

Indices de formes : de la 2D vers la 3D

Application au classement de noyaux de cellules

Guillaume THIBAUT, Caroline DEVIC, Bernard FERTIL, Jean-Luc MARI et Jean SEQUEIRA

Université Aix-Marseille II
Laboratoire LSIS - UMR CNRS 6168
Equipe I&M (Images et Modèles)
ESIL - Campus de Luminy case 925
163 Avenue de Luminy 13288 Marseille

Abstract

In this paper, we present a study on the characterization and the classification of binary digital shapes. This study is performed using a set of values obtained by the computation of "shape indexes". To get these indexes, we extract a set of data called "measures" from the 2D shape, like for example a shape's surface and perimeter. Then these measures are used as parameters of a function bringing a real value which gives information about geometrical and morphological features on the shape to analyze. A model of shape and texture characterizations is built based on this numerical information. It is used to classify cells nuclei in order to diagnose patients affected by the Progeria disease.

Résumé : *Dans cet article, il est présenté une étude sur la caractérisation et la classification de formes numériques binaires, effectuée à l'aide d'un ensemble de valeurs obtenues par le calcul d'indices de formes. Pour calculer ces indices, il est nécessaire d'extraire de la forme un ensemble de données appelées "mesures", par exemple l'aire et le périmètre de la forme. Ces mesures sont ensuite utilisées comme paramètres d'une fonction apportant une valeur réelle ou discrète qui nous renseigne sur les caractéristiques géométriques et morphologiques de la forme ou du volume à analyser. Toutes ces informations numériques permettent de construire un modèle qui décrit la forme et qui peut être ensuite utilisé par exemple pour classer des formes et une application est proposée pour diagnostiquer des noyaux de cellules chez des patients atteints par la maladie de la Progeria.*

1. Introduction

La reconnaissance de formes est une partie majeure de l'intelligence artificielle qui vise à automatiser le discernement de situations typiques au niveau de la perception. Elle est un enjeu majeur pour de très nombreuses applications : la reconnaissance des caractères manuscrits (numérisation des livres, lecture automatique des lettres postales et des chèques bancaires, etc.), la vidéo surveillance (reconnaissance faciale), l'imagerie médicale (échographie, scanner, imagerie par résonance magnétique), la télédétection, etc.

Au cœur de la reconnaissance de formes, il y a une première étape incontournable : la caractérisation de formes. En effet, afin de pouvoir reconnaître un objet ou un individu, il faut tout d'abord le décrire et donc définir ses caractéristiques (morphologiques, géométriques, textuelles, etc.)

et ensuite retrouver et identifier ces mêmes caractéristiques sur la source numérique à analyser. Pour cela il est souvent nécessaire d'étudier les objets selon deux critères : le premier est la forme avec des méthodes d'analyse globale (par exemple signature polaire [THK06], histogrammes de projections [SR04], etc.) ou de contours (par exemple chaînes de Freeman [IPSV97], MSGPR [KR06], etc.) et le deuxième la texture [CDT95, Cos01, Mav01].

Pour réaliser l'analyse de la forme, la caractérisation à l'aide d'indices de formes est de plus en plus utilisée [AAB07, Ben02, IP97, TLG*03, URW*02, ZR04] et notamment en classification avec l'utilisation fréquente des méthodes par apprentissage [AGW97, SEB*03]. La souplesse d'utilisation, la simplicité de mise en oeuvre et la facilité d'utilisation avec un classificateur font de cette méthode un

choix pertinent pour de nombreux problèmes.

Le but de cet article est de créer un modèle afin de classer des noyaux de cellules de patients atteints par la maladie de la Progeria (également connue sous le nom de syndrome de Hutchinson-Gilford). Cette maladie orpheline (une centaine de cas dans le monde) de type laminopathie [SGBC*03] provoque un vieillissement accéléré du patient. Nous disposons d'un ensemble de noyaux de cellules prélevés chez les patients, ainsi que d'une expertise de ces noyaux. Cette expertise révèle que la forme du noyau (normale ou bien *boursouflée*) joue un rôle principal dans le diagnostic, mais des informations complémentaires sont obtenues par une analyse de la texture concernant l'homogénéité du noyau. Pour observer la texture, les images de noyaux sont acquises à l'aide d'un microscope à fluorescence et on utilise un marqueur de type FITC afin de voir la répartition des lamines AC.

La première partie de ce document définit les notions de mesure, d'indice de forme et leurs principales propriétés (section 2) ainsi que la façon d'utiliser les indices pour la reconnaissance de formes (section 2.1). Ensuite, le modèle mis en œuvre pour résoudre le problème est présenté (sections 3 et 4), étudié et validé.

2. Indices de formes et mesures

Les indices de formes ont été présentés pour la première fois en 1976 dans le livre de Santalo [San76] relatif aux propriétés mathématiques des formes convexes. On trouve la définition et les propriétés des indices de formes dans [CC85, Fil95].

