

Simulation d'images en microscopie électronique pour la détection de la contamination de cellules par un virus

Grégoire COLSON¹, Séverine BALMAND², Corinne BOYER²,
Manuel ROSA-CALATRAVA³, Carole FRINDEL¹, David ROUSSEAU¹

¹Université de Lyon, CREATIS, CNRS UMR5220, INSERM U1206, Université Lyon 1, INSA-Lyon, 69621 Villeurbanne, France

²Biologie Fonctionnelle, Insectes et Interactions (BF2I), INRA, INSA-Lyon, Villeurbanne, France

³Virologie et Pathologie Humaine VIRPATH, Université Lyon 1, Lyon, France
david.rousseau@univ-lyon1.fr

Résumé – Nous présentons un simulateur d'images de cellules infestées par des virus en microscopie électronique. Ce simulateur, qui mixe une approche de génération de textures et d'empilement de structures cellulaires, permet de générer des images automatiquement annotées pour tester les performances d'algorithmes de détections de ces virus. Nous testons un algorithme simple de détection sur des images de synthèses et sur des images réelles et discutons le réalisme du simulateur.

Abstract – A simulator of images of cells including virus in electronic microscopy is presented. This simulator, which mixes texture and accumulation of cellular-like structures, enables to generate automatically annotated images that can serve to test the performances of schemes for the detection of these virus. We test a specific simple detection algorithm on synthetic images as well as on real images and discuss the realism of the simulator.

1 Introduction

Comme le souligne Danuser dans un article datant de 2011 intitulé “Computer Vision In Cell Biology” [1] : “même si les outils informatiques sont très utilisés pour l'acquisition d'images en microscopie, l'interprétation du contenu de ces images est la plupart du temps laissée à l'intuition humaine”. Cette situation est en train d'évoluer. En effet, l'augmentation du nombre de techniques de microscopie (tomographie optique de cohérence, microscopie à feuille de lumière, microscopie confocale, microtomographie à contraste de phase, ...) couplées à des systèmes de scan à haut débit révolutionne la problématique de l'analyse d'images en biologie [2] et nécessite donc des outils informatiques adaptés et performants comme ceux développés dans d'autres domaines de la vision par ordinateur.

Notre travail s'inscrit dans ce contexte d'application de la vision par ordinateur à la biologie. Nous considérons en particulier la problématique de la validation d'un algorithme pour une tâche de reconnaissance de formes à l'échelle microscopique. La méthode la plus courante pour valider des algorithmes de détection est d'avoir recours aux vérités terrains associées. Celles-ci doivent préalablement être établies par des experts, qui épluchent les images et réalisent à la main le travail de l'algorithme de détection. On compare ensuite la détection des algorithmes avec cette vérité. Néanmoins, face à un trop grand nombre d'images, il devient difficile d'établir les vérités terrains. Une solution peut être de construire un simulateur générant automatiquement des images labellisée ayant des propriétés pro-

ches des images réelles afin d'établir une vérité terrain synthétique. Il est alors possible de générer un nombre illimité d'images, les vérités terrains associées et ainsi de mesurer les performances des algorithmes. Cette approche de validation *in silico* est répandue en microscopie photonique [3, 4]. Notre travail constitue une contribution originale pour des images de microscopie électronique et en lien avec des structures virales.

On considère une tâche de détection de formes en microscopie électronique par transmission (TEM) pour la caractérisation de la contamination de cellules par un virus. Les virus sont des agents pathogènes dont la taille évolue entre 0.01 et 0.3 μm ce qui a empêché leur étude approfondie avant l'avènement de la microscopie électronique. Notre étude porte sur deux virus en particulier : le baculovirus, ainsi que le virus de la grippe infectant des cellules humaines [5]. Ces deux virus entraînent une modification de l'ultra-structure de la cellule comme une augmentation de la taille et de la structure du noyau et des éléments du cytoplasme, une apparition de structures virales telles des virus sous forme de polyèdres, etc. Dans cette étude, nous nous focalisons sur la détection des éléments viraux lorsqu'ils sont sous forme de polyèdres.

