



Évaluation de la robustesse des descripteurs de texture pour la segmentation d'images de documents anciens

Maroua Mehri, van Cuong Kieu, Mohamed Mhiri, Pierre Héroux, Petra Gomez-Krämer, Mohamed Ali Mahjoub, Rémy Mullot

► To cite this version:

Maroua Mehri, van Cuong Kieu, Mohamed Mhiri, Pierre Héroux, Petra Gomez-Krämer, et al.. Évaluation de la robustesse des descripteurs de texture pour la segmentation d'images de documents anciens. CIFED 2014 - Actes du treizième Colloque International Francophone sur l'Écrit et le Document, Mar 2014, Nancy, France. pp.25-40. hal-00965160

HAL Id: hal-00965160

<https://hal.science/hal-00965160>

Submitted on 24 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Évaluation de la robustesse des descripteurs de texture pour la segmentation d'images de documents anciens

Maroua Mehri^{*,} — Van Cuong Kieu^{***} — Mohamed Mhiri^{****}
— Pierre Héroux^{**} — Petra Gomez-Krämer^{*}
— Mohamed Ali Mahjoub^{****} — Rémy Mullot^{*}**

**Laboratoire Informatique, Image et Interaction - L3i, Université de Rochelle, France, emails: {maroua.mehri, petra.gomez, remy.mullot}@univ-lr.fr*

***Laboratoire d'Informatique, du Traitement de l'Information et des Systèmes - LI-TIS, Université de Rouen, France, email: pierre.heroux@univ-rouen.fr*

****Laboratoire Bordelais de Recherche en Informatique - LaBRI, Université de Bordeaux I, Bordeaux, France, email: vkieu@labri.fr*

*****Systèmes Avancés en Génie Électrique - SAGE, Université de Sousse, Tunisie, emails: mhirimohamed@hotmail.com, medali.mahjoub@ipeim.rnu.tn*

RÉSUMÉ. Dans la littérature, le choix d'utiliser des descripteurs de texture pour guider la segmentation d'images de documents anciens soumises à des dégradations importantes et variées s'est déjà révélé pertinent. Par ailleurs, il a également été démontré que l'usage de tels descripteurs permet de se dispenser de connaissances préalables sur le modèle du document ou des paramètres typographiques. Nous proposons de rechercher des régions homogènes du point de vue des descripteurs de texture en se basant sur une analyse multi-résolution pour guider la segmentation d'images de documents. Les résultats préliminaires montrent l'efficacité des descripteurs de texture extraits de la fonction d'auto-corrélation, de la matrice de co-occurrence des niveaux de gris et des filtres de Gabor. Afin d'évaluer la robustesse de ces descripteurs, nous évaluons l'influence de différents modèles de dégradation et de prétraitements visant l'amélioration de la qualité d'image (filtrage par moyennes non-locales et technique de superpixels) au travers de plusieurs mesures de performances. Cette étude montre la robustesse des descripteurs de texture pour la segmentation en présence de bruit, rendant inutile les prétraitements de débruitage.

ABSTRACT. For the segmentation of ancient digitized document images, it has been shown that texture feature analysis is a consistent choice for meeting the need to segment a page layout under significant and various degradations. In addition, it has been proven that the texture-based approaches work effectively without hypothesis on the document structure, neither on the document model nor the typographical parameters. Thus, by investigating the use of texture as a

tool for automatically segmenting images, we propose to search homogeneous and similar content regions by analyzing texture features based on a multiresolution analysis. The preliminary results show the effectiveness of the texture features extracted from the autocorrelation function, the Grey Level Co-occurrence Matrix (GLCM) and the Gabor filters. In order to assess the robustness of the proposed texture-based approaches, images under numerous degradation models are generated and two image enhancement algorithms (non-local means filtering and superpixel techniques) are evaluated by several accuracy metrics. This study shows the robustness of texture feature extraction for segmentation in the case of noise and the uselessness of a denoising step.

MOTS-CLÉS : Images numériques de documents anciens, texture, multi-résolution, modèles de dégradation, débruitage, moyennes non-locales, superpixels

KEYWORDS: Ancient digitized document images, Texture, Multiresolution, Degradation Models, Enhancement, Non-local means, Superpixel

1. Introduction et approche proposée

Au sein du projet DIGIDOC¹, nous proposons des approches visant à simplifier et améliorer l’archivage, le traitement, la comparaison et l’indexation des images d’ouvrages anciens extraits de bibliothèques numériques telles que Gallica². En effet, de nombreuses difficultés sont susceptibles de se poser lors d’une analyse automatique d’images de documents anciens. Leur caractérisation s’avère compliquée par la présence de bruit, de dégradations et d’une grande variabilité dans leur organisation physique : mises en forme non-standardisées, alignement aléatoire, superposition de différents objets ou de différentes couches d’information (tampons, notes manuscrites, retombes, tâches, interférence du verso, ...). Pour ces raisons, les approches classiques de segmentation (par exemple XY-CUT (Khedekar *et al.*, 2003)) basées sur des connaissances *a priori* telles que la redondance d’une structure à travers le corpus, ne sont plus efficaces dans le traitement des documents anciens. Dans ce contexte, nous proposons d’analyser l’apport des descripteurs de texture dans les tâches de segmentation, de caractérisation de la structure physique en termes de zones de texture homogène permettant de concevoir finalement des mesures de similarité entre ouvrages numérisés selon différents critères.

