvre de dispositifs d'acquisition et de traitement d'image permettant de contrôler différents systèmes de la salle d'opération. Un utilisateur stérile serait alors en mesure d'interagir directement avec un équipement non-stérile ou inversement. L'objectif est donc de développer une Interface Homme-Machine (IHM) sans contact à destination prioritaire du chirurgien tout en demeurant potentiellement extensible au restant de l'équipe chirurgicale. Cette interface doit permettre de piloter les fonctionnalités de certains équipements présents dans le bloc opératoire de manière intuitive et sans générer de gêne pour l'équipe chirurgicale. Les travaux existant à ce jour sur le contrôle gestuel d'équipements de bloc opératoire sont presque exclusivement dédiés à la commande sans contact de dispositifs de visualisation d'images médicales pour la consultation de dossiers patients [3, 4, 5]. Basés sur l'utilisation de systèmes d'acquisition fixes, ils imposent à l'utilisateur de se déplacer pour se rapprocher du dispositif. Le projet en cours a pour objectif de permettre au chirurgien d'interagir avec les équipements en restant au plus près du patient, sans avoir à quitter la table d'opération.

Les contraintes imposées au système sont d'une part liées à l'activité, d'autre part au milieu de fonctionnement : l'al-

gorithme de commande gestuelle devra fonctionner en temps réel ; il devra de plus s'avérer robuste à la forte intensité et aux variations dynamiques de l'éclairage au cours de l'opération chirurgicale ainsi qu'aux variations de l'apparence des mains de l'utilisateur, en terme de couleur et de points de vue considérés. La chaîne de traitement d'images développée dans ce cadre est présentée figure 1. Nous nous intéressons plus spécifiquement dans cet article à l'étape de segmentation permettant de séparer la main de l'arrière-plan.

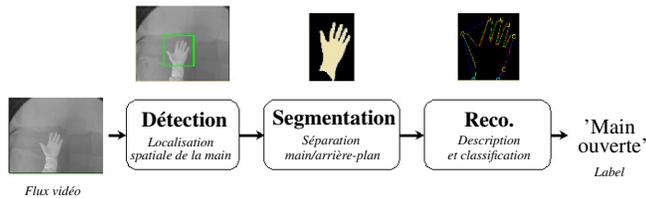


FIGURE 1 – Chaîne de traitement mise en place pour l'analyse de postures de mains

## 2 Protocole d'évaluation

### 2.1 Base d'images et critère d'évaluation

Le contexte de travail du bloc opératoire impose la prise en compte des dégradations liées à l'environnement, qu'elles soient de nature lumineuse (éclairage dynamique variable et de forte intensité) ou textuelle (gants pouvant être entachés de sang ou de sérum physiologique par exemple). La couleur n'étant par conséquent pas une information fiable, les acquisitions ont été réalisées grâce à une caméra en niveaux de gris portable. Le vocabulaire gestuel utilisé est constitué de six postures de main distinctes. Des vignettes extraites de la base d'images expertisée composée des couples image/vérité terrain correspondante a été créée. Au total, 20 variantes de chacune des six postures ont été enregistrées, représentant un total de 120 images et un nombre égal de vérités terrain. Acquisées dans des conditions proches des conditions réelles (utilisation de gants et draps chirurgicaux, d'un dôme d'éclairage opératoire), les images considérées présentent la réalisation des postures par différents locuteurs et diverses orientations de la main et rotations du poignet autour des trois axes. Elles correspondent à des situations pouvant rendre délicate l'étape de segmentation : doigts repliés, présence de plis marqués sur les gants, arrière-plan de couleur similaire au gant entre autres.

Travaillant dans un contexte supervisé, nous avons utilisé le critère du *F-score* [10] pour quantifier la qualité des résultats de segmentation obtenus. Défini par la moyenne harmonique de la précision et du rappel, la valeur du *F-score* varie entre 0 et 1, où 1 est le meilleur score. Seule une sous-région de chaque image, correspondant à la zone identifiée comme contenant la main lors de l'étape de détection et étendue de 55 % sur chaque côté, est segmentée afin de réduire le temps de calcul.

