



RAPPORT DE MASTER RECHERCHE – Juin 2004

Caractérisation de tracés manuscrits et recherches de similarités en vue de la compression par approche fractale

Abir EL ABED

Dea STIR « Signal, Télécommunications, Image et Radar » Option
Signal et Image

LIRIS UMR 5205, INSA de Lyon

abir_elabed@yahoo.fr

Encadrement :

Véronique EGLIN & Frank LEBOURGEOIS

REMERCIEMENTS

REMERCIER est nécessaire à tout travail qui réussit à voir le jour. Une page de ma vie est en train de se tourner, me laissant seule face à un avenir difficile à imaginer, car imprécis au moment où ces moments sortent de mon esprit. Tous ceux que je remercie sincèrement le savent bien, et ceci sans avoir à préciser le moindre nom, prénom ou surnom. Mais, il vous serait difficile à vous, lecteurs de ce manuscrit, de le savoir, aussi vais-je vous y aider en vous confiant quelques indices.

Je tiens tout particulièrement à remercier Monsieur Hubert EMPTOZ (Professeur et chef du laboratoire RFV/LIRIS à INSA de Lyon) qui est à l'origine de ce travail, qui m'a accueillie au laboratoire et a apporté son soutien en tant que directeur de stage.

Je remercie également Madame Véronique EGLIN et Monsieur Frank LEBOURGEOIS (Maîtres de conférence au laboratoire RFV/LIRIS à INSA de LYON) d'avoir accepté de diriger ce stage, d'avoir mis le doigt sur les « points durs », de m'avoir donnée des conseils précieux, et de m'avoir permis de mener ce travail à terme.

L'aboutissement de ce travail a vu également le jour grâce au jury, qui a bien voulu y apporter sa caution scientifique. De ce fait, je vous remercie tous.

Tout au long de ce stage, j'ai vraisemblablement été « invivable » par mes sautes d'humeur et négligente envers ma famille, en laissant passer le travail au premier plan. Qu'ils me pardonnent de tout ce que je leur ai fait subir et je les remercie vivement de la patience dont ils ont fait preuve à mon égard.

Enfin, merci à ceux dont le nom m'est parfois inconnu et qui ont contribué efficacement, par leur sourire, à rendre mes journées plus sereines et à vous qui êtes en train de lire ce début du rapport.

Merci à vous tous...

*« L'auteur, dans son œuvre,
doit être présent partout,
et visible nulle part »*

Flaubert

SOMMAIRE

REMERCIEMENTS.....	1
SOMMAIRE	4
TABLE DE FIGURES	6
I. INTRODUCTION GENERALE.....	7
I.1 Généralité autour de la numérisation du patrimoine.....	7
I.2 Autour d'un projet national : ACTION CONCENTREE INCITATIVE MADONNE.....	8
I.3 Présentation générale du problème	8
I.4 Plan de la mémoire.....	10
II. LA COMPRESSION DES DOCUMENTS	11
II.1 La compression générique (avec perte) : JPEG.....	11
II.2 La compression adaptée avec minimum de pertes d'informations utiles : DjVu.....	12
II.3 La compression dans le cadre du projet DEBORA	13
II.4 Recherche d'auto-similarités - la compression fractale.....	14
II.5 Vers une méthodologie sans segmentation et adaptée aux caractéristiques des tracés ..	16
III. NOTRE APPROCHE : LOCALISATION DE ZONES D'INTERET AU NIVEAU DU GRAPHEME.....	17
III.1 Première approche : Analyse différentielle des manuscrits	18
III.2 Notre seconde approche : Analyse fréquentielle des manuscrits et filtrage des données selon les directions émergentes de l'image	20
III.2.1 Passage par la fonction d'autocorrélation	21
III.2.2 Principe de la détection des orientations	21
III.2.3 La Rose de Directions : aspects techniques	22
III.3 Quantification et analyse des esquisses directionnelles	23
III.3.1 Extraction automatique des directions principales.....	23
III.3.2 Le filtre de Gabor : un outil inspiré de la perception visuelle humaine	23
III.3.3 Paramétrage des filtres de Gabor sur des images de manuscrits en niveaux de gris..	24
III.3.4 Exploitation des réponses de Gabor et construction des cartes directionnelles	26
III.4 Conclusion.....	27

IV. EXPLOITATION DES ESQUISSES ET MESURES DE SIMILARITES PARTIELLES.....	28
IV.1 Traitement appliqué aux cartes directionnelles.....	28
IV.2 Formation de connexités par opérations morphologiques et classification des esquisses	30
IV.3 Estimation du taux de similarité entre esquisses.....	30
IV.3.1 Taux de similarité et transformations affines au sein d'une même carte directionnelle (<i>intra-carte</i>).....	30
IV.3.2 Taux de similarités et transformations affines entre cartes directionnelles (<i>inter-carte</i>)	32
IV.4 Critère de recherche de l'esquisse de référence pour chaque carte directionnelle.....	32
IV.5 Intégration de l'étape de recherche de similarités dans un processus de compression adaptée aux tracés manuscrits	34
IV.6 Discussion et perspectives.....	35
IV.6.1 Synthèse et bilan positif des travaux.....	35
IV.6.2 Limitations dans la recherche de similarités	36
IV.6.3 Perspectives directes et travaux à venir	37
IV.7 Conclusion générale	38
REFERENCES BIBLIOGRAPHIQUES.....	39