Définition 2.1 (Indice de formes) *On appelle indice de formes tout paramètre, coefficient ou combinaison de coefficients permettant de donner des renseignements chiffrés sur la forme de l'objet.*

De plus, les indices doivent avoir les propriétés suivantes :

Propriété 2.2 (Indice de formes 1)

1. *Etre sans dimension.*
2. *Etre invariants par homothétie.*
3. *Etre invariants par rotation et translation.*
4. *S'appliquer à des formes connexes simples donc homéomorphes au disque.*

Le calcul d'un indice de formes est équivalent au calcul de la valeur d'une fonction à plusieurs variables. On aimerait alors que cette fonction soit bijective, mais deux contre-exemples sont donnés par les indices de concavité qui valent 1 pour toutes les formes convexes et l'indice d'allongement par le diamètre (défini plus loin [CC85]) pour le disque et le carré :

$$\text{Allongement}(\text{Disque}) = \text{Allongement}(\text{Carré}) = \frac{1}{2}$$

En revanche, la propriété suivante est vraie :

Propriété 2.3 (Indice de formes 2) *Soit F_1 une forme et I un indice de formes tel que $I(F_1) = \alpha$, avec $\alpha = cste \in \mathbb{R}$. Alors si pour toute forme F_2 quelconque, on a $I(F_2) \neq \alpha$ cela implique $F_1 \neq F_2$.*

Ce qui peut s'écrire :

$$\text{Si } \exists F_1, \exists F_2, \exists I / I(F_1) \neq I(F_2) \Rightarrow F_1 \neq F_2$$

Cette dernière propriété implique qu'un indice seul ne peut pas identifier une forme, mais qu'il "renseigne" sur la forme étudiée et sur les caractéristiques qu'elle ne possède pas.

La majorité des indices sont créés à partir d'une égalité ou d'une inégalité inhérente à la forme que l'on souhaite caractériser. Dans [Bon20] et [San76] les auteurs démontrent une série d'inégalités propres aux formes convexes dans un espace continu :

$$P^2 - 4\pi A \geq \frac{1}{2} \left[(P - 2\pi\rho_i)^2 + (2\pi\rho_e - P)^2 \right]$$

$$P^2 - 4\pi A \geq \pi^2 (\rho_e - \rho_i)^2 \text{ ou } P^2 - 4\pi A \geq \pi^2 (\rho_{max} - \rho_{min})^2$$

avec A l'aire, P le périmètre, ρ_i (resp. ρ_e) rayon de la plus grande (resp. petite) boule inscrite (resp. circonscrite [Gär99]), ρ_{max} (resp. ρ_{min}) le rayon de courbure maximum (resp. minimum). Toutes ces inégalités font intervenir différents paramètres de la forme que l'on appelle "*mesures*" (figure 1). Le calcul de ces mesures est une étape incontournable pour pouvoir calculer des indices de formes, car tous les indices utilisent au moins une mesure.

La possibilité de construire un indice à partir d'une égalité (ou inégalité) propre à la forme que l'on souhaite caractériser est l'avantage majeur des indices de formes. Cela permet d'utiliser les indices pour tout type de forme. De plus, la complexité de calcul d'un indice réside dans la complexité d'extraction des mesures qui le compose. Ces avantages rendent cette technique particulièrement bien adaptée en classification.

Définition 2.4 (Mesure) *On appelle "mesure" d'une forme toute valeur ou ensemble de valeurs numériques mesurées sur la forme.*

Les mesures possèdent ou non des dimensions : trois dimensions (le volume), deux dimensions (la surface), une dimension (le périmètre, le diamètre euclidien, la longueur de l'axe principal, etc.) et aucune dimension (le nombre de composantes connexes, le nombre de trous, etc.). Les mesures sans dimension possèdent toutes les caractéristiques de la propriété 2.2 et sont par conséquent des indices de formes.

2.1. Indices de formes et classement

Le but de la classification est d'associer à chaque individu étudié une classe d'appartenance. Dans le problème présent l'objectif est de déterminer si un noyau de cellule est "*sain*" ou "*pathologique*."

Les méthodes de classement se divisent en deux grandes

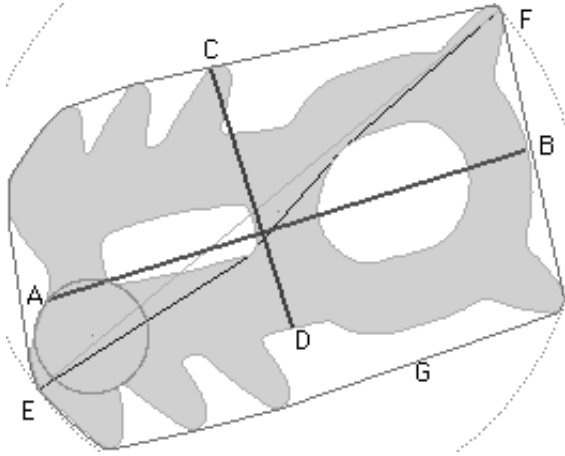


Figure 1: Exemples de mesures : surface (gris clair), AB et CD axes principaux, G enveloppe convexe, EF diamètre géodésique (noir) et diamètre Euclidien (gris clair), plus petite boule circonsrite (pointillés), A plus grande boule inscrite.

familles : les méthodes dites supervisées (appelées aussi méthodes de classement) et non supervisées. Les méthodes supervisées font intervenir une expertise et sont par conséquent généralement beaucoup plus puissantes. Dans notre cas, nous bénéficions de l'expertise des biologistes et généticiens qui ont déterminé les classes (sain et pathologique) et les sous classes (forme normale et forme boursoufflée, texture homogène et texture non homogène, etc.), ce qui nous permet d'utiliser des méthodes supervisées.