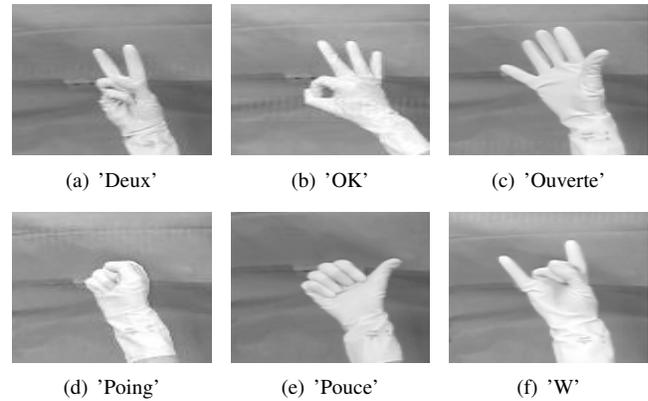


FIGURE 2 – Vocabulaire gestuel.

### 2.2 Algorithmes de segmentation comparés

Trois méthodes de segmentation permettant de traiter des images en niveaux de gris ont été considérées pour cette étude.

#### 2.2.1 K-Means

La segmentation par K-moyennes (*K-means*) [6] est un outil de classification de données regroupant celles-ci en  $K$  classes homogènes. Cet algorithme peut produire une segmentation cohérente de l'image sous réserve d'une intensité lumineuse localement homogène. Les paramètres  $K = 2$  à  $K = 5$  ont été considérés.

#### 2.2.2 Watershed

La segmentation par ligne de partage des eaux (*Watershed*) [8] considère une image en niveaux de gris comme un relief topographique où les valeurs des pixels correspondent à des altitudes. Partant de régions initiales, l'inondation de ce relief en suivant l'intensité du gradient de niveau de gris permet de segmenter l'image en régions. Cette méthode nécessite comme paramètres d'entrée des labels *a priori* permettant d'initialiser arrière- et avant-plans, ceux-ci correspondent respectivement à une extension de la sous-région de l'image issue de l'étape de détection et à une sous-région réduite à 1/6 ou 1/3 de la détection initiale. Les figures 3(b) et 3(c) illustrent, sur une image de la base, les deux initialisations prises en compte.

#### 2.2.3 GrabCut

Enfin, la méthode *GrabCut* [9] constitue l'une des applications à la segmentation d'images de la théorie des graphes, et plus particulièrement de la coupure de graphes. L'algorithme peut être initialisé sur deux régions indiquant les pixels appartenant vraisemblablement à l'avant-plan et à l'arrière-plan de l'image. Les paramètres utilisés sont les mêmes que pour le *watershed*. L'initialisation peut être affinée en définissant deux régions supplémentaires indiquant les pixels dont l'appartenance est connue avec certitude. Deux (avant- et arrière-plans probables), trois (avant- et arrière-plans probables, arrière-plan certain) et quatre (avant- et arrière-plans probables et certains)

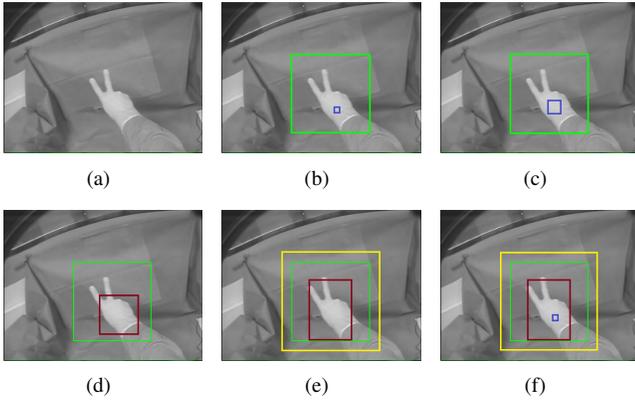


FIGURE 3 – Exemples d’image originale (a), et d’initialisations des algorithmes *Watershed* (b) et (c) et *GrabCut* (d), (e) et (f).

régions d’initialisation seront utilisées dans cet article. Des exemples de ces initialisations sont visibles sur les figures 3(d) à 3(f).

### 2.2.4 Implémentations

Les algorithmes *K-Means* et *Watershed* ont été tous deux fournis par la bibliothèque open source *OPENCV* [7]. Bien que cette bibliothèque comprenne également une implémentation du *GrabCut*, les graphes ont été calculés par un algorithme parallélisé sur GPU à l’aide de la librairie *NVIDIA Performance Primitives* pour des raisons de coût calculatoire. Les régions des images d’entrée identifiées par l’étape de détection et fournies aux algorithmes de segmentation ont vu leur taille réduite à 120x120 pixels afin de permettre un traitement temps réel.