TABLE DE FIGURES

Figure 1 : Extraits de deux ouvrages manuscrits de scripteurs différents présentant de faibles espaces interlignes et une non linéarité apparente des mots.....	9
Figure 2 : Image originale (a).....	12
Figure 3 : Différences visuelles entre les modes de compression (AT&T)	13
Figure 4 : Comparaison entre les taux de compression sur des documents anciens	14
Figure 5 : (a) image originale, (b) image décompressée par la méthode de la compression fractale.....	15
Figure 6 : (a) image originale, (b) partitionnement carré de la compression fractale, (c) image décompressée selon la méthode fractale	15
Figure 7 : Localisation des points d'intérêt sur une image de manuscrit arabe à l'aide du détecteur de Harris sur des images en niveaux de gris.....	19
Figure 8 : Rose de direction de la sélection entière et zoom sur les directions principales représentées par des pics locaux.....	22
Figure 9 : Spectre du filtre de Gabor traduisant l'amplitude des réponses des filtres $\theta = 0^\circ$ et $\theta = 45^\circ$ avec respectivement une sélection en fréquences $\omega = 30$ Hz et $\theta = 120$ Hz.	24
Figure 10 : Filtrage de GABOR, séparation en plans directionnels binarisés.....	26
Figure 11 : a) réponse de Gabor selon la direction 90° (et 0° dans (c)) ; b) et d) Dilatation adaptative des formes connexes	29
Figure 12 : Erosion adaptative des petites surfaces.....	29
Figure 13 : illustration des traits similaires à une rotation près (pour la direction de 90°)	31
Figure 14 : illustration de traits de similarités entre différente carte directionnelle à des transformations géométriques près (rotation et changement d'échelle).....	32
Figure 15 : illustration des liens existant entre esquisses (estimées par filtre de Gabor à 40°) et les graphèmes originaux de l'image en niveaux de gris.....	34
Figure 16 : filtrage d'une écriture latine médiévale dans la direction de 90° par le filtre de Gabor.....	36
Figure 17 : application du filtre de Gabor sur une écriture de Montesquieu dans la direction de 45° et résultat de la dilatation adaptative aux traits de la réponse.....	37

I. INTRODUCTION GENERALE

I.1 Généralité autour de la numérisation du patrimoine

Nos bibliothèques regorgent d'ouvrages riches en informations (images de documents anciens ou historiques, des catalogues de vente par correspondance, des formulaires, des archives gouvernementales ou commerciales, des publications scientifiques, techniques,...). La numérisation de documents du patrimoine scientifique et culturel facilite l'accès à un public plus large et offre de nouveaux services comme la consultation en ligne de documents rares, la duplication rapide et économique des ouvrages, la navigation simplifiée, une recherche d'information plus performante et enfin la possibilité de partager des connaissances avec d'autres lecteurs ([1], [2], [4]).

A l'heure actuelle, il n'existe pas d'outils réellement adaptés à une exploitation à distance des collections numérisées. Jusqu'à maintenant, le besoin majeur était de stocker et d'archiver les données provenant généralement de collections et de corpus volumineux, sans se soucier réellement de la qualité de la transmission puisque les documents étaient généralement consultés à partir de CD Rom ou sur des réseaux locaux sans risque de pertes de paquets de données ([4]).

Du point de vue de l'utilisateur de la bibliothèque numérique, on peut constater un besoin de plus en plus grand de fonctionnalités qui ne peuvent pas encore être offertes en temps réel. Les usagers cherchent d'abord une plus grande accessibilité aux ouvrages, aux images et à leurs contenus. L'accessibilité traduit une possibilité de consultation en ligne facile et rapide, ainsi qu'une possibilité d'interrogation avancée des contenus pour trouver l'information désirée. Ces besoins nécessitent la mise en place de méthodes de compression des images adaptées plus efficaces ainsi que la définition de formats de données qui permettent une interrogation à distance de leurs contenus. Les formats doivent aussi permettre de gérer tous les contenus par définition très hétérogènes d'une bibliothèque numérique (textes, images, structures, annotations par des objets multimédia, hyperliens...) et doivent pouvoir être robuste aux perturbations et aux erreurs de transmission.