Les méthodes de classement sont ce que l'on appelle des méthodes d'apprentissages. Leur but est de construire un modèle de classement en fonction des données à classer. Bien que s'appliquant à un problème spécifique, le modèle doit être capable de généraliser (au sens des données). Pour cela, les données sont séparées en deux groupes : un échantillon d'apprentissage et un échantillon de validation. Le classificateur doit garder les mêmes performances lors de l'apprentissage et de la validation.

Pour mettre en œuvre une méthode de classement, il faut au préalable construire un vecteur caractéristique de l'individu étudié. Le vecteur doit être pertinent au problème posé afin de permettre un bon classement et une bonne prédiction. Le risque majeur lorsque l'on fournit trop de caractéristiques au classificateur est l'apprentissage par cœur. Plus la dimension du vecteur est grande, plus le modèle sera adaptable et donc plus le classement sera bon, mais plus la validation du modèle à l'aide d'individus non utilisés dans la phase d'apprentissage sera mauvaise. Il faut alors systématiquement valider chaque modèle construit et obtenir le meilleur compromis entre bonne classification et bonne prédiction. Dans notre problème, le vecteur caractéristique sera composé des indices de formes afin de caractériser la forme et

des caractéristiques Haralick [HSD73, Har79] pour la texture.

La méthode de classification choisie pour le modèle est la régression logistique [DS89]. C'est un modèle linéaire particulièrement adapté pour les problèmes de classement à deux classes. Elle réalise une analyse statistique des données de l'ensemble d'apprentissage et utilise une fonction de distribution logistique pour discriminer les données :

$$P = P(Y/x) = \frac{e^{f(x)}}{1 + e^{f(x)}}$$

avec $x = (x_1, \dots, x_n)$ le vecteur caractéristique de la donnée en entrée, $f(x) = \sum_i \alpha_i x_i$ et $P(Y/x)$ la probabilité conditionnelle P de la variable x d'appartenir à la classe Y . Pour estimer les coefficients α_i du modèle, on utilise le plus souvent la méthode du maximum de vraisemblance qui maximise la probabilité d'obtenir les valeurs observées sur les échantillons de l'ensemble d'apprentissage. Elle consiste à rechercher les paramètres qui optimisent la fonction de vraisemblance $\mathcal{L}(\alpha, Y) = P^Y [1 - P]^{1-Y}$.

3. Caractérisation de la forme des noyaux de cellules à l'aide des indices de formes

Nous disposons d'un ensemble de plus de trois mille noyaux de cellules (figure 2) prélevés chez des patients atteints par la maladie de la Progéria. Ces noyaux ont tout d'abord été classés manuellement et il est apparu que le critère de forme est l'élément de diagnostic décisif dans plus de 90% des noyaux. Il est ainsi primordial de construire un modèle de classification qui caractérise la forme du noyau. Pour cela, nous disposons d'un ensemble de douze indices de formes existant dans la littérature scientifique, mais il est nécessaire de créer trois nouveaux indices pour répondre plus spécifiquement au problème posé. La liste complète des indices utilisés est en annexe de cet article.

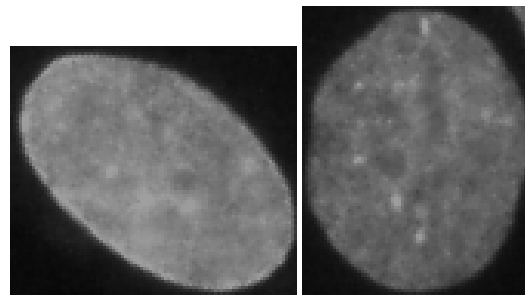


Figure 2: Deux noyaux de cellules.

3.1. Deux nouveaux indices de formes

Dans la section 2, nous avons vu que les indices de formes peuvent être créés à partir d'une égalité ou d'une inégalité

inhérente à la forme que l'on souhaite caractériser. Cette propriété est l'avantage majeur de cette technique, car elle rend l'utilisation des indices de formes applicable à tout type de forme.

Les noyaux de cellules étudiés ont une forme quasi elliptique lorsqu'ils sont sains. Il est donc utile de construire un indice qui permet de caractériser des ellipses. Il est basé sur la formule de l'aire d'une ellipse : $A = \pi ab$, avec a le demi grand axe et b le demi petit axe. Or dans le cas d'une ellipse, certaines mesures sont égales aux paramètres de l'aire : $R_{max} = a$ et $R_{min} = b$ ou encore $R_{max} = \frac{1}{2}L_{AP1}$ et $R_{min} = \frac{1}{2}L_{AP2}$, avec R_{max} (resp. R_{min}) le plus grand rayon (resp. plus petit) et L_{AP1} (resp. L_{AP2}) la longueur de l'axe principal (resp. la longueur de l'axe secondaire). Ces égalités permettent de construire deux indices de formes :

$$\Psi_{ellipse} = \frac{\pi R_{min} R_{max}}{A} \text{ ou } \frac{\pi L_{AP1} L_{AP2}}{4A} \in [0, 1]$$

Le dénominateur et le numérateur sont égaux dans le cas d'une ellipse et l'indice vaut 1.

