

Diagnostic du cancer à partir d'images de coupes de tissus non-colorées à résolution nanométrique en utilisant des statistiques d'ordre supérieur

ANDRY RAJAONARIVÉLO¹, ERIC PERRIN¹, GABRIEL POPESCU², KRISHNARAO TANGELLA³, VALERIU VRABIE¹

¹ CReSTIC, Université de Reims Champagne-Ardenne,
Moulin de la Housse - B.P. 1039, 51687 REIMS Cedex 2, France

² QLI Laboratory, University of Illinois at Urbana-Champaign,
405 North Mathews Avenue, Urbana, IL 61801, USA

³ Department of Pathology, Christie Clinic,
1400 West Park St., Urbana, IL 61801, USA

maminiaina.rajaonarivelo@etudiant.univ-reims.fr, eric.perrin@univ-reims.fr, ¹valeriu.vrabie@univ-reims.fr

²gpopescu@illinois.edu

³Ktangella@christieclinic.com

Résumé - Le but de ce travail prospectif est d'étudier si la dégradation introduite par le cancer dans l'architecture des cellules peut être mise en évidence par l'analyse statistique aux ordres supérieurs de l'information enregistrée à l'aide d'une nouvelle technique microscopique : la Spatial Light Interference Microscopy (SLIM). Cette technique mesure le contraste intrinsèque des structures cellulaires de coupes de tissus non-colorées avec une résolution nanométrique, une précision topographique de 0.3 nm et des temps d'acquisition rapides. En se limitant aux statistiques jusqu'à l'ordre 4 et en les combinant à l'aide d'un classifieur non-supervisé de type fuzzy C-means, les études qualitatives et quantitatives entre les résultats obtenus sur des adénocarcinomes de la prostate et des images des coupes colorées H&E analysées au préalable par notre anatomopathologiste montrent le fort potentiel de cette méthodologie pour le diagnostic du cancer à partir d'images SLIM de coupes de tissus non-colorées.

Abstract – The aim of this preliminary study is to investigate if nanoscale changes caused by cancer to the cells' interior architecture could be highlighted by higher order statistical analysis of the information recorded by a new microscopy technique: the Spatial Light Interference Microscopy (SLIM). This technique can reveal the intrinsic contrast of cell structures from unstained tissue sections with nanometric resolution, a topographic accuracy of 0.3 nm and fast acquisition time. Using statistics up to 4th order and combining them with an unsupervised classifier such as fuzzy C-means, qualitative and quantitative comparisons between the results on high-grade prostate adenocarcinomas and images of H&E stained sections previously analyzed by our pathologist show the potential of this methodology to diagnose cancer from SLIM images of unstained tissue sections.

1 Introduction

Les techniques optiques sont couramment utilisées dans l'imagerie du cancer. La méthode classique employée par les anatomopathologistes consiste à analyser à l'aide d'un microscope des coupes tissulaires fines (quelques μm) après leur coloration à Hématoxyline-Eosine (H&E). En analysant la structure morphologique et l'architecture des tissus, ils valident ou non le diagnostic de cette maladie. Cette procédure, considérée comme le "gold standard" [1], est cependant relativement coûteuse, la coloration représentant environ 60-80% du coût total d'une analyse.

Différentes techniques complémentaires permettant d'analyser les coupes tissulaires sans coloration ont été explorées. La plus répandue est l'analyse par spectroscopie vibrationnelle, Raman ou Infrarouge. Ces techniques ont démontré un fort potentiel pour l'analyse histologique numérique non-destructive à résolution élevée (quelques μm) dans le diagnostic de différents types de cancer [2]. Des études récentes ont démontré qu'il est également possible d'analyser les coupes sans déparaffinage chimique en utilisant des techniques avancées de traitement du signal [3]. Cependant,

l'analyse par spectroscopie vibrationnelle a un désavantage important : le temps d'acquisition est long (quelques heures pour une biopsie d'environ 1 mm^2).