I.2 Autour d'un projet national : ACTION CONCENTREE INCITATIVE MADONNE

Le patrimoine culturel et scientifique de l'Europe est un bien public unique qui représente la mémoire collective et vivante de nos différentes sociétés. Le projet ACI MADONNE s'inscrit dans une démarche de sauvegarde et de valorisation de données patrimoniales (<http://I3iexp.univ-lr.fr/madonne/>). L'aspect « masse de données » est adressé sous l'angle des collections d'ouvrages numérisés, représentés sous forme d'images scannées, pour lesquelles les techniques traditionnelles des bases de données sont inopérantes. L'exploitation et la valorisation à venir de ces collections d'images n'ont pas encore trouvé de réponses satisfaisantes, du fait même de leur caractère faiblement structuré. La génération de ces entrepôts de données, présentés sous forme de collections de documents hétérogènes faiblement structurés soulève le problème de la recherche d'information et de la navigation au sein de ces corpus. Dans le cadre de ce vaste projet, il a été proposé une démarche commune d'exploitation et de valorisation des corpus basée sur la recherche d'indices s'adaptant aux différentes représentations de l'information que l'on peut rencontrer dans ces documents patrimoniaux, telles que les zones textuelles, imprimées ou manuscrites, les images, ou encore les illustrations graphiques. Ces signatures sont sensées apporter des connaissances spécifiques et aider la navigation et la recherche d'informations.

Les indices caractéristiques des formes en présence sont extraits de manière automatique en exploitant des méthodes et outils propres à l'analyse d'images, étendus et adaptés aux spécificités des grands corpus. La consultation en mode image des documents patrimoniaux suppose dans un premier temps un archivage qui exige d'examiner de manière approfondie les possibilités spécifiques de compression de ces masses de documents.

I.3 Présentation générale du problème

Le volume des images de documents numérisés pose des problèmes importants en termes de stockage et de transmission sur un réseau limité en débit d'où la nécessité de les compresser. Les formats de compression existants comme JPEG¹, DjVu et DEBORA² (ce dernier a été développé au sein du laboratoire **LIRIS**) ne sont pas efficaces sur les images particulières des documents manuscrits ([4]). L'échec de ces méthodes est lié, d'une part, à la grande complexité des traits présents dans les textes manuscrits et à leur localisation souvent bien difficile, d'autre part, à la difficulté de segmenter les documents manuscrits du fait même de l'absence de redondance de formes d'objets connexes, de leur grande variabilité ainsi que du manque de similarités entre les formes en présence.

Les documents manuscrits présentent une vraie difficulté en terme de segmentation en lignes et en mots. L'étape de la localisation des lignes des textes à l'aide des méthodes existantes devient inopérante vu la non linéarité des mots et des lignes rencontrés ([6], [10], [13], [24], [25], [26], [27], [28], [30], [31]). La première difficulté est avant tout liée à la mise en page de cette classe de documents manuscrits qui n'est pas a priori modélisable et peut même être très complexe, et peu prévisible en présentant tantôt une colonne unique, tantôt plusieurs colonnes de taille et d'interlignage différents au gré des besoins et de l'expression de l'auteur (scripteur). Le second problème relevé concerne la forme du tracé proprement dit, lié plus particulièrement à l'allure générale et à la courbure des lignes de texte. Les

¹ JPEG: Joint Photographic Expert Group.

² DEBORA: Digital accEss to BOoks of the RenAissance. Il s'agit d'un projet européen dont l'objectif est de concevoir un ensemble d'outils permettant l'accès distant et collaborative à des livres numérisés du XVI^{ème} siècle.

irrégularités rencontrées proviennent bien souvent des faibles espaces entre les lignes qui entraînent de nombreux contacts entre les caractères de lignes voisines (*Figure 1*). De même, le faible espace qui existe entre les mots, parfois presque collés, pose un vrai problème de segmentation en mots. Sans rentrer dans le détail des difficultés rencontrées à l'observation de ces corpus, nous pouvons toutefois souligner l'inefficacité des approches usuellement utilisées en segmentation des formes dans le cas de textes imprimés ou manuscrits très réguliers (approches par recherche de connexités binaires, filtrages directionnels strictement horizontaux de type RLSA³, projections en lignes et en colonnes...).