2 Une tâche de détection

Soit H_0 correspondant à la situation où “La cellule ne contient pas de virus sous forme de polyèdre” (Fig. 1 Gauche) et l'hypothèse alternative H_1 “La cellule contient au moins un vi-

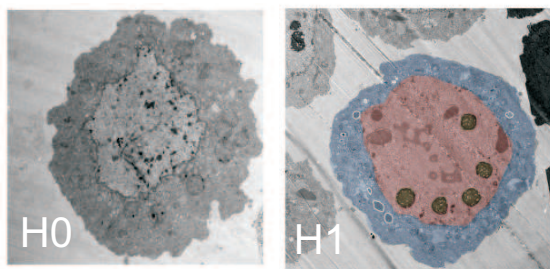


FIGURE 1 – Gauche image de cellule humaine sans la présence de virus. Droite, image de cellule humaine avec présence de virus sous la forme de polyèdres. Dans l'image de droite, on trouve en bleu un masque du cytoplasme, en rose celui du noyau et en marron celui des polyèdres.

rus sous forme de polyèdre”. Le travail des algorithmes de détection est ainsi de valider l’une ou l’autre de ces hypothèses. Nous avons acquis deux bases d’images décrites dans le Tableau 1. Des vérités terrains ont été réalisées manuellement sur ces deux bases comme le montre la Fig. 1 de droite. Ces vérités terrain servent classiquement de référence pour la validation d’algorithmes de classification visant à décider entre l’hypothèse H0 ou H1. Ici nous les utilisons pour réaliser un simulateur d’images.

Base	1	2
Virus	Baculovirus	Grippe
Nombre d’images	28	11
Images sans polyèdres (H1)	13	2
Images avec polyèdres (H0)	15	9

TABLE 1 – Bases d’images annotées manuellement utilisées. Une image de chaque base est donnée sur la Fig. 4.

3 Simulateur de cellules avec virus

Nous présentons dans cette section la méthode utilisée pour valider des machines de vision dédiées à la détection de polyèdres comme montré dans la Fig. 1. Le processus global est donné en Fig. 2.

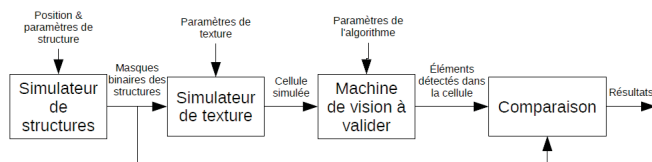


FIGURE 2 – Processus de synthèse d’images et de validation des algorithmes de détection de virus.

Pour générer des images synthétiques proches de celles de la Fig. 1 nous proposons un modèle hybride entre un modèle

de type feuilles mortes [6], i.e. superposition de couches, et un modèle purement statistique, i.e. remplissage avec une texture aux propriétés statistiques stationnaires. Nous avons séparé les éléments cellulaires en deux types : les éléments principaux et les éléments secondaires. Les éléments principaux sont le cytoplasme, le noyau et les éléments à détecter (les polyèdres dans notre cas). Ils sont placés dans l’image pour créer des structures. Dans le cas de la cellule, contrairement à l’approche classique de type feuilles mortes, où les objets sont entassés les uns sur les autres sans ordre, une relation d’inclusion existe entre les éléments principaux et il y a la contrainte de n’avoir aucun élément se chevauchant entre ces structures. Les formes des masques de ces éléments principaux sont soit tirés aléatoirement au sein de la base des vérités terrain segmentées manuellement ou bien sous la forme d’ellipses. Les éléments secondaires sont d’autres composants cellulaires tels que la polyédrine du noyau ou les vacuoles du cytoplasme et sont modélisés sous forme de texture générée avec un processus stochastique (bruit blanc gaussien filtré passe-bas) dont les statistiques du premier et second ordre sont celles estimées sur les bases d’images du Tableau 1. À ce simulateur de structures et textures imbriquées, nous ajoutons deux étapes supplémentaires de natures plus physiques. On réalise une convolution par un masque gaussien qui génère un flou simulant la réponse impulsionnelle du microscope électronique. Aussi, on ajoute un bruit de type poivre (pixel noir) qui simule la présence d’agrégats métalliques liés à la fixation de l’échantillon qui peut se concentrer sur certaines structures comme le pourtour des polyèdres mais aussi se disperser dans toute l’image. Certains autres bruits pourraient être ajoutés comme les artefacts liés à la découpe de l’échantillon qui produisent des traits orientés très caractéristiques. Il existe de nombreux algorithmes efficaces pour éliminer ces artefacts (par exemple [7]) c’est pourquoi nous n’avons pas augmenté la complexité de notre simulateur afin de rester sur une version “minimaliste” en termes de nombre de paramètres. Le réalisme des images ainsi simulées peut être apprécié sur la Fig. 3 pour une vue d’ensemble et plus en détail sur la Fig. 5.