Lors des vingt dernières années, plusieurs études (Journet *et al.*, 2008 ; Mehri *et al.*, 2013) se sont penchées sur la caractérisation et l’indexation des images de documents anciens et se basant sur des descripteurs de texture. Dans nos travaux précédents (Mehri *et al.*, 2013), nous avons présenté une étude comparative de trois descripteurs classiques liés à l’analyse de texture (la fonction d’auto-corrélation (Petrou et Sevilla, 2006), la matrice de co-occurrence des niveaux de gris (Haralick *et al.*, 1973), les filtres de Gabor multi-canaux (Gabor, 1946)) pour la segmentation d’images de documents anciens. Ces descripteurs de texture sont extraits uniquement pour les pixels correspondant au tracé (obtenus par un algorithme de binarisation) d’une image en niveaux de gris. Les descripteurs de texture sont calculés sur quatre fenêtres d’analyse de différentes tailles afin d’implémenter une approche multi-résolution et centrées sur les pixels sélectionnés. Une classification ascendante hiérarchique (CAH) (Lance et Williams, 1967) est ensuite appliquée sur les valeurs normalisées des descripteurs de texture. Chaque pixel, en fonction des valeurs de descripteurs de texture qui lui sont associées, intègre alors l’un des regroupements de la CAH. Le nombre de regroupements maximum étant fixé au nombre de types de contenu défini dans la vérité terrain. Dans ce même article, les différents descripteurs de texture sont présentés, les performances qu’elles permettent d’atteindre sont évaluées par le biais de différentes mesures d’évaluation de classification non-supervisée mais uniquement sur une base réduite d’images de documents anciens volontairement simplifiées. Cette étude a notamment montré que les meilleures performances sont obtenues grâce aux indices de texture produits par les filtres de Gabor.

1. DIGIDOC ANR-10-CORD-0020 est financée par l’ANR

2. <http://gallica.bnf.fr>

Le présent article prolonge l'étude précédente en évaluant la robustesse des descripteurs présentés. La robustesse est tout d'abord attestée en examinant l'évolution des mesures de performance en présence de plusieurs modèles de dégradation dont l'intensité est contrôlée. Par ailleurs, la robustesse est également évaluée en montrant que des étapes préalables de filtrage du bruit sur les images n'ont que très peu d'influence sur la qualité des résultats.

Cet article est structuré de la façon suivante. La section 2 présente succinctement deux prétraitements pour l'amélioration et le débruitage des images. La génération d'images semi-synthétiques de documents anciens basée sur différents modèles de dégradation est détaillée en section 3. La section 4 présente les protocoles expérimentaux mis en œuvre pour évaluer la robustesse des descripteurs de texture ainsi que les résultats obtenus. Finalement, les conclusions de cette étude et les perspectives envisagées sont données en section 5.

2. Prétraitements

La problématique de l'amélioration des images de document a déjà fait l'objet de nombreuses études. Likforman-Sulem *et al.* (Likforman-Sulem *et al.*, 2009), par exemple, comparent et combinent deux techniques de prétraitement (filtrage par moyennes non-locales (Buades *et al.*, 2005) et la variation totale (Darbon et Sigelle, 2006)) pour l'amélioration et le débruitage des images de documents imprimés. Dans cette section, nous présentons succinctement deux techniques de prétraitement utilisées par la suite pour l'étude de la robustesse. Nous présentons d'une part, la technique de filtrage par moyennes non-locales, utilisée dans de nombreux systèmes, et d'autre part, un prétraitement d'amélioration/débruitage basé sur une technique de superpixels.

2.1. Technique de filtrage par moyennes non-locales

La technique de filtrage par moyennes non-locales (Buades *et al.*, 2005) associe au pixel traité une moyenne pondérée des pixels du voisinage. Les pondérations des pixels sont déduites des similarités entre le voisinage du pixel analysé et les voisinages des voisins situés dans la fenêtre de recherche. Likforman-Sulem *et al.* (Likforman-Sulem *et al.*, 2009) fixent les paramètres par défaut des moyennes non-locales à 3 et 4 respectivement pour la taille du voisinage et la taille de la fenêtre de recherche, et ceci pour l'amélioration de documents historiques imprimés. Par l'utilisation de l'approche basée sur les moyennes non-locales, les auteurs améliorent les résultats de la reconnaissance optique de caractères pour un faible niveau de dégradation. Cette approche de débruitage par moyennes non-locales ayant montré son efficacité dans le traitement des documents anciens, nous l'intégrons en fixant ses paramètres aux valeurs les plus communément admises dans la littérature (Likforman-Sulem *et al.*, 2009). La Figure 2(d) montre l'exemple d'une image ayant été améliorée par le biais de cette technique. On y remarque notamment que le fond du document fait encore

apparaître du bruit et qu'à l'œil nu, il n'existe pas de différence notable avec l'image originale (cf. Figure 2(c)).

2.2. Technique de superpixels

La technique dite de superpixels a récemment été introduite et utilisée dans des applications de vision par ordinateur, notamment pour la segmentation (Fulkerson *et al.*, 2009), l'estimation de bruit gaussien (Wu et Chang, 2012),... Les approches à base de superpixels sont devenues des alternatives crédibles à l'utilisation d'une grille régulière et rigide de pixels. Plusieurs algorithmes ont ainsi été proposés dans la littérature (Achanta *et al.*, 2012). Ces approches permettent des traitements plus rapides, requérant une capacité mémoire moindre, en offrant la possibilité de ne calculer les descripteurs de texture qu'à l'échelle des superpixels plutôt qu'à celui, plus dense, des pixels. Achanta *et al.* (Achanta *et al.*, 2012) proposent une catégorisation des méthodes existantes en trois classes : les méthodes à base de graphes, les méthodes basées sur le gradient et les méthodes SLIC (Simple Linear Iterative Clustering). Les techniques de superpixels reposent sur le regroupement de pixels voisins partageant des caractéristiques similaires (texture, contour, couleur,...) au sein de régions polygonales. Par conséquent, les techniques de superpixels produisent une image sur-segmentée représentant une carte compacte du contenu initial de l'image (Achanta *et al.*, 2012).