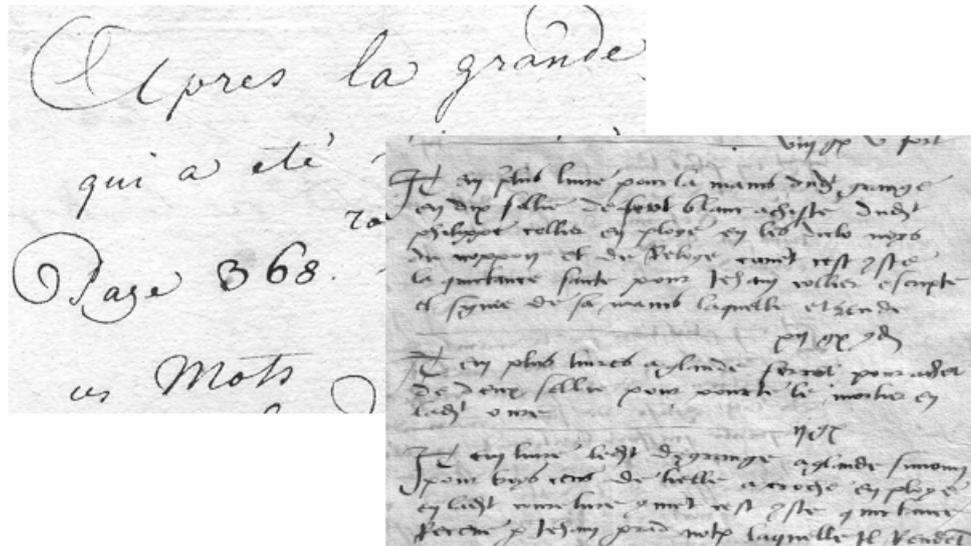


Figure 1 : Extraits de deux ouvrages manuscrits de scripteurs différents présentant de faibles espaces interlignes et une non linéarité apparente des mots

Pour remédier à ces difficultés, il a fallu s'orienter vers une nouvelle méthodologie permettant de localiser précisément les zones de textes, de repérer les formes redondantes sans avoir recours à aucune segmentation de la page et conduisant ainsi à une forme de partitionnement intelligent de l'image centré autour des zones informantes. Rappelons ici que le partitionnement de l'image est à la base de nombreuses méthodes de compression et qu'il peut être réalisé de façon idéale par une recherche guidée par les régions d'intérêt. A ce jour, peu de techniques permettent un tel partitionnement : on peut toutefois citer l'approche de la compression fractale basée sur la recherche d'auto-similarités entre blocs (à partir d'un partitionnement sous forme géométrique, rectangulaire ou triangulaire,[34]) et les techniques de compression DjVu et DEBORA qui procèdent en une séparation par couches d'informations et un partitionnement de zones binaires exclusivement textuelles ([4]). Nous reviendrons dans le détail de ces techniques dans le chapitre 2. Quoi qu'il en soit, les modèles de partitionnement rencontrés ne s'adaptent pas aux tracés manuscrits. Ces modèles de partitionnement ne sont pas adaptés ni guidés par le contenu de l'image. C'est précisément sur ce problème du choix du modèle de partitionnement que se situe notre travail et par conséquent notre apport personnel. En nous inspirant des lacunes des méthodes actuelles de

³ RLSA: Run Length Smoothing Algorithm. Il s'applique aux séquences binaires. Il remplace les pixels blancs par des pixels noirs dans les séquences de pixels blancs de longueur inférieure ou égale à un seuil donné. Cette technique permet de regrouper les connexités rapprochées en une connexité de plus grande taille. Dans le cas de lignes de texte, cet algorithme fournit des blocs noirs et blancs dans l'image du document et de la résolution de numérisation.

partitionnement et de caractérisation des formes présentes sur les images de traits, nous proposons dans cette mémoire une nouvelle approche de localisation et de caractérisation des régions informantes contenues dans les tracés manuscrits et conduisant à un partitionnement intelligent de l'image selon la distribution non aléatoire de l'information sur la page. La clé de notre proposition consiste donc à caractériser, dans un premier temps, les tracés en localisant une version simplifiée des formes que nous nommerons « *Esquisses* » et sur lesquelles nous rechercherons des indices de similarité et de redondance caractéristiques des redondances du tracé d'origine. Notre approche repose sur une représentation robuste des tracés basée sur une recherche de similarités entre formes incluant des hypothèses non restrictives sur les changements d'échelle, les possibilités de rotation et de translation. Pour trouver une similarité entre deux formes semblables (représentée par leurs esquisses), nous avons procédé au repérage des points saillants de contours par le détecteur de Harris et exploité les régions de saillance directionnelle par les bancs de filtres de Gabor paramétrés dans les directions privilégiées des tracés. Pour cela, nous avons exploité la répartition spectrale de l'image à l'aide d'une fonction d'autocorrélation permettant de relever les directions répétées caractéristiques des tracés en présence. Nous avons choisi de développer une approche basée sur une comparaison flexible de similarités partielles pour localiser les différentes configurations redondantes. La recherche de similarités entre les zones d'images que nous avons mise en oeuvre dépend des certaines transformations géométriques affines comme la translation, la rotation et l'homothétie.

I.4 Plan de la mémoire

Cette mémoire s'articule autour de la description et l'exploitation d'une nouvelle approche de recherche de similarités entre formes présentes dans les tracés manuscrits en vue de leur compression. Après avoir présenté dans une introduction générale la problématique de la numérisation du patrimoine et le besoin majeur de stocker et d'archiver ces grandes masses de données sans se soucier réellement de la qualité de la transmission, nous abordons dans deuxième chapitre, à l'aide de plusieurs exemples, l'échec des différentes méthodes de compression existantes appliquées aux images de manuscrits. Avec ces résultats, nous prouvons la nécessité, très urgente, de nous orienter vers des nouvelles méthodologies permettant de pallier aux différents problèmes rencontrés.