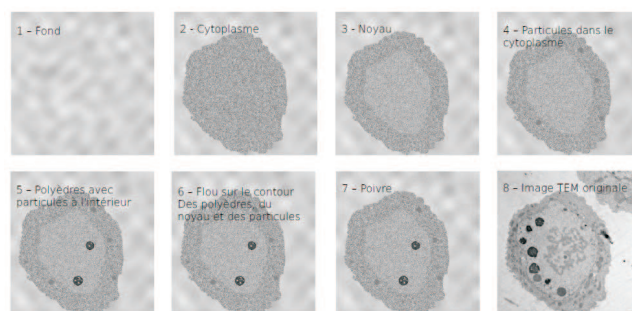


FIGURE 3 – Frise illustrant de 1 à 7 la production de cellules de synthèses proches d’images de cellules contenant des virus sous forme de polyèdres et acquises en microscopie électronique. L’image 8 est une image réelle pour comparaison avec l’image de synthèse 7 finalisée.

4 Schéma de détection

Nous sommes prêts pour tester un schéma de détection de polyèdres dans des images de microscopie électronique de cellule. Nous avons élaboré le schéma décrit sur la Fig. 4 qui enchaîne, dans un esprit d'analyse d'images, une succession de filtres avec une variante concernant l'étape 3 sur la forme pour les deux bases d'images. En effet, les polyèdres pour la base d'images du virus de la grippe apparaissent véritablement anguleux quand ceux observés pour le Bacculovirus les polyèdres ont des bords très émoussés et arrondis. Dans une première étape, on utilise l'a priori biologique que les polyèdres sont (i) des structures qui apparaissent en leur sein en hyposignal relativement stable, (ii) des structures qui fixe l'agent de contraste en leur bordure (iii) des structures qui constituent des formes convexes. Une étape de segmentation de type maximally stable extremal region [8] suivi d'une ouverture et d'un test de solidité ne conservent que les objets convexes. L'échelle d'observation dans les deux bases est constante avec systématiquement une unique cellule entière dans l'image, il est possible de se baser sur une connaissance a priori de la taille des virus. Un test de taille est donc appliqué pour ne conserver que les objets binaires correspondant à la taille attendue. On utilise un a priori de forme circulaire pour le bacculovirus et anguleuse (via un algorithme de détection de coin [9]) pour le virus de la grippe. Enfin, on tient compte du fait que les contours des polyèdres sont marqués par un hyposignal plus intense. L'ensemble des filtres correspond à des outils classique de traitement d'images.

Le schéma de détection de la Fig. 4 qui tient rigoureusement compte de l'ensemble des a priori peut sembler laborieux. Cependant, il faut garder en tête la formidable diversité des contrastes observés en microscopie électronique. Lors des tests des performances du schéma de détection de la Fig. 4 aucune des étapes n'est apparue inutile.

5 Performances de détection

L'algorithme de la section précédente valide pour chaque image l'hypothèse H0 ou H1 que l'on compare avec la classe réelle de l'image, établie par le simulateur ou les experts (C1 si l'image contient réellement des polyèdres, C0 sinon). Nous présentons les performances dans une matrice de confusion.

	C1	C0		C1	C0
H1	71	17	H1	69	39
H0	29	83	H0	31	61

TABLE 2 – Matrices de confusion exprimée en % pour l'algorithme de la Fig.4 sur les images réelles annotées (à gauche) et sur 200 images de synthèse (à droite) générées par le simulateur de la section 3.