Lorsque les noyaux ne sont pas sains, il est extrêmement fréquent qu'ils ne soient pas convexes et donc qu'ils possèdent des points de concavité. Pour compter ces points de concavité dans le cas des noyaux, il est possible de calculer le nombre de composantes connexes N_{Cce} issues de la soustraction de la forme à son enveloppe convexe. Cet indice sera utilisé sous sa forme normée $\Psi_{N_{Cce}}$ dans le classificateur :

$$N_{Cce} = \text{card}(\text{EnveloppeConvexe}(F) \setminus F)$$

$$\Psi_{N_{Cce}} = \frac{1}{1 + N_{Cce}} \in]0, 1]$$

Cet indice vaut 1 si aucune composante d'écart n'est trouvée et plus la forme a des points de concavité, plus l'indice tend vers 0.

Mais dans la pratique, on ne peut considérer comme composante connexe, des composantes dont la taille est de l'ordre du pixel et qui sont dues à des erreurs de discrétisation. De plus, dans les éléments de diagnostic des noyaux, la taille et le nombre des composantes connexes N_{Cce} doivent être pris en compte. Il faut trouver un ou plusieurs seuils de taille et de nombre à partir des éléments de diagnostic. Pour cela, nous avons réalisé une étude systématique du pourcentage de bonne classification en fonction de la taille et du nombre de composantes connexes d'écart (figure 3).

Cette analyse met en exergue l'utilité de l'indice $\Psi_{N_{Cce}}$ dans le cas de noyaux non convexes ayant un point de concavité d'au moins trente-deux pixels soit deux points de concavité d'au moins douze pixels (figure 4). Ces deux seuils sont utilisés de manière équivalente dans $\Psi_{N_{Cce}}$ et leur combinaison permet ainsi d'obtenir un taux de bonne classification de plus de 90% par rapport au sous diagnostic de forme avec ce seul indice.

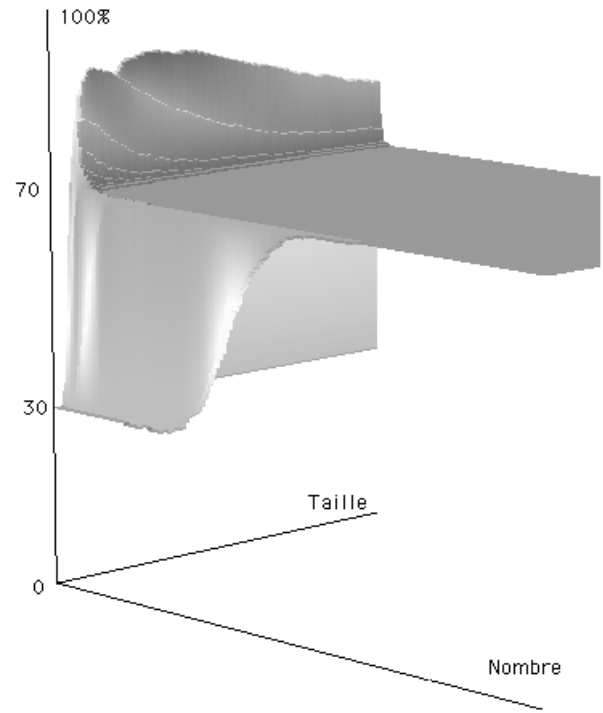


Figure 3: Surface représentant le pourcentage de bonne classification des noyaux en fonction du nombre et de la taille des composantes connexes d'écart.

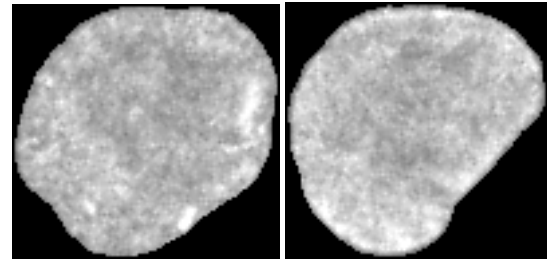


Figure 4: Concavité : à gauche, un noyau avec deux points de concavité d'au moins douze pixels, à droite, un noyau avec un seul point de concavité d'au moins 32 pixels.

3.2. Construction et validation du modèle de caractérisation de la forme des noyaux

Nous disposons maintenant d'un large éventail d'indices de formes afin de caractériser la forme des noyaux de cellules : les indices déjà existants dans la littérature et deux nouveaux indices créés pour répondre spécifiquement au problème traité. Il convient désormais de construire une combinaison linéaire des différents indices et de démontrer la pertinence du modèle créé pour répondre aux critères de formes.

Il faut choisir un nombre d'indices de formes pertinents afin d'obtenir un taux maximum de bonne classification et de validation, tout en conservant des propriétés de bonne généralisation (cf. section 2.1).