Dans le troisième chapitre, nous présentons l'analyse différentielle et fréquentielle des manuscrits étudiés pour détecter les zones d'intérêts dans l'image de document manuscrit. A partir de ces études nous proposons une méthode de localisation des zones textuelles en régions d'intérêt à partir de l'analyse des tracés présentant des similarités visuelles.

Dans le dernier chapitre, nous présentons, en premier lieu, le traitement appliqué aux images issu de l'exploitation des réponses des filtres directionnels de Gabor conduisant à la construction d'un ensemble de cartes directionnelles permettant de localiser et de caractériser les esquisses des formes des tracés. C'est sur ces esquisses que porte ensuite l'estimation de l'ensemble des similarités entre formes. En second lieu, nous présentons la définition du prototype référent appelé « *Esquisse de référence* » estimée pour chaque carte directionnelle et qui sert de base à la définition des indices de similarité entre esquisses. Nous proposons ensuite de relever l'ensemble des liens existant entre les différentes esquisses de références pour définir un « *taux de similarité* » entre formes. L'estimation du taux de similarité entre les différentes esquisses nous permet finalement de localiser précisément les zones de tracés redondantes dans l'image initiale. Ce travail aboutit dans une dernière partie à son intégration à un processus complet de compression.

II. LA COMPRESSION DES DOCUMENTS

Les usagers des bibliothèques numériques désirent de plus en plus de fonctionnalités qui ne peuvent pas encore être offertes en temps réel. Ils cherchent une grande accessibilité aux ouvrages, aux images et à leurs contenus. L'accessibilité traduit une possibilité de consultation en ligne facile et rapide. Mais, le volume de ces images pose des problèmes importants en termes de stockage et de transmission sur un réseau limité en débit, d'où la nécessité de compresser. Les formats traditionnels qui existent ne sont pas efficaces sur les images particulières de documents manuscrits. En effet, ces formats ne tiennent pas compte de la spécificité des tracés manuscrits, de la fréquence des dispositions et des caractéristiques des formes. D'où la nécessité de trouver une autre approche, une méthode de compression adaptée et efficace, qui pourra pallier l'échec des méthodes existantes. Dans ce chapitre, nous présentons quelques exemples qui montrent l'échec des différentes méthodes existantes, comme JPEG, DjVu, la méthode de DEBORA développée au sein du laboratoire LIRIS.

II.1 La compression générique (avec perte) : JPEG

Les méthodes de compression d'images avec perte utilisent les imperfections de l'œil humain pour détruire l'information non perceptible par l'homme. Cette méthode consiste à enlever toutes les hautes et moyennes fréquences d'une image sans modifier la perception de celle-ci. Dans une image de documents, ces hautes et moyennes fréquences sont des informations qui concernent les contours des traits et des caractères.

Le format le plus répandu de compression avec perte est JPEG ([4]). Il offre un gain moyen de compression autour de 1:10 avec des dégradations visuelles très importantes qui nuisent à la lecture et aux traitements automatiques par ordinateur.

Le nouveau standard JPEG2000, basé sur la compression en ondelettes, améliore très peu la compression des images de textes ([4]). L'échec de cette méthode s'explique par la forte présence de traits dans ces images dont la complexité peut être difficilement reproduite par un codage générique. JPEG n'est pas adapté à la complexité des traits des caractères et à leurs localisations aléatoires dans une image de texte.

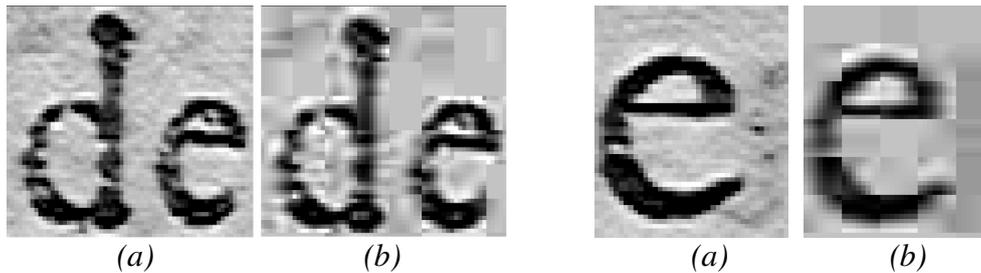


Figure 2 : Image originale (a)

Image Comprimée JPEG à 90% qualité (b)

La Figure 2 représente une image comprimée avec JPEG. Nous remarquons une délocalisation du niveau de gris des pixels, une réduction des nuances de gris due au filtrage fréquentiel passe-bas et une apparition de niveaux de gris qui n'existaient pas auparavant due au découpage en blocs carrés propres à l'algorithme de compression JPEG. De même, une destruction des traits dont la taille est inférieure à quelques pixels et qui ne sont pas restitués dans l'image comprimée ce qui implique une dégradation des performances des logiciels de reconnaissance des caractères (*OCR*⁴).