Une nouvelle technique microscopique, la Spatial Light Interference Microscopy (SLIM) a récemment été développée en combinant deux idées classiques : la microscopie à contraste de phase et l'holographie [4]. Cette technique permet de mesurer le contraste intrinsèque des structures cellulaires de coupes de tissus non-colorées avec une résolution nanométrique, une précision topographique de 0.3 nm et des temps d'acquisition très rapides. Elle permet de quantifier des informations cellulaires accessibles uniquement à très haute résolution.

En s'appuyant sur l'hypothèse que le cancer induit des altérations dans les structures cellulaires, la distribution de pixels appartenant à des zones tumorales devrait être modifiée par rapport aux distributions des pixels des zones saines.

Le but de cette étude prospective est d'examiner si la dégradation introduite par le cancer dans l'architecture cellulaire peut être mise en évidence par une analyse statistique d'images SLIM. Nous nous sommes limités

jusqu'aux statistiques d'ordres 4, variance, Skewness et Kurtosis, chacune pouvant être utilisée pour discriminer les zones tumorales du reste. Nous montrons également qu'en combinant ces informations via un classifieur non-supervisé de type fuzzy C-means, les zones tumorales sont mieux identifiées. Des études qualitatives et quantitatives sur deux adénocarcinomes de la prostate montrent le fort potentiel de cette méthodologie à diagnostiquer le cancer sur des coupes de tissus non-colorées.

2 Système d'acquisition et données

La Spatial Light Interference Microscopy (SLIM) est une nouvelle technique microscopique capable de mesurer les indices de réfraction de structures nanométriques dans des cellules (vivantes) via l'interférométrie [4]. Elle permet d'acquérir des images des structures transparentes avec une précision topographique pouvant aller jusqu'à 0.3 nm, ce qui dépasse les performances de tous les systèmes actuels. Cette technique révèle le contraste intrinsèque des structures cellulaires, la précision résultante étant comparable à celle de la microscopie à force atomique, alors que la vitesse d'acquisition est 1000 fois plus élevée. La SLIM permet dans son développement actuel des acquisitions très rapides, de quelques minutes, sur des coupes tissulaires de dimensions typiques en analyse histopathologique (environ 1cm²).

La SLIM a été utilisée pour acquérir des images sur des coupes des tissus d'adénocarcinomes (cancer) de la prostate à degré de malignité élevé. Les coupes ont été sectionnées (4 μm d'épaisseur), déparaffinées chimiquement, puis des images à résolution de 200 nm ont été acquises sans aucune coloration préalable. Les coupes analysées ayant des dimensions d'environ 4 cm², les images résultantes sont de très grandes dimensions (Giga-pixels). Les images SLIM ont été quantifiées sur 256 niveaux de gris pour une plage de valeurs mesurées (proportionnelles aux indices de réfraction) variant entre $-\pi/2$ (noir) et $3\pi/2$ (blanc). La figure 1 montre, à gauche, deux zooms d'une image acquise : zone saine en haut et zone tumorale en bas ; à droite, les histogrammes correspondant à ces deux zones.

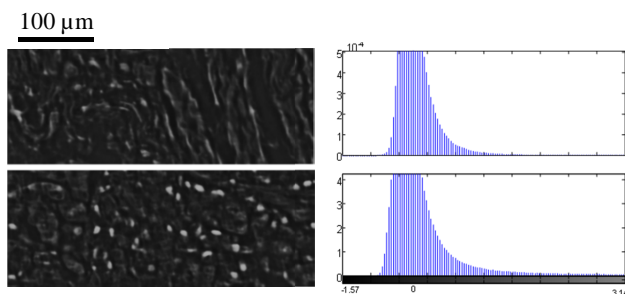


Figure 1 : A gauche : zooms zone saine en haut et zone tumorale en bas. A droite : histogrammes.

Afin de pouvoir comparer les résultats avec la méthode histologique classique, des coupes adjacentes ont été également sectionnées, colorées (H&E) puis scannées et analysées par notre anatomopathologiste. Sur ces images, 5 zones tumorales et 5 zones saines ont

été marquées par l'anatomopathologiste, ces zones servant pour des mesures quantitatives.