Les résultats de détection sur les images générées montrent une détection moins spécifique et sensible sur les images de synthèse issue de notre simulateur par rapport aux performances obtenues sur les images réelles. Cependant, on observe un ra-

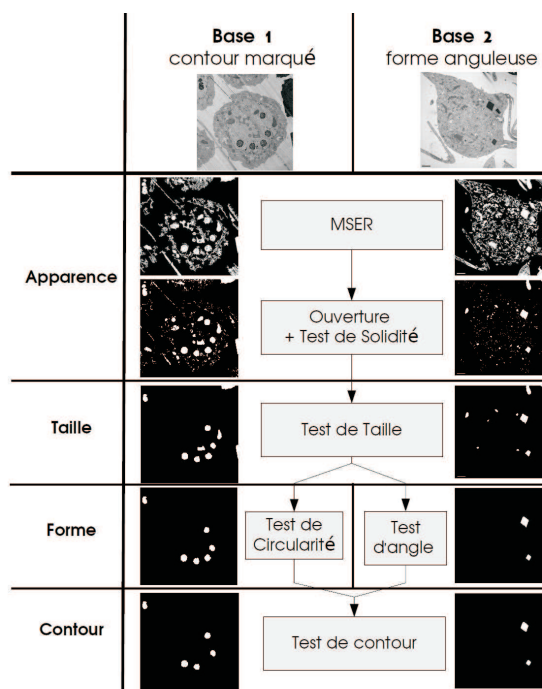


FIGURE 4 – Algorithmes de détection de virus sous forme de polyèdres en microscopie électronique.

tio similaire entre les différentes statistiques : une spécificité (probabilité de ne pas détecter le virus chez les cellules non affectées) faible, une sensibilité (probabilité de détecter le virus chez les cellules réellement affectées) forte et une précision moyenne. Cela pousse à penser que le générateur produit des images conduisant à des erreurs similaires à celles synthétisées sur les images réelles, en les exagérant. La Fig. 5 droite montre certaines erreurs de détection réalisées par l'algorithme sur des images de synthèse. Les faux positifs (J, K, L dans la Fig. 5) dans les images simulées correspondent à des faux positifs également rencontrés dans les images réelles (C, D, E dans la Fig. 5) mais dont la fréquence d'apparition est difficile à établir vu le faible nombre d'images réelles disponibles à ce stade de nos investigations. Les faux négatifs (G, H de la Fig. 5) correspondent à des contrastes très faibles en termes de niveau de gris ou bien à des tailles de polyèdres trop petits pour être détectés (Fig. 5I). Aussi, nous observons des structures à l'intérieur de polyèdres réels (Fig. 5A) que nous simulons (Fig. 5G et H) mais qui manifestement perturbent l'algorithme testé. Un test de texture supplémentaire, se concentrant sur la détection de ces particules blanches dans les polyèdres, pourrait régler le problème. L'algorithme de détection de la Fig. 4 montre des faiblesses, au niveau du test de contour et de la différenciation entre les polyèdres et les autres éléments de la cellule lorsque le contraste de niveau de gris entre noyau et les polyèdres est très faible comme le montre les Figs. 5G et H. Certaines situations de faux positifs sur les images réelles montrées Fig. 5 B et F correspondent à des polyèdres en train de se former, une situation qui n'est pas simulable simplement avec notre approche de

structure-texture. Pour tous les cas problématiques simulables, une étude approfondie sur un plus grand nombre d'images de l'ultrastructure de la cellule dans le cas de l'infection par un tel virus semble nécessaire. Elle permettrait d'établir des fréquences d'apparition de différents éléments cellulaires et des différents types d'erreurs de détection commises par l'algorithme.

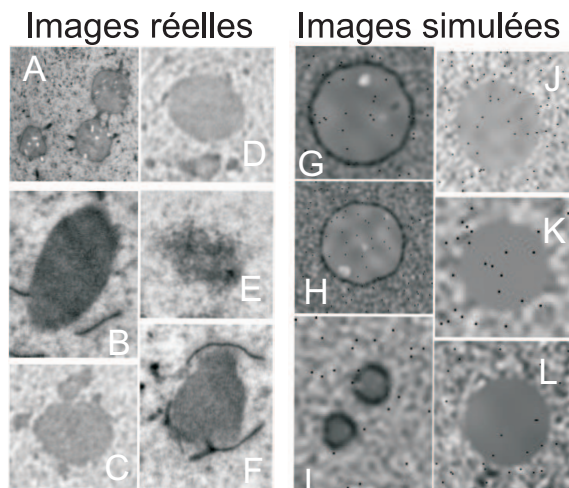


FIGURE 5 – Erreurs de détection réalisées par le schéma de détection. Sur la colonne de gauche dans le cas d'images réelles et dans celle de droite pour des images de synthèse. Panneaux A, G, H, I sont des faux négatifs et le reste des faux positifs.