Nous constatons aussi que la compression JPEG modifie la localisation des contours des caractères qui constitue une information très importante pour la segmentation des caractères et leur interprétation. Leurs déformations peuvent conduire à des ruptures de tracés ou des collages entre caractères consécutifs.

D'autre part, il n'existe pas de filtre inverse qui corrige les effets de la compression de JPEG ([4]). Même si certains artifices comme le filtrage des contours permettrait de les améliorer, nous ne pourrions jamais retrouver l'information perdue d'origine. Ainsi les fonds documentaires déjà numérisés et stockés sous ce format resteront difficilement interprétables par ordinateur. Nous pouvons poser le constat qu'il n'est pas possible de compresser ces images de façon efficace avec des taux de compression intéressants sans perte d'information, c'est la raison pour laquelle il faut donc trouver le moyen de perdre l'information la moins utile possible pour la restitution et l'interprétation par ordinateur.

La solution passe par des moyens de compressions adaptés aux images des documents et à la conservation des informations essentielles pour la lecture et le traitement informatique.

II.2 La compression adaptée avec minimum de pertes d'informations utiles : DjVu

L'idée consiste à séparer dans l'image, les traits du support papier. En appliquant une compression sans perte sur l'image de traits et une compression avec perte sur l'image du support papier, nous conservons à la fois l'information pour la lecture et nous augmentons simultanément la possibilité de compresser l'arrière-plan avec des facteurs de compression beaucoup plus importants sans risquer d'affecter le rendu de l'image. Cette forme adaptée de compression existe avec DjVu d'AT&T (<http://djvu.research.att.com/>, [4]).

La compression DjVu effectue tout d'abord une séparation entre la forme et le fond à l'aide d'une binarisation adaptative. Elle autorise un taux élevé de compression (entre 1:50 et 1:200).

⁴ OCR : Optical Character Recognition

Image Originale	JPEG, 300dpi,	JPEG, 100dpi,	Ondelettes 300dpi,	Compression forme/fond séparé
 20640K	 292K 70:1	 50K 412:1	 61K 338:1	 52K 396:1
 9534K	 116K 82:1	 17K 560:1	 20K 476:1	 19K 501:1
 22013K	 383K 57:1	 41K 536:1	 43K 511:1	 38K 579:1
 12990K	 250K 51:1	 38K 341:1	 42K 309:1	 40K 324:1
 12128K	 206K 58:1	 35K 346:1	 39K 310:1	 37K 327:1
 31059K	 388K 80:1	 77K 403:1	 78K 398:1	 73K 425:1

Figure 3 : Différences visuelles entre les modes de compression (AT&T)

La compression DjVu présente quelques inconvénients. En premier lieu, la séparation entre le fond et la forme s'effectue par traitement d'image sans aucune interprétation. L'appariement des formes se fait indistinctement sur des caractères comme sur des éléments graphiques ce qui rend difficile toute exploitation ultérieure de ces informations pour l'indexation et la construction automatique de liens. Les caractères qui n'ont pas été isolés sont détruits par la compression par ondelettes et inversement les éléments graphiques qui se retrouvent segmentés dans l'avant-plan, surchargeant la compression. Il est donc important de s'assurer que tous les caractères, qui doivent être préservés, soient correctement isolés ce que DjVu ne garantit pas. Enfin DjVu n'utilise pas la redondance des formes sur tout un livre et effectue la compression page par page de façon indépendante ce qui limite considérablement le nombre de formes redondantes caractérisables.

II.3 La compression dans le cadre du projet DEBORA

La méthode DEBORA est adaptée aux images de documents imprimés et repose sur la très forte redondance des caractères par rapport aux éléments graphiques de formes aléatoires ([4]). Comme pour les méthodes précédentes, on sépare l'arrière plan représentant l'image de support pour le compresser avec pertes. Le principe consiste à diviser l'image de document en trois plans dont chacun est compressé différemment avec un modèle de compression adapté.

- le plan textuel (image binaire) : il comporte les images de caractères qui ont la propriété d'être fortement redondants sur l'ensemble du livre. Les plans textuels peuvent être compressés avec un taux extrêmement élevé de compression en utilisant l'appariement des formes.
- le plan graphique (image binaire) : il contient les éléments graphiques qui ne sont pas des caractères et qui se caractérisent par une très faible redondance de forme. Il n'est pas utile de compresser ces éléments vu la faible redondance de formes dans le livre.
- l'arrière plan (image couleur) : il représente comme pour DjVu l'image du support papier privé des caractères et des éléments graphiques. Cette image couleur est fortement compressée avec perte en utilisant les codeurs existants comme JPEG.