Pour cela, nous avons commencé par tester le modèle contenant tous les indices, puis au fur et à mesure les indices les moins pertinents ont été retirés. Pour la validation, chaque modèle a été construit à partir de 50% des noyaux choisis aléatoirement (résultats d'apprentissage, soit environ 1400 noyaux), puis validé sur les 50% restant (résultats de validation). Les figures 5 et 6 contiennent les différentes courbes construites pendant le test du modèle, sa validation ainsi que son efficacité.

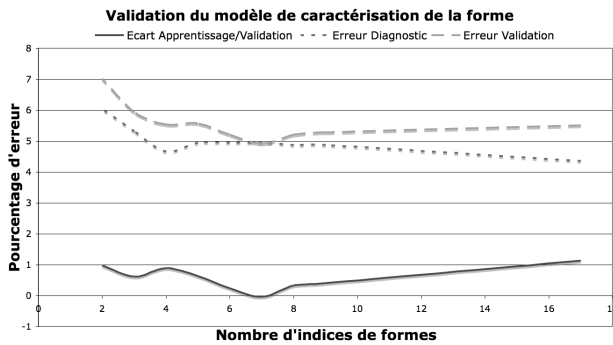


Figure 5: Pourcentage de bonne classification lors de la validation du modèle de caractérisation de la forme. Le graphe montre l'intérêt d'utiliser les sept indices les plus pertinents.

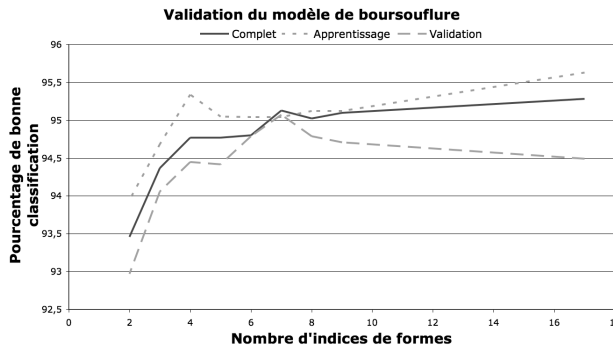


Figure 6: Pourcentage d'erreur lors de la validation du modèle : les deux courbes pointillées convergent l'une vers l'autre pour sept indices, puis diverge si on augmente le nombre d'indice.

Sur la figure 6 on peut constater que les courbes en pointillés représentant les erreurs de diagnostic et de validation convergent pour un nombre d'indices égal à sept. On voit également que l'écart entre ces courbes augmente de manière monotone au delà de sept indices. Ces deux

graphiques démontrent la validité de notre modèle qui permet un taux de bonne classification de la forme supérieur à 95% pour sept indices.

4. Caractérisation de la texture

Bien qu'ayant obtenu de très bons résultats avec le modèle de forme, celui-ci ne répond qu'à 89% du problème final, qui est la classification des noyaux en "sain" ou "pathologique". Pour compléter le modèle, il faut pratiquer une analyse de la texture et plus particulièrement une analyse de l'homogénéité du noyau.

Afin d'observer les noyaux, les experts utilisent un marqueur fluorescent de type FITC dans le microscope. Ce marqueur réagit à la présence des lamines AC dans le noyau et révèle leur répartition qui doit être homogène au centre et légèrement plus importante sur le bord. Mais en pratique, il est quasiment impossible que le marqueur se répartisse régulièrement sur la totalité du noyau. Cette inégalité de répartition pourrait laisser croire que les lamines AC ne sont pas réparties de manière homogène. Forts de cette information, les experts considèrent comme texture non homogène, uniquement les noyaux ayant une texture "fortement non homogène" (figure 7). Pour éviter ce problème pendant l'analyse, nous appliquons au préalable un filtrage passe-bas de type Gaussien sur la totalité de la texture :

$$I(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{\sigma^2}}$$

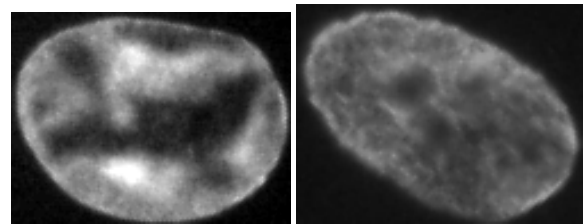


Figure 7: Deux noyaux de cellules possédant une texture fortement non homogène.