Cohen *et al.* (Cohen *et al.*, 2013) proposent de séparer le tracé du fond et du bruit présent sur les images de documents anciens en utilisant les caractéristiques spatiales et colorimétriques des superpixels. Par le biais d'une nouvelle fonction objectif, Liu *et al.* (Liu *et al.*, 2011) proposent un algorithme réalisant une segmentation en superpixels. Cette fonction objectif est basée sur l'équilibrage et l'entropie. Le terme de cette fonction dépendant de l'entropie vise à proposer des regroupements homogènes du point de vue de la texture, alors que celui relatif à l'équilibrage vise à assurer la compacité des regroupements en équilibrant leur taille. Dans une première étape, l'image sur-segmentée est représentée par un graphe pondéré dans lequel chaque nœud représente un pixel et dont les arcs sont pondérés par la similarité entre les pixels représentés par les nœuds extrémités. Le graphe est ensuite partitionné en n sous-graphes maximisant la fonction objectif par une approche glouton (Nemhauser *et al.*, 1978). Les auteurs démontrent que l'approche à base de graphes qu'ils proposent se comporte mieux que les approches alternatives de l'état de l'art lorsque le nombre de superpixels est faible (Liu *et al.*, 2011).

Afin d'obtenir une amélioration du fond sur les images de documents anciens bruitées, nous proposons un algorithme de séparation fond/forme basé sur une segmentation en superpixels utilisant le critère de (Liu *et al.*, 2011) et une classification des superpixels à base de k-moyennes (MacQueen, 1967). La segmentation en superpixels permet de réduire la combinatoire de la séparation fond/forme en ramenant la classification au niveau des superpixels plutôt qu'au niveau des pixels. Nous appliquons dans un premier temps à l'image en niveaux de gris du document une segmen-

tation en superpixels en fixant n , le nombre de superpixels, à 5% du nombre initial de pixels. Chaque superpixel est alors affecté de la moyenne des niveaux de gris des pixels qu'il contient. Une binarisation correspondant à une séparation fond/forme est obtenue par application de l'algorithme des k-moyennes aux superpixels affectés de leur niveau de gris sans tenir compte de leur localisation et en fixant le paramètre k à 2. L'image du document est débruitée en affectant le niveau de gris correspondant au blanc à tous les pixels contenus dans les superpixels associés au fond, tandis que les niveaux de gris originaux sont conservés inchangés pour les pixels contenus dans les superpixels de la couche forme. La Figure 2(e) illustre le résultat de l'application de la technique de débruitage proposée.

3. Modèles de dégradation

Le plus souvent dans la littérature, des dégradations synthétiques sont générées afin d'estimer le niveau de bruit ou pour l'évaluation des algorithmes d'amélioration ou de restauration d'images (Wu et Chang, 2012). Dans cet article, nous utilisons les modèles de dégradation afin d'évaluer la robustesse des descripteurs de texture. Ces modèles de dégradation permettent de générer des images semi-synthétiques modélisant le bruit et des effets liés à la dégradation des supports, et nous pouvons ainsi examiner l'évaluation des performances en fonction des niveaux de perturbation appliqués.

Nous utilisons d'une part un modèle de dégradation uniforme sur la globalité du document basé sur un bruit additif gaussien (*baG*). Les images semi-synthétiques (*cf.* Figure 2(l)) sont générées en ajoutant à chaque pixel de l'image initiale (*cf.* Figure 2(c)) une valeur issue d'une distribution normale centrée et ayant un écart-type de 35.

Par ailleurs, nous générons des images semi-synthétiques par application de modèles de dégradation locaux. Trois modèles de dégradation locaux ont été proposés dans (Kieu *et al.*, 2013) : tâches indépendantes (*It*), tâches recouvrantes (*Rt*), tâches de non-connexion (*NCt*). Le bruit *It* est généré en ajoutant des tâches sombres ou claires selon qu'elles sont situées sur le tracé ou sur le fond du document comme montré sur la Figure 1(b). Le bruit *Rt* est généré en ajoutant les tâches sombres ou claires connectées au bord des formes de tracé comme indiqué sur la Figure 1(c). Enfin, le bruit *NCt* est généré de telle sorte que des tâches claires soient ajoutées de façon à rompre la connexion de certains tracés comme indiqué sur la Figure 1(d). La génération d'une telle dégradation est opérée en sélectionnant des points germes sur l'image originale binarisée. Dès lors, une tâche claire ou sombre, elliptique et centrée sur chacun des points germes est ajoutée à l'image originale pour générer des images semi-synthétiques. Trois autres modèles de dégradation locaux : faible (*Fad*), moyen (*Md*) et fort (*Fod*), sont générés par combinaison des modèles *It*, *Rt* et *NCt*, dont les proportions sont tirées aléatoirement, en fixant le nombre de points germes et la taille des tâches. Pour chaque modèle parmi les six de dégradation locaux présentés précédemment, deux jeux d'images semi-synthétiques sont générés,

soit 12 jeux d'images au total. Les caractéristiques des modèles de dégradation sont présentées dans le Tableau 1.

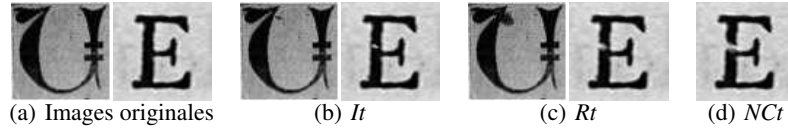


Figure 1. Illustration des trois types de dégradation locale sur des images semi-synthétiques.