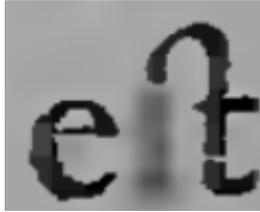
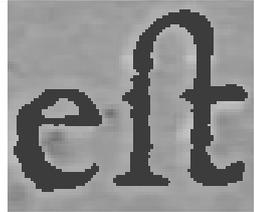
			
JPEG Q=60% 10: 1	DjVu (lossy) 120:1	DjVu (Sans perte) 20:1	DEBORA 60:1

Figure 4 : Comparaison entre les taux de compression sur des documents anciens

La Figure 4 montre les différences visuelles résultant de l'utilisation de différents algorithmes avec leurs taux de compression respectifs. Les résultats visuels de la compression de DEBORA/RFV sont comparables avec ceux de la version sans perte de la compression DjVu. Le taux moyen de compression sans perte de DEBORA/RFV est très supérieur à la compression sans perte de DjVu mais il reste inférieur aux taux de compression avec perte de DjVu en version «**aggressive**». Cependant, la compression avec perte de DjVu introduit de nombreuses erreurs qui peuvent ne pas satisfaire l'utilisateur. Pour conserver une qualité d'image suffisamment élevée pour le travail minutieux des historiens, les méthodes de compression modernes comme DjVu ou DEBORA/RFV compressent en moyenne 5 à 10 fois mieux que le standard JPEG à qualité visuelle constante.

La compression selon DEBORA utilise la méthode de compression par séparation de couches. Or, les documents manuscrits présentent une réelle difficulté de séparer les différentes couches d'informations et d'appliquer une approche de compression par redondance des formes (exploitation du pattern matching et de la substitution des formes) du fait de la grande variabilité et de la complexité des formes ainsi du manque de similarités existant entre les formes connexes en présence. Ces limitations nous permettent de conclure que la méthode de compression DEBORA n'est pas adaptée à la compression des documents manuscrits numérisés et qu'il faut envisager une approche plus robuste à la grande variabilité des formes présentes dans le tracé.

II.4 Recherche d'auto-similarités - la compression fractale

La compression fractale a pour principe qu'une image n'est qu'un ensemble de motifs identiques en nombre limité, sur lesquels on applique des transformations géométriques affines de types translations, rotations, symétries, agrandissements, ou encore réductions. Evidemment, plus l'image possède de motifs vérifiant ces propriétés, meilleur en est le résultat ([8], [9], [32], [33], [34], [35], [36]).

Comme pour le format JPEG, l'image est découpée en blocs de pixels, mais ils sont ici de tailles variables. Il faut ensuite détecter les redondances entre ces blocs à diverses résolutions. On parle alors de transformations fractales basées sur un opérateur contractant. Ces transformations décrivent l'image de plus en plus finement. A la fin de ce processus, on ne stocke pas le contenu d'un bloc autant de fois qu'il a été "vu" dans l'image mais seulement les équations mathématiques permettant de représenter le contenu de ces carrés ([32], [33], [35], [36]). Au final, on obtient une structure présentant des caractéristiques similaires à des échelles différentes. Pour retrouver l'image il suffit alors de décrire les transformations qui ont été appliquées aux blocs initiaux. Ce processus rend la compression **indépendante** de la taille de l'image. De plus, l'image produite est **vectorisée** et ne subit pas les effets de la pixellisation, contrairement au JPEG. Ce phénomène est surtout visible lors d'un zoom par

exemple, l'image fractale peut devenir floue mais ne pixellise pas (*Figure 5*). Ceci est dû au fait que lors de l'agrandissement, ce ne sont pas les pixels qui sont élargis, mais toute l'image qui est recalculée mathématiquement. Pour la compression fractale, on trouve plusieurs types de partitionnement, les plus connus sont en Quadtree (*Figure 6*) et en rectangle et Delaunay (triangles). La plus efficace de ces méthodes dans le rapport qualité/compression est la méthode par partitionnement Delaunay.