## 3 Résultats expérimentaux

Les résultats issus de la comparaison des méthodes de segmentation *K-Means*, *GrabCut* et *Watershed* sur notre base d’images à l’aide de la mesure du *F-score* sont récapitulés dans le tableau 1. Ces résultats montrent l’algorithme *K-Means* avec le paramètre  $K = 3$  comme le plus performant avec un score égal à 0.89. Les paramétrages  $K = 4$  et  $K = 5$  se révèlent moins efficaces de par leur tendance à sur-segmenter l’image, et l’image est souvent sous-segmentée lorsque  $K = 2$ . Le *Watershed* se place également bien avec un score de 0.88. Le *GrabCut* obtient le plus mauvais score lorsque seulement deux régions *a priori* sont définies ; l’augmentation du nombre de ces régions permet d’améliorer ses performances sans pour autant atteindre celles des autres algorithmes. Il est à noter que la prise en compte d’une quatrième région représentant l’avant-plan certain améliore peu les performances moyennes du *GrabCut*. Quelques exemples des segmentations obtenues par ces algorithmes sont visibles figure 4.

Par ailleurs, *K-means* et le *Watershed* présentent chacun des défauts spécifiques liés à leur nature. Une illumination de la main non-uniforme comme sur la figure 5(a) amène *K-means* à sous-segmenter la main en affectant les pixels ombragés à l’arrière-plan, produisant le résultat de segmentation figure 5(c).

K-Means K = 2	K-Means K = 3	K-Means K = 4	K-Means K = 5	-
0.74	<b>0.89</b>	0.84	0.68	-
8	10	11	14	-
Watershed 1/6	Watershed 1/3	GrabCut 2 régions	GrabCut 3 régions	GrabCut 4 régions
0.88	0.88	0.66	0.79	0.80
73	42	123	118	94

TABLE 1 – Performances et coûts calculatoires moyens des méthodes de segmentation étudiées, exprimées respectivement en termes de valeurs prises par le *F-score* et en millisecondes.

Un tel résultat peut perturber l’étape de reconnaissance de la main sans pour autant être rédhibitoire. Le *Watershed* quant à lui voit sa progression suivant les gradients de niveaux de gris de l’image stoppée par des plis importants du gant comme le montrent les figures 5(d) à 5(f). Ce dernier défaut est problématique pour la reconnaissance de la main car des doigts entiers peuvent disparaître ainsi.

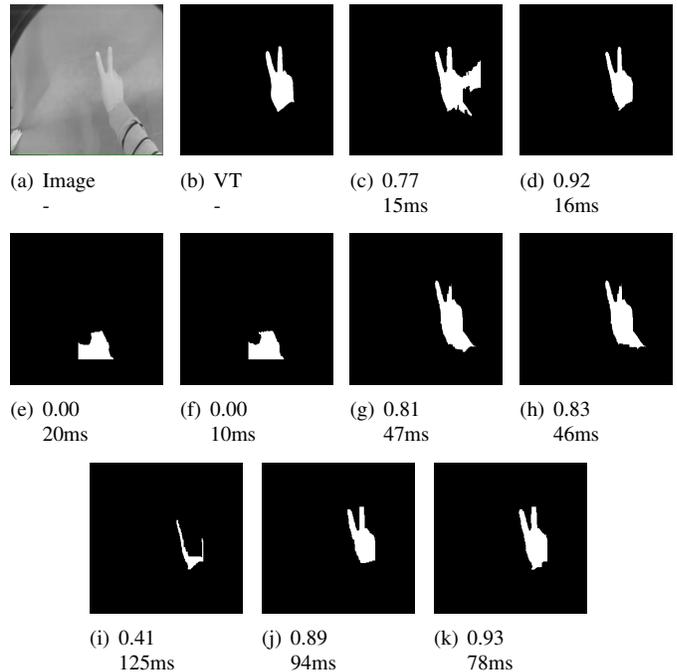


FIGURE 4 – Exemples d’image originale (a), vérité terrain associée (b), et segmentations obtenues par les algorithmes *K-means*  $K = 2$  (c) à  $K = 5$  (f), *Watershed* 1/6 (g) et 1/3 (h) et *GrabCut* 2 régions (i) à 4 régions (k). Les valeurs du *F-score* associées ainsi que les coûts calculatoires de chaque algorithme sont indiqués en dessous des images.

Le coût calculatoire d’un algorithme possède un poids important dans toute application au fonctionnement temps réel. Notre étude montre la supériorité de l’algorithme *K-Means* sur ce point ; en effet il ne lui faut qu’une dizaine de millisecondes pour fournir une segmentation efficace, tandis que le *Watershed* et le *GrabCut* requièrent respectivement quatre et dix fois plus

de temps pour obtenir un résultat inférieur ou équivalent.