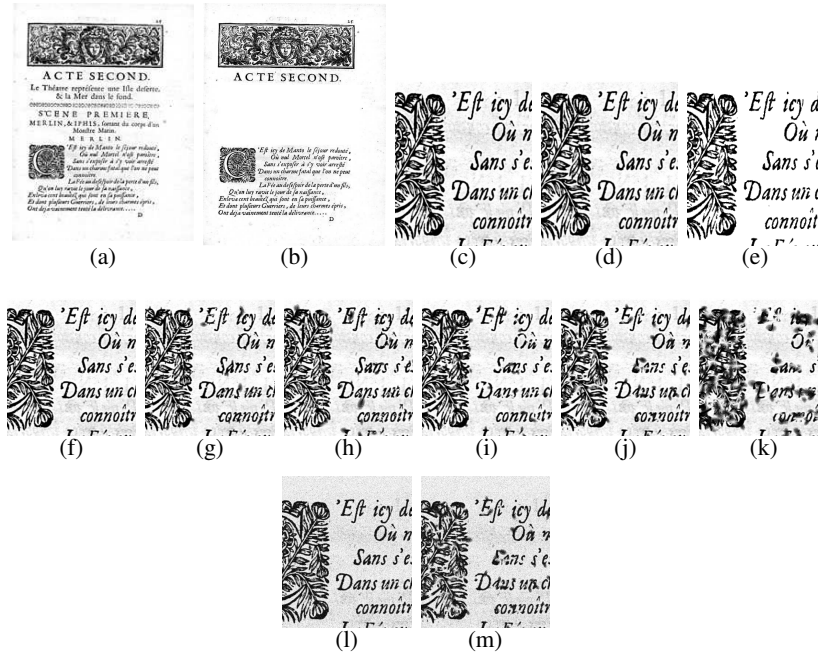


Figure 2. Constitution de la base d'images utilisée pour l'évaluation : (a) représente une image d'un document ancien réel, (b) illustre un exemple d'image simplifiée par occlusion de zones, (c) représente un détail de l'image (b). Les Figures (d) et (e) illustrent respectivement le résultat du prétraitement par le filtre à base de moyennes non-locales et par la technique de superpixels. Les Figures (f), (g), (h), (i), (j) et (k) illustrent respectivement les modèles de dégradation appliqués à (b) via It2, Rt2, NCt2, Fad2, Md2 et Fod2. Les Figures (l) et (m) montrent les résultats de la dégradation de (b) respectivement par le bruit additif gaussien et par baGMd2 (un mélange du bruit additif gaussien baG et de Md2).

Finalement, un mélange (noté *baGMd2*) du modèle de dégradation *Md2* et du bruit additif gaussien *baG* est également appliqué aux images afin d'évaluer la robustesse des descripteurs de texture au mélange de bruit uniforme et local.

Tableau 1. *Modèles de dégradation locaux*

	Nombre de points germes	Modèles / Taille des tâches
<i>It1</i>	2000	Tâches indépendantes
<i>It2</i>	2500	Tâches indépendantes
<i>Rt1</i>	1500	Tâches recouvrantes
<i>Rt2</i>	2000	Tâches recouvrantes
<i>NCt1</i>	1000	Tâches de non-connexion
<i>NCt2</i>	1500	Tâches de non-connexion
<i>Fad1</i>	1000	Faible / 5
<i>Fad2</i>	1000	Faible / 8
<i>Md1</i>	1500	Moyen / 7
<i>Md2</i>	1500	Moyen / 9,5
<i>Fod1</i>	2500	Fort / 10
<i>Fod2</i>	2500	Fort / 12

4. Évaluation et résultats

4.1. Protocole expérimental

Considérant la variabilité en termes de types de contenu et de mise en forme existant dans les documents anciens, l'évaluation fine des performances des méthodes de segmentation est une opération difficile. En effet, la diversité des documents anciens réels ne permet pas de contrôler le contexte dans lequel les méthodes sont évaluées. Afin de remédier à cela, nous avons, à partir de documents réels (*c.f* Figure 2(a)), généré une base de documents dans lesquels certaines zones ont été occultées. Nous obtenons ainsi des documents (*c.f* Figure 2(b)) dont la complexité est certes réduite mais dont on connaît les éléments. Il nous est alors possible de déterminer les performances des différents descripteurs de texture étudiés en fonction du contexte. À partir de quatre images de documents réels (*c.f* Figure 2(a)), contenant des éléments graphiques (ornements, textures, gravures,...) et du texte (avec différentes polices de caractères), nous générons un jeu de 25 images simplifiées (*c.f* Figure 2(b)). À partir de ce jeu d'images simplifiées nous générons :

- 50 images prétraitées par le filtre des moyennes non-locales et la technique de superpixels
- 350 images produites par application d'un modèle de dégradation :
 - 25 images dégradées par un bruit additif gaussien
 - 300 images dégradées par application de modèles de dégradation locaux (25 images simplifiées $\times$ 6 modèles de dégradation locaux $\times$ 2 jeux d'images semi-synthétiques)
 - 25 images bruitées par *baGMd2* : un mélange de *Md2* et du bruit additif gaussien *baG*

Dans les premières expériences, nous comparons de façon qualitative les performances de la segmentation des images utilisant les jeux de descripteurs de texture décrits dans (Mehri *et al.*, 2013) (auto-corrélation, matrice de co-occurrence des niveaux de gris et filtre de Gabor). Par la suite, afin d'évaluer la robustesse au bruit des descripteurs de texture sur les images prétraitées et les images bruitées, chaque image est segmentée en analysant les indices de descripteurs extraits par une approche multi-résolution. L'extraction des trois descripteurs sur les images prétraitées produit un total de 150 images analysées (50 images prétraitées $\times$ 3 descripteurs). De même, l'extraction de ces mêmes descripteurs de texture sur les images bruitées produit un total de 1050 images analysées (350 images dégradées $\times$ 3 descripteurs). Les résultats de la segmentation sont comparés à la vérité terrain, et sont finalement caractérisés via plusieurs mesures de performance d'évaluation de classification non-supervisée et classification supervisée. D'une part, nous évaluons des résultats de la classification non-supervisée au regard de la largeur de silhouette *SW* et de la pureté par bloc *PPB* (Mehri *et al.*, 2013). Par ailleurs, en associant les regroupements produits par la classification non-supervisée aux classes de la vérité terrain, nous évaluons les résultats au regard de la précision *P*, du rappel *R* et du taux de bonne classification. Pour chacune de mesures, une forte valeur indique de bons résultats.