Pour caractériser la texture, nous construisons la matrice de co-occurrence, puis nous calculons neuf caractéristiques Haralick [HSD73, Har79] :

1. La moyenne $m = \sum_x \sum_y p(x, y)$
2. L'écart type $\sum_x \sum_y (p(x, y) - m)^2$
3. Le contraste $\sum_{n=0..N_g} n^2 \{ \sum_{i=1..N_g, |i-j|=n} \sum_{j=1..N_g} p(i, j) \}$
4. L'angle du moment du second ordre
5. La corrélation
6. L'entropie $\sum_x \sum_y p(x, y) \log(p(x, y))$
7. L'homogénéité $\sum_x \sum_y \frac{1}{1+(x-y)^2} p(x, y)$
8. La dissimilarité $\sum_x \sum_y |x - y| p(x, y)$
9. L'inertie $\sum_x \sum_y (x - y)^2 p(x, y)$

avec $p(x,y)$ l'élément (x,y) de la matrice de co-occurrence, N le nombre de pixels qui compose la texture à analyser, N_g le nombre de niveaux de gris. L'angle du moment du second ordre (que l'on appelle aussi *énergie*) mesure l'uniformité de la texture. Plus la texture est uniforme, moins il y a de transitions de niveaux de gris et donc plus la somme des carrés des éléments de la matrice est faible. L'homogénéité est d'autant plus élevée que l'on retrouve souvent le même couple de pixels, ce qui est le cas lorsque le niveau de gris est uniforme ou quand il y a périodicité spatiale. Les autres valeurs caractéristiques apportent d'autres informations sur la texture, moins intéressantes mais tout de même utilisées dans notre modèle.

Contrairement à l'expertise de la forme, l'expertise de la texture a fourni deux classes ("*homogène*" et "*non homogène*") totalement déséquilibrées ; il y a environ vingt fois plus d'individus dans la classe homogène. Pour construire efficacement le modèle, il faut équilibrer les classes pendant la phase d'apprentissage. Pour cela, l'apprentissage a été réalisé avec les noyaux dont la texture est "non homogène" et un nombre égal de noyaux à texture homogène choisis aléatoirement.

Le modèle de texture ainsi construit obtient un taux de bonne classification de 85.7% dans la phase d'apprentissage et 85.2% lors de la validation. Ce taux est bien inférieur à celui obtenu pour la forme[†]. Toutefois, il y a plusieurs explications à cet écart : la principale raison est le faible nombre d'individus appartenant à la classe "non homogène". Ce manque d'individu réduit les possibilités d'apprentissage. La deuxième raison est le biais introduit par le marqueur employé qui diminue la fiabilité du diagnostic.

5. Résultats, modèle complet : diagnostic des noyaux

Nous venons de construire deux modèles de classification qui caractérisent les deux principaux paramètres de diagnostic : la forme et la texture du noyau. Il faut maintenant combiner ces modèles afin de construire le modèle final qui prédit l'aspect anormal des noyaux. Ce dernier est une nouvelle combinaison linéaire des sept indices de formes du modèle de forme ainsi que des neuf coefficients du modèle de texture. Le taux de bonne classification obtenu par le modèle final est de 89.4% sur l'échantillon de validation.

6. Conclusions et perspectives

Nous venons de présenter une méthode de classement de noyaux de cellules de patients atteints par la maladie de la Progéria. Cette méthode repose essentiellement sur l'étude de la forme des noyaux à l'aide d'indices de formes. Dans un premier temps, nous avons construit des indices spécifiques au problème posé. Par la suite, ces indices ont

parfaitement répondu au sous problème et ont apporté un taux de bon classement supérieur à 95%. Dans un souci de fiabilité et de validation, il n'est pas envisagé de chercher à améliorer ce résultat pour deux raisons : la première raison est qu'il faudrait apporter de nouvelles caractéristiques et que l'on risquerait de faire de l'apprentissage par cœur sur les données. Les quelques dixièmes de bonne classification gagnés lors des tests seraient perdus lors de la validation. La deuxième raison est que ce résultat est équivalent à celui produit en général par un expert.

Par la suite nous avons construit un modèle basé sur les caractéristiques Haralick afin de répondre au sous problème de caractérisation de la texture. Ce modèle n'a pas répondu de manière très satisfaisante au problème posé[‡] et n'a pas apporté d'amélioration probante dans le modèle final. En effet, le modèle de forme obtient à lui seul un taux de 88.9% de bonne classification pour le problème final. L'ajout du modèle de texture n'a permis d'améliorer le diagnostic que de 0.5%. Par conséquent, il nous faudra améliorer notre modèle afin de répondre de manière plus pertinente au sous problème de texture.

Le modèle final permet un taux de bonne classification proche de 90% en n'utilisant que des méthodes de classification linéaires dont la puissance est limitée. On pourrait améliorer ce taux en employant des méthodes moins contraintes telles que les réseaux de neurones, mais il est aussi utile de chercher à améliorer les caractéristiques extraites des noyaux pour récupérer plus d'informations pertinentes pour la classification.

Au vu des résultats apportés, il est évident qu'il nous faut améliorer le modèle de texture. Pour cela, nous souhaitons étendre la notion d'indices de formes et l'utiliser dans l'analyse de textures. La solution que nous souhaitons aborder pour y parvenir est de considérer une texture comme une carte de hauteur ; chaque pixel ne code plus un niveau de gris, mais une hauteur (image 8).

Nous souhaiterions alors extraire les caractéristiques de ce volume sous la nappe. Ces caractéristiques seraient essentiellement les variations d'altitude tels que les pics (figure 9), les crêtes, les lacs (figure 10) et les failles. Une fois extraites et isolées, ces informations sont de type volumique. Il serait alors possible d'utiliser des indices de formes 3D afin de les caractériser. Ces indices seraient soit les indices existant en 2D que nous étendrions en 3D, soit de nouveaux indices spécifiques au problème posé en 3D. Ces caractéristiques de volumes renseigneraient sur les variations d'intensité de la texture et permettraient de la classer.