3 Méthodologie d'analyse : statistiques d'ordre supérieur et classification floue

La Figure 2 montre le principe de la méthodologie proposée. L'hypothèse de base est que le cancer induit plus de variations aléatoires dans les mesures SLIM (cf. histogrammes de la Figure 1) du fait de l'altération des structures cellulaires. Ceci est vrai notamment dans les structures contenues dans le noyau, uniquement accessibles par des techniques à très haute résolution, comme le nucléolinus (structure très proche du nucléole, riche en ARN, et relativement mal connue, bien que découverte il y a 150 ans).

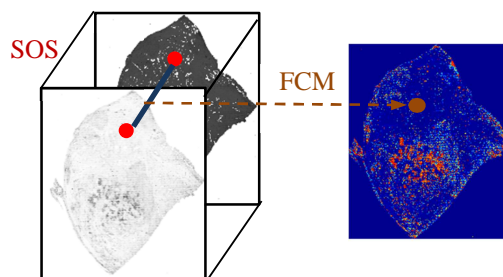


Figure 2 : Méthodologie : Extraction des caractéristiques par des statistiques d'ordre supérieur et classification en plusieurs zones tissulaires par fuzzy C-means

Une façon simple d'estimer les distributions de zones tissulaires est de calculer un certain nombre de statistiques d'ordre supérieur (SOS), qui approximent asymptotiquement les distributions étudiées [5]. Cependant, du fait de la variance des estimateurs proportionnelle à la taille de la population sur laquelle l'approximation est faite, nous nous limitons ici jusqu'aux statistiques d'ordre 4 (variance, Skewness et Kurtosis). Comme nous allons le montrer, l'utilisation des SOS apporte un meilleur résultat que la seule utilisation des statistiques d'ordre 2. De plus, les distributions de probabilités mises en jeu étant non-symétriques, l'utilisant du cumulatif d'ordre 3 normalisé (le Skewness) s'avère nécessaire. Il est à noter ici que les SOS peuvent être estimées à l'aide des estimateurs non biaisés : les k -statistiques [5].

Afin de pouvoir mettre en évidence les différences entre les zones tissulaires d'une coupe analysée, ces statistiques sont calculées par fenêtrage glissant. Nous obtenons ainsi un jeu de plusieurs caractéristiques qui peuvent être considérées comme une image multidimensionnelle (cf. Figure 2 à gauche). La taille de la fenêtre d'analyse est un paramètre que nous avons étudié quantitativement par rapport aux informations indiquées par l'anatomopathologiste.

La classification des différentes zones tissulaires a été ensuite réalisée en mode non-supervisé à l'aide d'un algorithme de classification de type fuzzy C-means (FCM) [3]. Cette méthode a l'avantage de pouvoir mettre en évidence les hétérogénéités tumorales, tout en étant moins sensible à l'initialisation aléatoire des

centres de classes [6]. Cependant, comme les caractéristiques extraites ont des plages de valeurs relativement différentes, une normalisation de chacune d'entre elles à une moyenne nulle et une variance unitaire a été réalisée. Le nombre de classes est également un paramètre important du fait de l'hétérogénéité des tissus analysés. Néanmoins, nous avons choisi empiriquement d'utiliser 4 classes, choix qui fournit des résultats relativement consistants lors de l'extraction des zones tumorales en comparaison avec l'analyse histologique classique effectuée sur des coupes adjacentes colorées à H&E.