Les résultats de détection du Tableau 2 correspondent à des images générées en utilisant les paramètres de structure et de texture estimés sur les images réelles. Le simulateur permet également de faire varier ces paramètres indépendamment. Afin de tester la robustesse d'un algorithme de détection, il est ainsi possible de faire évoluer certains paramètres du simulateur un à un et de mesurer leur influence sur la sensibilité et la spécificité d'un schéma de détection. Pour illustration, nous testons la robustesse du schéma de détection de la Fig. 4 au contraste entre le noyau et les polyèdres. Pour ce faire, on calcule entre le noyau et les polyèdres (en excluant les bords sombres des polyèdres) le ratio de Fisher $FR = \frac{(\lambda_A - \lambda_B)^2}{\sigma_A + \sigma_B}$ avec λ et σ respectivement la moyenne et la variance des valeurs de gris dans la région considérée. On observe sur la Fig. 6 que le ratio de Fisher FR a un effet sur la sensibilité de la détection mais pas sur la spécificité. Lorsque les histogrammes des deux éléments sont semblables, l'algorithme a plus de mal à détecter les polyèdres, mais l'effet est visible seulement lorsque les histogrammes sont proches ($FR < 0.2$), comme sur les faux négatifs visibles en Fig. 5.

6 Conclusion

Nous avons présenté un simulateur d'images de cellules infectées de virus en microscopie électronique. Ce simulateur

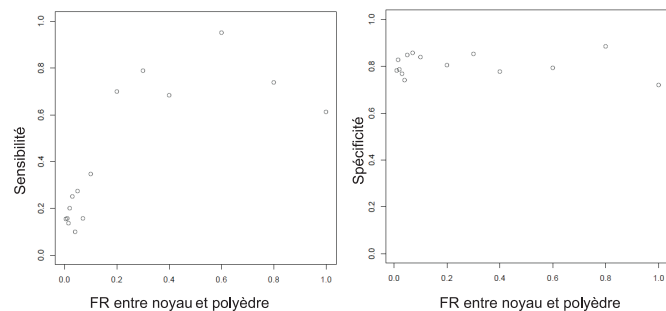


FIGURE 6 – Sensibilité et spécificité de l'algorithme de détection de la Fig. 4 en fonction du contraste mesuré par le Ratio de Fisher (FR) entre le noyau et les polyèdres.

peut être utilisé pour validation objective et automatique de la robustesse d'algorithme de traitement d'images. L'algorithme testé ici pour illustration était classiquement basé sur l'expertise issue de l'observation d'invariants par des humains, mais une extension naturelle serait l'utilisation de données simulées pour servir de base d'apprentissage dans des approches supervisées d'intelligence artificielle comme, bien sûr, celles connexionnistes de type "deep learning" moyennant une adaptation de domaine qui requièrent de grands jeux d'images annotées.

Références

- [1] G. Danuser, Computer vision in cell biology. *Cell*, vol. 5, pp. 973–978, 2011.
- [2] E. Meijering, *et al*, Imaging the future of bioimage analysis. *Nature Biotechnology*, vol. 34, pp. 1250–1255, 2016.
- [3] S. Rajaram, *et al*, SimuCell : a flexible framework for creating synthetic microscopy images. *Nature methods*, vol. 9, pp. 634–639, 2012.
- [4] V. Ulman, *et al*, Virtual cell imaging : A review on simulation methods employed in image cytometry, *Cytometry Part A*, John Wiley Sons, 16 pp., 2017.
- [5] O. Terrier, *et al*, Ultrastructural fingerprints of avian influenza A (H7N9) virus in infected human lung cells. *Virology*, vol. 456, pp. 39–42, 2014.
- [6] C. Bordenave, *et al*, The dead leaves model : a general tessellation modeling occlusion. *Advances in applied probability*, vol. 38, pp. 31–46, 2006.
- [7] J. Fehrenbach, *et al*, Variational algorithms to remove stationary noise : applications to microscopy imaging. *IEEE Transactions on Image Processing*, vol. 21, pp. 4420–4430, 2012.
- [8] J. Matas, *et al*, Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, vol. 22, pp. 761–767, 2004.
- [9] C. Harris, *et al*, A combined corner and edge detector. *Image and vision computing*, vol. 15, pp. 44–52, 1988.