[†] supérieur à 95%

[‡] 85.7% de bonne classification

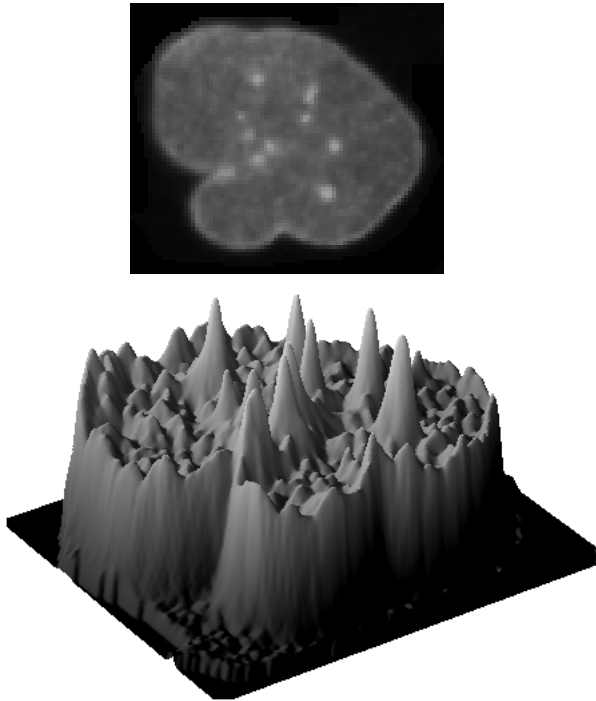


Figure 8: Exemple d'un noyau de cellule et de son volume sous la nappe.

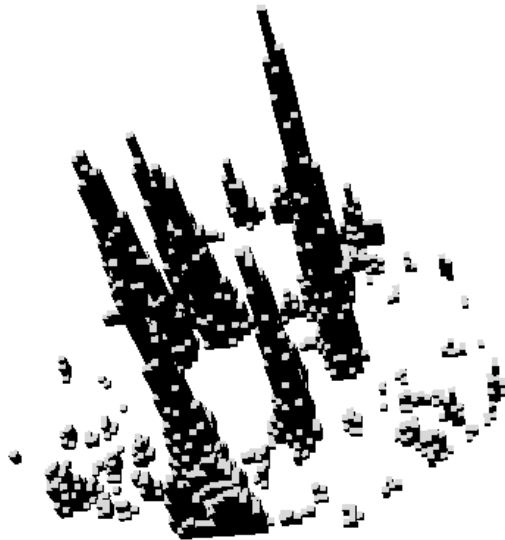


Figure 9: Les pics extraient du volume sous la nappe de la figure 8.

References

[AAB07] ARAGON C. R., ARAGON D. B., BERKELEY L.: A fast contour descriptor algorithm for supernova im-



Figure 10: Les lacs extraient du volume sous la nappe de la figure 8.

age classification. In *Real-Time Image Processing 2007* (1 2007), vol. 6496.

[AGW97] AMIT Y., GEMAN D., WILDER K.: Joint induction of shape features and tree classifiers. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1997), vol. 19, IEEE.

[Ben02] BENDJOUDI H.: The gravelius compactness coefficient: critical analysis of a shape index for drainage basins. *Hydrological Sciences* 47, 6 (12 2002), 921–930.

[Bon20] BONNESEN T.: *Les problèmes des Iso-périmètres et des iséphanes*. Gauthier Villars, 1920.

[CC85] COSTER M., CHERMANT J.-L.: *Précis d'analyse d'images*. Editions du CNRS, 1985.

[CDT95] CHEN Y. Q., DIXON M. S., THOMAS D. W.: Statistical geometric features for texture classification. In *Pattern recognition* (1995), vol. 28, pp. 537–552.

[Cos01] COSTA J.-P. D.: *Analyse statistiques de textures directionnelles*. PhD thesis, Université Bordeaux I, 12 2001.

[DS89] D. W. H., S. L.: *Applied Logistic Regression*. John Wiley & Sons, Toronto, 1989.

[Fil95] FILLERE I.: *Outils mathématiques pour la reconnaissance de formes*. PhD thesis, Université de St Etienne, Septembre 1995.

[Gär99] GÄRTNER B.: Fast and robust smallest enclosing ball. In *ESA '99: Proceedings of the 7th Annual European Symposium on Algorithms* (London, U, 1999), vol. 1643, Springer-Verla, pp. 325–338.

[Har79] HARALICK R. M.: Statistical and structural approaches to texture. In *Proceedings of the IEEE* (1979), vol. 67, pp. 786–804.

[HSD73] HARALICK R. M., SHANMUGAM K., DINSTEN I.: Textural features for image classification. In *IEEE Transactions on Systems, Man and Cybernetics* (1973), vol. 3, pp. 610–621.