4 Résultats

Les Figure 3 et 4 montrent les résultats obtenus pour deux adénocarcinomes de la prostate à degré de malignité élevé. Sur les images histologiques (H&E), notre anatomopathologiste a entourée 5 zones saines (en bleu) et 5 zones tumorales (en vert). Sur la Figure 3, la zone saine représentant la vessie a été entourée en orange. Les figures notées par b) représentent les valeurs des variances calculées pour chaque position de la fenêtre glissante, celles par c) les valeurs du Skewness tandis que les images d) représentent les zones tumorales obtenues par la méthodologie proposée. Ces résultats ont été obtenus avec une fenêtre rectangulaire de dimensions $100 \times 100 \mu\text{m}$ (couvrant environ une vingtaine de cellules), soit 500×500 pixels, et un recouvrement de $50 \mu\text{m}$. Du fait de l'hétérogénéité de l'échantillon, la classification a été effectuée en considérant 4 classes : zone tumorale (affichée sur les Figure 3 d) et 4 d)) et le reste (non affiché ici).

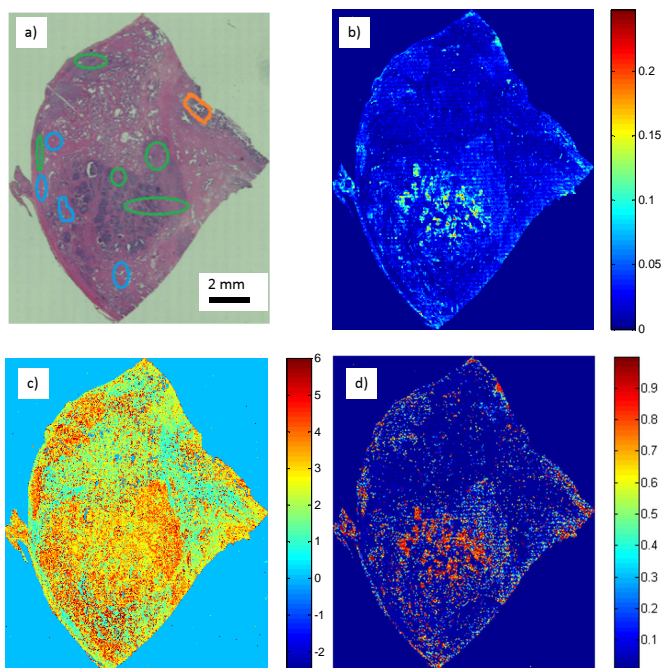


Figure 3 : Premier adénocarcinome ; a) image d'une coupe colorée (H&E) : zones saines entourées en bleu et zones tumorales en vert. La vessie, zone saine qui donne généralement des faux positifs, est entourée en orange ; b) et c) valeurs de variance et de Skewness pour une fenêtre rectangulaire de dimensions 500×500 pixels ; d) résultat de la méthode proposée, classe représentant les zones tumorales.

La variance seule peut être considérée comme une information pertinente dans la détection des zones tumorales. Bien que les statistiques d'ordre supérieur prises à part sont moins informatives puisque différentes zones saines présentent des valeurs élevées de Skewness ou Kurtosis (voir les valeurs de Skewness affichées sur les Figures 3 c) en bas et Fig. 4 c) en haut), leur prise en compte dans la méthodologie proposée apporte un gain dans la détection des zones tumorales, cf. ci-dessous.

Nous constatons une forte corrélation entre les résultats finaux et les images des coupes colorées à H&E. Ceci indique que les dégradations causées par le cancer à l'intérieur des cellules, notamment celles correspondant à un degré de malignité élevé, peuvent être mises en évidence avec succès en étudiant la distribution des valeurs des pixels.

Nous remarquons aussi que la zone correspondant à la vessie (entourée en orange dans l'image H&E) et qui peut introduire des fausses alarmes dans l'analyse histologique classique à H&E, n'est pas identifiée par notre méthode comme étant une zone tumorale.