4.2. Descripteurs de texture

Les Figures 3(a), 3(f) et 3(k) illustrent de façon qualitative les performances obtenues par les descripteurs de texture dans la tâche de discrimination des pixels correspondant à la couche tracé sur les images de documents anciens et notamment leur capacité à séparer la couche graphique de la couche texte. Sur ces illustrations, on remarque en particulier que les descripteurs à base d'auto-corrélation et ceux basés sur la matrice de co-occurrence des niveaux de gris (*cf* Figures 3(a) et 3(f)) donnent des résultats satisfaisant pour la séparation texte/graphique. On constate qu'en revanche, les descripteurs basés sur les filtres de Gabor regroupent les éléments graphiques et les caractères majuscules des parties de texte correspondant aux titres. Ce type d'erreur peut s'expliquer par les difficultés qu'ont les descripteurs de Gabor à séparer des éléments proches mais de types différents. En l'occurrence, l'espace vertical entre les deux zones est trop faible. En effet, les descripteurs de Gabor sont extraits pour certaines bandes de fréquence et dans des directions choisies. Les performances des descripteurs de Gabor sont donc dépendantes du contexte dans lequel apparaît la zone à extraire. Cependant, en analysant les descripteurs de Gabor, on observe qu'ils permettent la distinction de différents types de graphiques (Figure 3(k)). Au sein d'un même document, chaque regroupement produit par un algorithme de classification non-supervisée appliqué au descripteur de texture est identifié par une couleur et représente des textures considérées comme homogènes du point de vue des descripteurs concernés.

Les résultats quantitatifs obtenus sur les 25 images de documents simplifiées et mesurés via plusieurs mesures de performance sont présentés sur la Figure 4 (*cf*.

barre bleue). Les meilleures performances sont obtenues via les descripteurs de Gabor pour la quasi-totalité des mesures avec 0.33%(*SW*), 96%(*PPB*), 93%(*P*), 90%(*R*) et 93%(*CA*) et ce, sans prendre en considération les relations de voisinage entre les pixels. Cette information contextuelle n'est considérée que lors de l'extraction des descripteurs de texture qui est opérée sur une fenêtre d'analyse centrée sur le pixel considéré. Ceci contredit la conclusion déduite à partir de la Figure 3(k) en utilisant l'approche Gabor, mais on peut expliquer ce résultat par la dominance des pixels de graphique qui sont bien classés par rapport à ceux du texte. En revanche, on observe également que cette qualité de résultats est obtenue au prix d'une consommation importante des ressources informatiques. En effet, les descripteurs de Gabor sont ceux qui, parmi ceux présentés dans cette étude requièrent le plus de mémoire et de temps de traitement. Le temps requis pour le traitement d'une page (1965×2750 pixels) en utilisant l'auto-corrélation et les filtres de Gabor est de 2 et 6 minutes respectivement alors que le traitement de la même page via les descripteurs basés sur la co-occurrence ne requiert que 14 secondes. Les valeurs des mesures de performances de classification supervisée observées dans ce travail sont en concordance avec celles mentionnées dans (Mehri *et al.*, 2013), *i.e.* les valeurs des mesures de performances de classification non-supervisée.

4.3. Prétraitements

Dans les premières expériences, nous comparons de façon qualitative les performances de la segmentation des images utilisant les jeux de descripteurs de textures décrits dans (Mehri *et al.*, 2013) (auto-corrélation, matrice de co-occurrence des niveaux de gris et filtre de Gabor), sur des images nonprétraitées (*cf.* Figures 3(a), 3(f) et 3(k)), sur des images prétraitées par le filtre à base de moyennes non-locales (*cf.* Figures 3(b), 3(g) et 3(l)) et sur celles prétraitées par la technique de superpixels (*cf.* Figures 3(c), 3(h) et 3(m)). L'examen visuel des résultats montre des résultats relativement similaires que les images aient été prétraitées ou pas. On constate même une légère détérioration des résultats produits à base de descripteurs de Gabor sur les images prétraitées par la technique de superpixels (*cf.* Figure 3(m)) ainsi que sur ceux produits à partir des matrices de co-occurrence des niveaux de gris sur les images prétraitées par le filtre à base de moyennes non-locales (*cf.* Figure 3(g)). À l'opposé, une amélioration est obtenue sur les résultats produits par la matrices des co-occurrences des niveaux de gris prétraitées par la technique de superpixels (*cf.* Figure 3(h)). Cette amélioration s'explique par le fait que le prétraitement ramène l'ensemble des pixels du fond du document à une valeur unique. Ce prétraitement rend alors plus représentatif et tend à stabiliser les estimateurs des probabilités de co-occurrence de niveaux de gris faisant intervenir les pixels représentant le fond du document.

La Figure 4 (*cf.* colonne de gauche) présente les résultats quantitatifs sur les images originales, sur celles prétraitées par le filtre des moyennes non-locales et sur celles prétraitées par la technique de superpixels. On remarque un léger gain de performance

(1%(SW), 0%(PPB), 2%(P), 1%(R) et 3%(CA)) pour les descripteurs basés sur la matrice de co-occurrence des niveaux de gris sur les images prétraitées par la techniques de superpixels. Cette constatation confirme celle exprimée dans l'analyse qualitative des résultats. Cependant, nous observons dans la plupart des cas des performances similaires lorsqu'on compare les résultats obtenus sur les images originales et ceux sur les images prétraitées. Nous en concluons que les prétraitements par le filtre des moyennes non-locales et la technique de superpixels ne permettent pas d'améliorer les résultats obtenus via les descripteurs de texture présentés.