- [IP97] IIVARINEN J., PEURA M.: Efficiency of simple shape descriptors. In *International Workshop on Visual Form* (5 1997).
- [IPSV97] IIVARINEN J., PEURA M., SÄRELÄ J., VISA A.: Comparison of combined shape descriptors for irregular objects. In *8th British Machine Vision Conference, BMVC'97* (Essex, Great Britain, 1997), vol. 2, pp. 430–439.
- [KR06] KPALMA K., RONSIN J.: Multiscale contour description for pattern recognition. In *IEEE Transactions on Pattern Recognition Letters* (2006), IEEE, pp. 1545–1559.
- [Mav01] MAVROMATIS S.: *Analyse de texture et Visualisation scientifique*. PhD thesis, Université de la Méditerranée, 2001.
- [San76] SANTALO L.: *Integral Geometry and Geometric Probability*. Addison Wesley, 1976.
- [SEB*03] SBONER A., ECCHER C., BLANZIERI E., BAUER P., CRISTOFOLINI M., ZUMIANI G., FORTI S.: A multiple classifier system for early melanoma diagnosis. *Published in Artificial Intelligence in Medicine* 27, 1 (2003), 29–44.
- [SGBC*03] SANDRE-GIOVANNOLI A. D., BERNARD R., CAU P., NAVARRO C., AMIEL J., BOCCACCIO I., LYONNET S., STEWART C. L., MUNNICH A., MERRER M. L., LEVY N.: Lamin a truncation in progeria, 2003.
- [SR04] SOLTANZADEH H., RAHMATI M.: Recognition of persian handwritten digits using image profiles of multiple orientations. In *IEEE Transactions on Pattern Recognition Letters* (2004), IEEE, pp. 1569–1576.
- [THK06] TOUMI A., HOELTZENER B., KHENCHAF A.: Classification des images isar pour la reconnaissance des cibles. In *XIIIème Rencontres de la Société Francophone de Classification (SFC)* (2006).
- [TLG*03] TUSET V. M., LOZANO I. J., GONZÁLEZ J. A., PERTUSA J. F., GARCÍA-DÍA M. M.: Shape indices to identify regional differences in otolith morphology of comber. In *Journal of Applied Ichthyology* (4 2003), vol. 19, p. 88.
- [URW*02] UNG A., RANCHIN T., WALD L., WEBER C., HIRSCH J., PERRON G., KLEINPETER J.: Cartographie de la pollution de l'air: une nouvelle approche basée sur la télédétection et les bases de données géographiques. application à la ville de Strasbourg. In *Journées CASSINI 2002 du GDR CASSINI-SIGMA* (September 2002).
- [ZR04] ZUNIC J., ROSIN P. L.: A new convexity measure for polygons. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2004), vol. 26, pp. 923–934.

Annexe : liste des indices de formes utilisés dans cet article

$$\text{Allongement}_{\text{Axes principaux}} = \frac{L_{AP2}}{L_{AP1}} \in [0, 1]$$

$$\text{Allongement}_{\text{Diametre}} = \frac{E}{D} \in [0, 1]$$

$$\text{Allongement}_{\text{Rayons}} = \frac{\rho_i}{\rho_e} \in [0, 1]$$

$$\text{Allongement}_{\text{Geodesique}} = \frac{4}{\pi} \frac{A}{D_g^2} \in [0, 1]$$

$$\text{Circularite} = \frac{R_{\min}}{R_{\max}} \in [0, 1]$$

$$\text{Circularite}_{\text{Courbure}} = \frac{\rho_{\min}}{\rho_{\max}}$$

$$\text{Convexite}_{\text{Perimetrique}} = \frac{P(\text{ConvexHull}(F))}{P(F)}$$

$$\text{Convexite}_{\text{Surfacique}} = \frac{A(F)}{A(\text{ConvexHull}(F))}$$

$$\text{Deficit} = 1 - \pi \frac{(\rho_e - \rho_i)^2}{\rho_e^2} \in [1 - \frac{\pi^2}{16}, 1]$$

$$\text{Deficit}_{\text{IsoPerimetrique}} = 4\pi \frac{A}{P^2} \in [0, 1]$$

$$\text{Ecart disque inscrit} = \frac{\pi \rho_i^2}{A} \in [0, 1]$$

$$\text{Etalement}_{\text{Morton}} = \frac{4}{\pi} \frac{A}{L_{AP}^2} \in [0, 1]$$

$$\text{Symetrie}_{\text{Besicovitch}} = \sup_{x \in F} \frac{A(F \cap \text{Symetrique}(F, x))}{A(F)}$$

$$\text{Variance circulaire} = \frac{1}{S_{\mu_r^2}} \sum_{p_i \in F} (\|p_i - B\| - \mu_r)^2$$

$$\Psi_{\text{Ellipse}} = \frac{\pi R_{\min} R_{\max}}{A} \text{ ou } \frac{\pi}{4} \frac{L_{AP1} L_{AP2}}{A} \in [0, 1]$$

$$\Psi_{N_{Cce}} = \frac{1}{1 + N_{Cce}} \in [0, 1]$$

$$\Psi_{\text{Parallelogramme}} = \frac{A}{E_{AP} L_{AP}} \text{ ou } \frac{A}{E \times D}$$