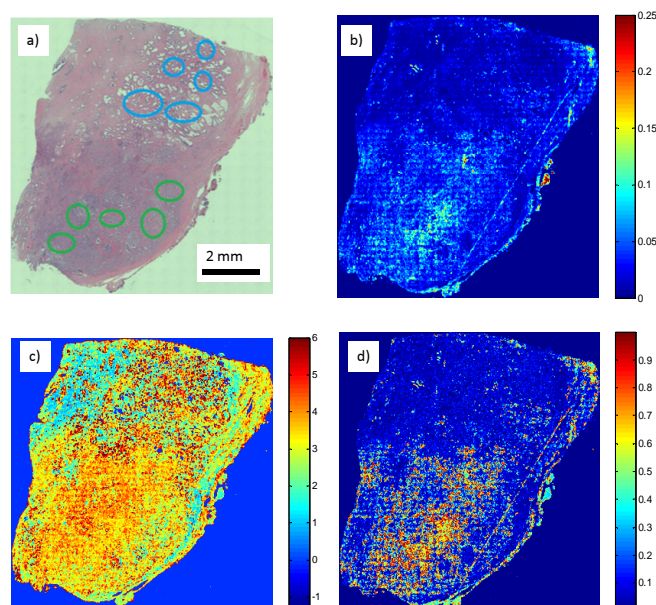


Figure 4 : Deuxième adénocarcinome : a) image d'une coupe colorée (H&E) : zones saines entourées en bleu et zones tumorales en vert ; b) et c) valeurs de variance et de Skewness pour une fenêtre rectangulaire de dimensions 500×500 pixels ; d) résultat de la méthode proposée.

Néanmoins, des artefacts apparaissent au bord de l'image. Ils sont principalement dus à la manipulation des coupes tissulaires. Il est à noter que les bords des tissus ne sont également pas pris en compte par les anatomopathologistes dans le cas des analyses classiques des coupes H&E.

Les zones saines/tumorales indiquées par notre anatomopathologiste (entourées sur les Figures 3 a) et 4 a)) ont été utilisées pour réaliser des mesures quantitatives de performance. Nous avons choisi ici comme mesures quantitatives de performance les courbes ROC [7] afin de s'intéresser à la capacité à discriminer les zones tumorales des zones saines.

Nous nous sommes intéressés dans un premier temps à étudier l'apport des SOS (Skewness et Kurtosis) par rapport à l'utilisation seule de l'information de variance. La figure 5 montre les courbes ROC obtenues pour le premier adénocarcinome (affiché sur la Figure 3) : en noir en prenant en compte uniquement la variance, en rouge les courbes ROC construites sur le résultat de classification en utilisant la variance, le Skewness et le Kurtosis. L'image à gauche a été obtenue pour une fenêtre d'analyse rectangulaire de dimensions 500x500 pixels et recouvrement de 50%, celle à droite pour une fenêtre rectangulaire de dimensions 200x200 pixels et même recouvrement.

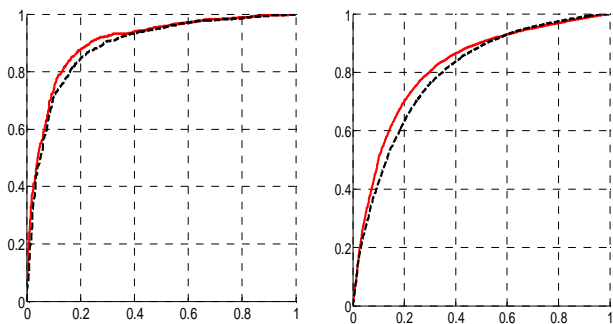


Figure 5 : Courbes ROC : variance seule ; variance, Skewness et Kurtosis pour l'adénocarcinome présenté sur la Figure 3. A gauche fenêtre rectangulaire de dimensions 500x500 pixels, à droite de dimension 200x200 pixels.

La figure 5 montre que l'utilisation des statistiques d'ordre supérieur apporte une information supplémentaire qui pourrait s'avérer utile pour le diagnostic du cancer par rapport à l'utilisation seule des statistiques d'ordre 2.

Nous avons également étudié la robustesse de la classification par rapport aux dimensions de la fenêtre glissante. La figure 6 montre les courbes ROC obtenues pour les deux adénocarcinomes en fonction de la taille de la fenêtre d'analyse.