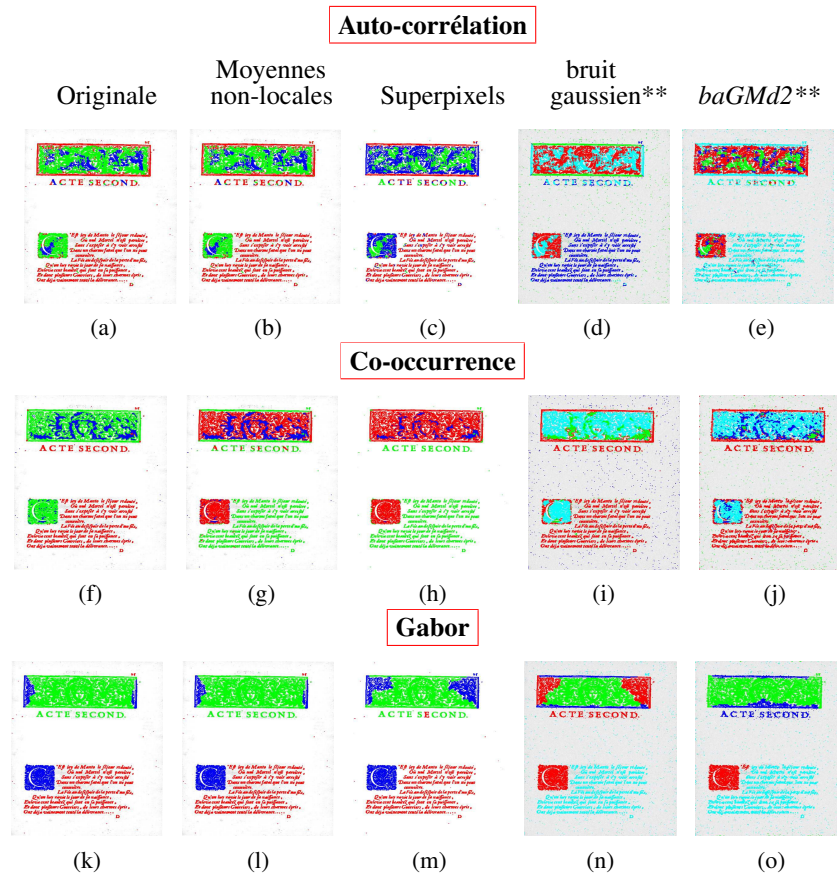


Figure 3. Résultats de segmentation d'images simplifiées de documents anciens (cf. Figures 3(a), 3(f) et 3(k)), sur des images prétraitées par le filtre basé sur les moyennes non-locales (cf. Figures 3(b), 3(g) et 3(l)), sur la technique de superpixels (cf. Figures 3(c), 3(h) et 3(m)) et sur des images dégradées par un bruit additif gaussien (cf. Figures 3(d), 3(i) et 3(n)). Les Figures 3(e), 3(j) et 3(o) montrent le résultat de la classification non-supervisée sur des images soumises à baGMd2 : un mélange du bruit Md2 et du bruit additif gaussien baG. Du fait que le traitement soit non-supervisé, les couleurs associées aux regroupements peuvent changer d'une image à l'autre.

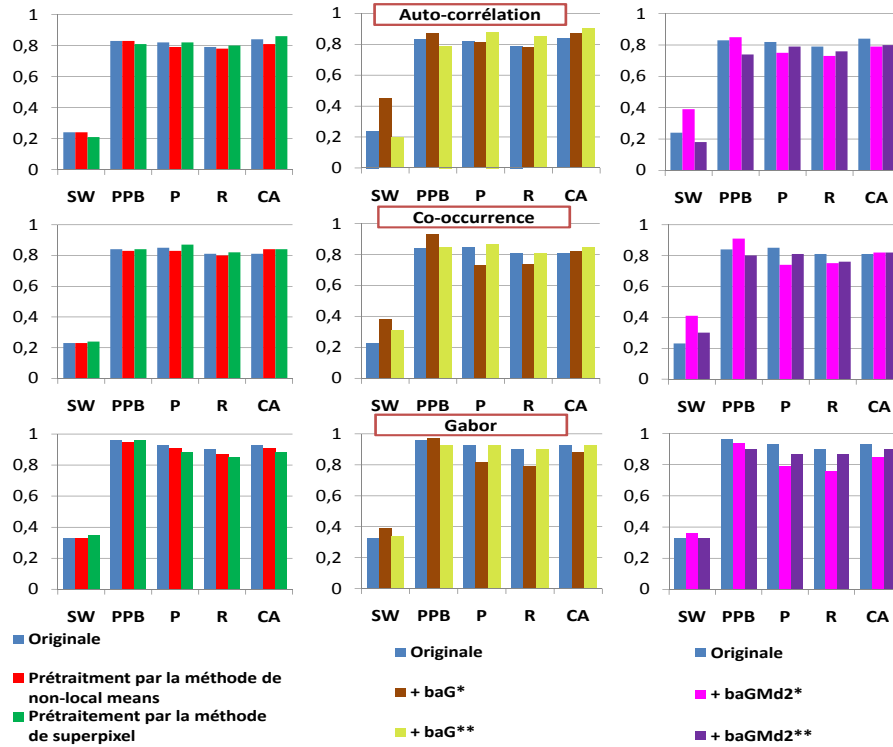


Figure 4. Évaluation des performances obtenues via les descripteurs de texture sur les images de document prétraitées (filtre à base de moyennes non-locales et technique de superpixels) et sur des images dégradées de façon synthétique : largeur de silhouette (SW), pureté par bloc (PPB), précision (P), rappel (R) et taux de classification (CA). Les “*” et “**” représentent respectivement les cas où la classification est opérée en fixant le nombre de types de contenu égal à celui défini dans la vérité terrain, et ce même nombre augmenté d’une unité (afin de prendre en considération un type de contenu associé au bruit). baGMd2 est un mélange de Md2 et du bruit additif gaussien baG.