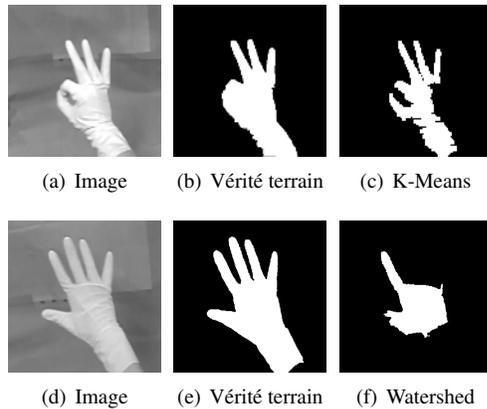


FIGURE 5 – Exemples de mauvais résultats de segmentation. Sur la première ligne, K-Means ( $K = 3$ ) interprète mal les ombres portées sur gant ; la seconde ligne montre la sensibilité du Watershed aux plis du gant.

## 4 Conclusion et perspectives

### 4.1 Conclusion

L'étude comparative présentée dans cet article dans le contexte de la commande sans contact dans un bloc opératoire vise à évaluer les performances de trois méthodes de segmentation sur des images en niveaux de gris. L'évaluation a été conduite sur une base d'images *ad hoc* s'approchant des conditions environnementales du bloc opératoire. La mesure du *F-score* utilisée pour quantifier la qualité des résultats obtenus montre l'algorithme *K-Means* comme le plus performant avec un score égal à 0.89, suivi par le *Watershed* (0.88) et le *GrabCut* (0.80). La contrainte temps réel de notre application favorise également *K-Means* de par sa rapidité : 10 ms par image de 120x120 pixels, contre plus de 40 ms pour ses concurrents.

### 4.2 Perspectives

Les résultats présentés dans cet article seront complétés par l'évaluation des autres étapes de la chaîne de traitement avant de quantifier les performances du système complet. Le vocabulaire utilisé, défini à des fins expérimentales, sera étendu de manière à optimiser les performances de la chaîne de traitement et offrir une expérience ergonomique et intuitive aux chirurgiens.

## Remerciements

Les auteurs souhaitent remercier le Ministère de l'Industrie, la Région Centre, le FEDER et l'OSEO dans le cadre du programme FUI 9 soutenu par le Pôle de Compétitivité S2E2.

## Références

- [1] M. Saleem, A.K. Vaish, M.Z. Idris, A.A. Sonkar, J. Agarwal, M. Singh, I. Ahmad et M.K. Gupta. *Prevalence of nosocomial infections in surgical wards of tertiary care hospital at Lucknow*. Indian Journal of Scientific Research, 3(2) :79-84, 2012.
- [2] J. Reilly, S. Stewart, G. Allardice, A. Noone, C. Robertson, A. Walker et S. Coubrough. *NHS Scotland national HAI prevalence survey*. Health Protection Scotland [Report], 2007.
- [3] C.R. Dressler, T. Neumuth, M. Fischer, O. Abri et G. Strauss. *Intraoperative use of electronic patient records by the surgeon. An evaluation of the "MI Report" in ENT surgery*. HNO, 59(9) :900-907, 2011.
- [4] T. Kipshagen, M. Graw, V. Tronnier, M. Bonsanto et U.G. Hofmann. *Touch- and marker-free interaction with medical software*. IFMBE Proceedings of the World Congress on Medical Physics and Biomedical engineering, 2009.
- [5] J.P. Wachs, H.I. Stern, Y. Edan, M. Gillam, J. Handler, C. Feied et M. Smith. *A gesture-based tool for sterile browsing of radiology images*. Journal of the American Medical Informatics Association, 15(3), 2008.
- [6] J.B. MacQueen. *Some methods for classification and analysis of multivariate observations*. Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability, L. M. Le Cam and J. Neyman, Eds. 1967, 1 :281-297, University of California Press.
- [7] G.R. Bradski et V. Pisarevsky. *Intel's computer vision library : Applications in calibration, stereo segmentation, tracking, gesture, face and object recognition*. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2000.
- [8] S. Beucher et F. Meyer. *The Morphological approach of segmentation : the watershed transformation, in Mathematical Morphology in Image Processing*. Marcel Dekker, 1992.
- [9] C. Rother, V. Kolmogorov et A. Blake. *Grabcut : interactive foreground extraction using iterated graph cuts*. Proceedings of ACM Transactions on Graph, 309-314, 2004.
- [10] C. Wolf et J.M. Jolion. *Object count/area graphs for the evaluation of object detection and segmentation algorithms*. International Journal on Document Analysis and Recognition, 8(4) :280-296, 2006.