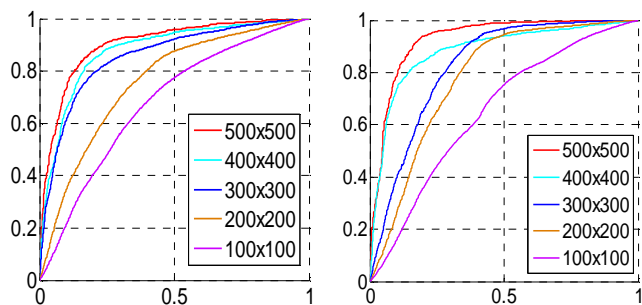


Figure 6 : Courbes ROC en fonction de la dimension de la fenêtre d'analyse (en pixels). A gauche : pour l'adénocarcinome présenté sur la Figure 3 ; à droite pour celui présenté sur la Figure 4. Dans les deux cas, les fenêtres glissantes sont rectangulaires et le recouvrement est de 50%.

Nous pouvons ainsi constater que plus la taille de la fenêtre est grande, plus la classification, et donc la détection de zones tumorales, est efficace. Cependant, nous constatons un effet de tassement dès lors que les dimensions des fenêtres sont assez importantes. Le

choix d'une fenêtre de dimensions 500x500 pixels, soit 100x100 μm , semble donc un bon compromis qualité (capacité à diagnostiquer le cancer) – résolution spatiale (50 μm pour un recouvrement de 50%) du résultat final. A cette échelle, la fenêtre d'analyse couvre une vingtaine de cellules environ.

5 Conclusion

Cette étude prospective montre qu'il est possible de mettre en évidence les zones tumorales d'une coupe tissulaire sans coloration préalable à l'aide d'une nouvelle technique : la SLIM, via l'étude de la distribution des valeurs mesurées à l'aide des statistiques d'ordre supérieur. Cette étude étant faite sur la distribution globale des pixels à l'intérieur d'une fenêtre d'analyse, les zones à fort degré de malignité ont principalement été mises en évidence. En effet, dans ce cas, le cancer modifie fortement la structure tissulaire et donc l'architecture des cellules. D'autres études sont à mener afin de réaliser une analyse plus fine des dégradations causées par le cancer à l'intérieur des cellules, comme par exemple en étudiant les statistiques locales dans des voisinages de taille relativement réduite. Ceci peut être par exemple réalisé à l'aide des matrices de cooccurrence, qui ont été introduites pour l'analyse des textures et utilisées récemment avec succès dans l'analyse et l'interprétation d'images des biopsies colorées à H&E [8]. En extrayant d'autres caractéristiques, et en les combinant d'une façon optimale, nous pouvons espérer pouvoir localiser des tumeurs à faible degré de malignité tout en gardant les principaux avantages de cette méthode d'acquisition : rapidité d'acquisition, pas de coloration préalable des tissus et donc réduction importante du coût total d'une analyse.

Références

- [1] LB. Rorke: Pathologic diagnosis as the gold standard. *Cancer* Vol. 79, No 4, pp. 665-667, 1997.
- [2] C. Kendall, M. Isabelle, *et al.* Vibrational spectroscopy: a clinical tool for cancer diagnostics. *Analyst*. Vol. 134. No 6. pp. 1029-1045, 2009.
- [3] D. Sebiskveradze, V. Vrabie, *et al.* Automation of an algorithm based on fuzzy clustering for analyzing tumoral heterogeneity in human skin carcinoma tissue sections, *Laboratory Investigation*, 2011, InPress.
- [4] Z. Wang, L. J. Millet, *et al.* Spatial light interference microscopy (SLIM). *Optics Express*, Vol. 19, N° 2, pp. 1016-1026, 2011.
- [5] V. Vrabie. Statistiques d'ordre supérieur : application en géophysique et électrotechnique. EUE, 2010.
- [6] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.
- [7] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, Vol. 27, pp. 861-874, 2006.
- [8] O.S Al-Kadi. Texture measures combination for improved meningioma classification of histopathological images. *Pattern Recognition*, Vol. 43, No 6, pp.2043-2053, 2010.