4.4. Modèles de dégradation

Afin d’évaluer la robustesse au bruit des descripteurs de texture, nous comparons les performances que ces descripteurs permettent d’atteindre sur les images originales et sur celles soumises à un modèle de dégradation tel que décrit en section 3. L’examen visuel des résultats (cf. Figures 3(d), 3(i), 3(n), 3(e), 3(j) et 3(i)) montre des résultats identiques sur les images bruitées lorsque le nombre de regroupements est fixé à 4, soit le nombre de types de contenu dans la vérité terrain augmenté d’une unité afin de

prendre en considération un type de contenu pour le bruit. On remarque notamment que les descripteurs permettent de discriminer les pixels correspondant au bruit de ceux correspondant au texte et au graphique. On observe sur la Figure 3(o), que les descripteurs issus des filtres de Gabor permettent en partie de séparer les différents types de graphiques ainsi que les différentes polices de caractères. Les résultats de la classification non-supervisée sont meilleurs sur les images bruitées (cf. Figure 3(o)) à partir des descripteurs de Gabor que ceux sur les images originales (cf. Figure 3(k)). Ceci peut être justifié par le fait que les directions et les bandes de fréquence spécifiées pour paramétrer ce descripteur sont totalement décorréliées du bruit aléatoire ajouté aux images. Ainsi, les pixels correspondant au bruit sont tous associés au même regroupement.

Les Figures 4 et 5 donnent les résultats quantitatifs permettant de comparer les performances obtenues sur les images originales et celles obtenues sur les images soumises à des modèles de dégradation. On remarque que lorsqu'on applique la dégradation par un bruit additif gaussien on obtient des performances similaires en prenant en considération un regroupement pour le bruit. Une légère baisse de performance au regard des mesures *SW* et *PPB* peut être observée sur les approches basées sur la fonction d'auto-corrélation lorsque les images sont soumises à un bruit additif gaussien. Cependant, on obtient en moyenne de meilleures performances au regard des mesures *P*, *R* et *CA*. Ce gain de performance provient d'une bonne classification des pixels liés au bruit qui ne sont pas présents dans la vérité terrain des images originales. Ce phénomène illustre le fait que pour une évaluation plus juste, il nous aurait fallu disposer au lieu d'une information de vérité terrain définie au niveau du bloc comme c'est le cas actuellement, d'une information de vérité terrain définie au niveau des pixels. Mais la définition de cette information est une tâche fastidieuse, coûteuse et sujette à des erreurs et des interprétations différentes. On observe également une légère baisse de performance moyenne pour les trois jeux de descripteurs lorsque les images sont soumises aux modèles de dégradation locaux : *Fad1*, *Fad2*, *Md1*, *Md2*, *Fod1* et *Fod2*.

Nous présentons dans le Tableau 2 quelques unes des différences selon les différentes mesures de performances obtenues sur les images originales et celles provenant d'images soumises à des dégradations. En moyenne, pour les trois types de descripteurs, les performances chutent pour les forts niveaux de dégradation : *Fod1* et *Fod2*. Les performances baissent légèrement et progressivement avec le niveau de dégradation. Nous observons également un léger gain pour l'ensemble de descripteurs issu de la matrice de co-occurrence des niveaux de gris lorsque les images sont soumises au modèle local *It2*. Cette évolution, bien que peu significative, peut tout de même être expliquée par la faculté de ce descripteur à discriminer les différentes textures et en particulier à identifier les pixels correspondant au bruit. Lors de l'ajout du bruit produit par *baGMd2*, la différence en termes de performance est relativement faible quel que soit le jeu de descripteurs considéré. Nous concluons que les descripteurs de texture étudiés sont robustes au bruit de natures et d'intensités différentes.

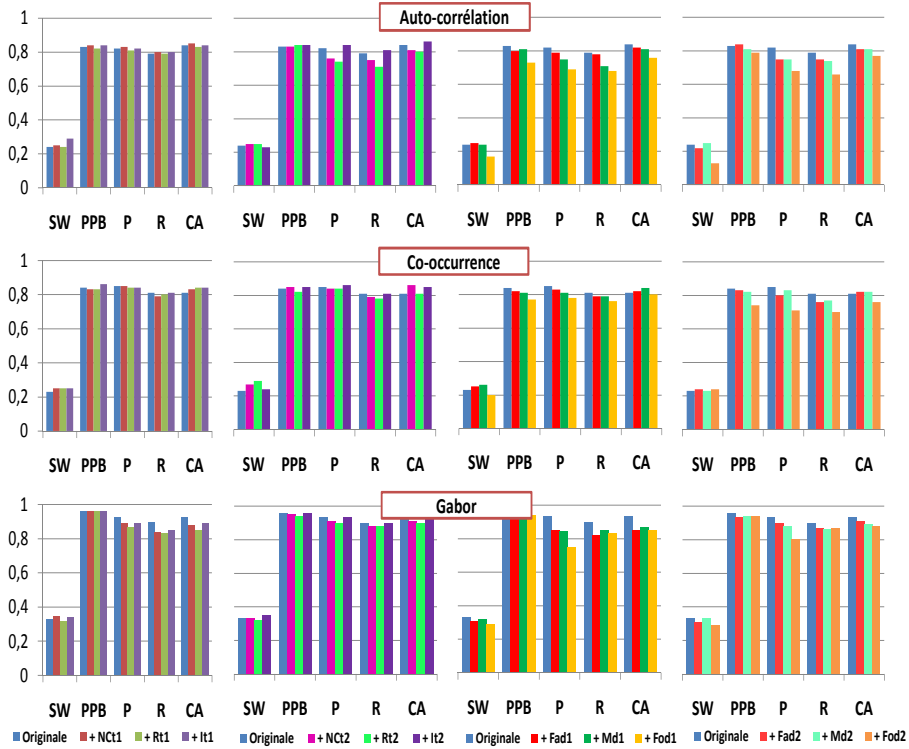


Figure 5. Évaluation des performances obtenues via les descripteurs de texture sur des images soumises à des modèles de dégradation : largeur de silhouette (SW), pureté par bloc (PPB), précision (P), rappel (R) et taux de bonne classification (CA).

Tableau 2. Différences de performance observées sur les images bruitées ou dégradées par rapport à celles obtenues sur les images originales.

	Descripteur	SW	PPB	P	R	CA
<i>baG</i>	Gabor	1	-3	0	0	0
<i>baG</i>	Auto-corrélation	-4	-4	6	6	6
<i>Fod1</i>	Gabor	-4	-2	-18	-7	-8
<i>Fod2</i>	Gabor	-4	-2	-13	-3	-5
<i>It2</i>	Co-occurrence	1	1	1	0	4
<i>baGMd2</i>	Auto-corrélation	-6	-9	-3	-3	-4
<i>baGMd2</i>	Co-occurrence	-7	-4	-4	-5	1
<i>baGMd2</i>	Gabor	0	-6	-6	-3	-3

5. Conclusions et perspectives

L'étude décrite dans cet article vise à évaluer la robustesse des descripteurs de texture pour une tâche de segmentation d'image de document anciens. Trois jeux de descripteurs sont évalués : la fonction d'auto-corrélation, la matrice de co-occurrence des niveaux de gris et les filtres de Gabor multi-canaux. Dans un premier temps, l'étude a porté sur l'influence des algorithmes de prétraitements visant à améliorer la qualité de l'image. Les algorithmes étudiés sont d'une part, le filtrage à base de moyennes non-locales et, d'autre part, une technique à base de superpixels. En nous appuyant sur le calcul de plusieurs mesures de la qualité de la classification supervisée ou non-supervisée, nous concluons sur le fait que ces étapes de débruitage se révèlent inutiles pour les descripteurs de texture étudiés. En effet, l'étape de débruitage n'apporte pas d'amélioration significative des résultats de la segmentation. L'évaluation de la robustesse a également été opérée en mesurant l'évolution des mesures de performance lorsque les descripteurs étaient extraits sur des images de documents sur lesquelles ont été appliqués différents modèles de dégradation. Les observations ont également indiqué que les descripteurs de texture étudiés étaient robustes à plusieurs modèles de dégradation de différentes intensités.

Nos travaux futurs dans la lignée de ceux décrits dans cet article viseront à évaluer la robustesse d'autres descripteurs de texture tels que ceux basés sur les ondelettes, les LBP et les indices de Tamura. À plus long terme, nous envisageons de combiner plusieurs descripteurs afin de proposer un jeu de descripteurs optimal.

6. Bibliographie

- Achanta R., Shaji A., Lucchi A., Fua P., Süsstrunk S., « SLIC Superpixels Compared to State-of-the-Art Superpixel Methods », *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, n° 11, p. 2274-2282, 2012.
- Buades A., Coll B., Morel J. M., « A Non-Local Algorithm for Image Denoising », *CVPR*, p. 60-65, 2005.
- Cohen R., Asi A., Kedem K., El-Sana J., Dinstein I., « Robust text and drawing segmentation algorithm for historical documents », *HIP*, p. 110-117, 2013.
- Darbon J., Sigelle M., « Image Restoration with Discrete Constrained Total Variation Part I : Fast and Exact Optimization », *JMIV*, vol. 26, p. 261-276, 2006.
- Fulkerson B., Vedaldi A., Soatto S., « Class segmentation and object localization with superpixel neighborhoods », *ICCV*, p. 670-677, 2009.
- Gabor D., « Theory of communication. Part 1 : The analysis of information », *Journal of the Institution of Electrical Engineers - Part III : Radio and Communication Engineering*, vol. 93, n° 26, p. 429-441, 1946.
- Haralick R., Shanmugam K., Dinstein I., « Textural features for image classification », *SMC, IEEE*, p. 610-621, 1973.

- Journet N., Ramel J., Mullot R., Eglin V., « Document image characterization using a multi-resolution analysis of the texture : application to old documents », *IJDAR*, vol. 11, n° 1, p. 9-18, 2008.
- Khedekar S., Ramanaprasad V., Setlur S., Govindaraju V., « Text - Image Separation in Devanagari Documents », *ICDAR*, p. 1265-1269, 2003.
- Kieu V. C., Visani M., Journet N., Mullot R., Domenger J. P., « An Efficient Parametrization of Character Degradation Model for Semi-synthetic Image Generation », *HIP*, p. 102-109, 2013.
- Lance G. N., Williams W. T., « A general theory of classificatory sorting strategies 1. Hierarchical systems », *The Computer Journal*, vol. 9, n° 4, p. 373-380, 1967.
- Likforman-Sulem L., Darbon J., Smith E. H. B., « Pre-Processing of Degraded Printed Documents by Non-local Means and Total Variation », *ICDAR*, p. 758-762, 2009.
- Liu M. Y., Tuzel O., Ramalingam S., Chellappa R., « Entropy Rate Superpixel Segmentation », *CVPR*, p. 2097-2104, 2011.
- MacQueen J. B., « Some Methods for Classification and Analysis of MultiVariate Observations », *Berkeley Symposium on Mathematical Statistics and Probability*, p. 281-297, 1967.
- Mehri M., Gomez-Krämer P., Héroux P., Boucher A., Mullot R., « Texture Feature Evaluation for Segmentation of Historical Document Images », *HIP*, p. 102-109, 2013.
- Nemhauser G. L., Wolsey L. A., Fisher M. L., « An analysis of approximations for maximizing submodular set functions-I », *Mathematical Programming*, vol. 14, n° 1, p. 265-294, 1978.
- Petrou M., Sevilla P. G., *Image Processing : Dealing with texture*, John Wiley & Sons, 2006.
- Wu C. H., Chang H. H., « Gaussian noise estimation with superpixel classification in digital images », *CISP*, p. 373-377, 2012.

Remerciements

Ce travail s'appuie sur le support de l'Agence Nationale de la Recherche Française (ANR) dans la cadre du projet DIGIDOC sous contrat ANR – 10 – CORD – 0020. Les auteurs tiennent également à remercier Geneviève Cron de la Bibliothèque Nationale de France (BnF).