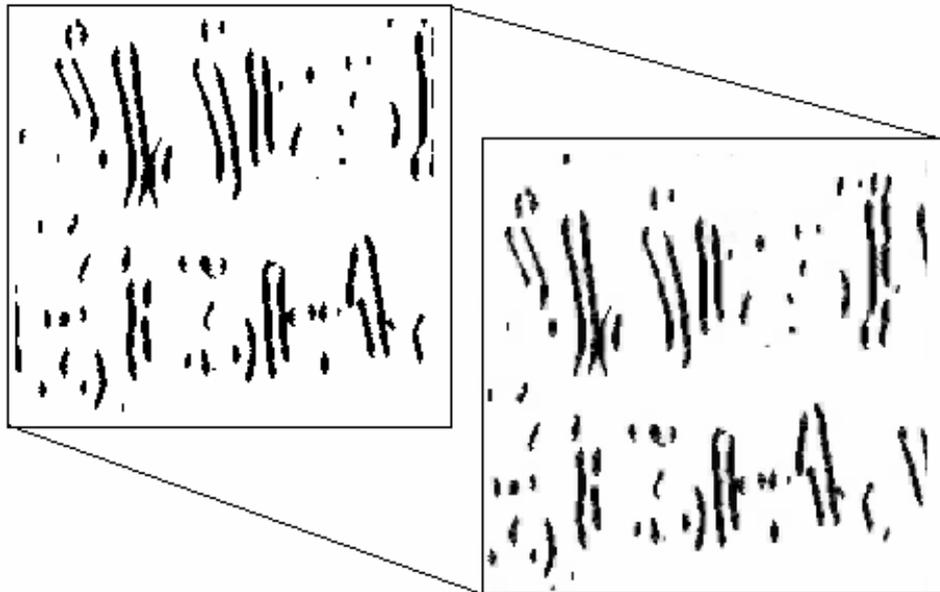


Figure 5 : (a) image originale, (b) image décompressée par la méthode de la compression fractale

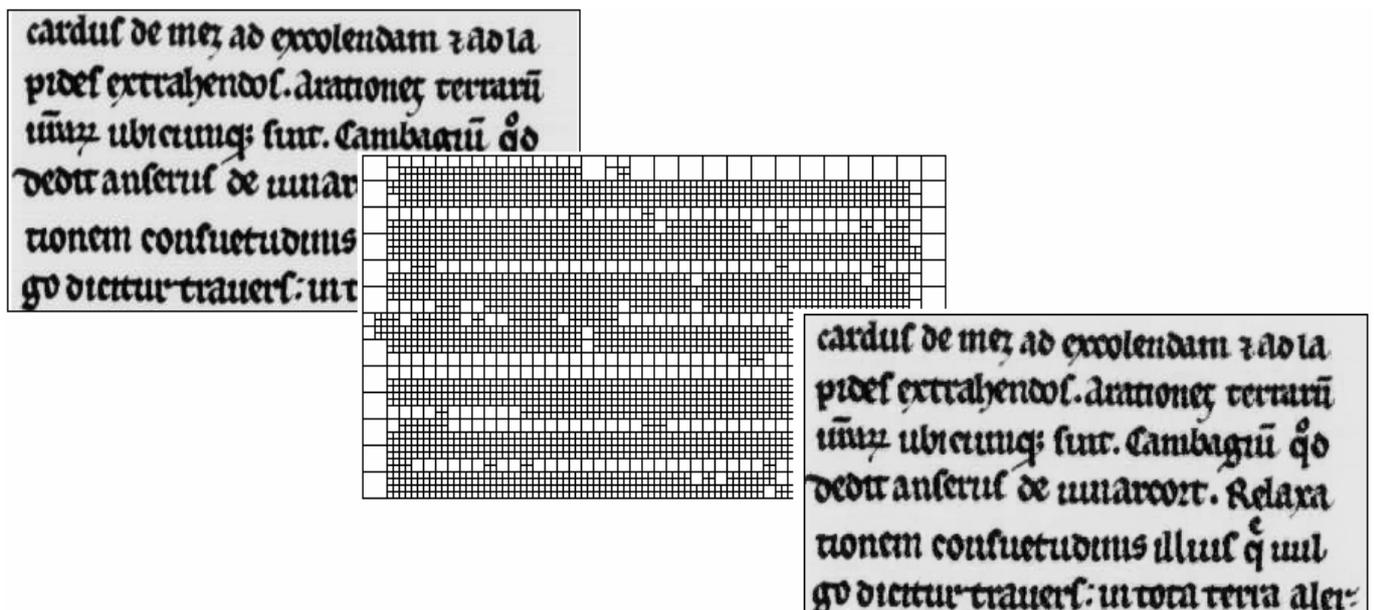


Figure 6 : (a) image originale, (b) partitionnement carré de la compression fractale, (c) image décompressée selon la méthode fractale

La méthode de compression fractale appliquée aux images de textes (textes imprimés principalement) permet de trouver dans les imagerie (blocs de pixels qui partitionnent l'image et que nous appellerons *Domaines*) les éléments qui peuvent être retrouvés à des échelles différentes dans l'image. Avec l'ensemble des Domaines extraits et utilisés dans la

phase initiale de compression, on peut reconstruire une base de référence des traits pour l'alphabet utilisé sur ces images de textes. Cette base de référence représente les similarités internes contenues dans l'écriture ([7]). Il faut noter que la forme géométrique des imageries de domaines dépend du type de partitionnement utilisé.

La complexité des écritures de documents manuscrits et l'absence de similarité interne dans les écritures montrent l'échec de cette méthode et le mauvais taux de compression (1 :16) obtenu sur différentes images.

II.5 Vers une méthodologie sans segmentation et adaptée aux caractéristiques des tracés

Tous les formats de compression, cités précédemment, sont performants sur des documents imprimés, mais les taux de compression se dégradent rapidement sur des documents dont le contenu est trop complexe à analyser et les formes ne sont pas suffisamment redondantes (documents manuscrits). Pour pouvoir comprimer de façon intelligente et efficace, il faut savoir segmenter correctement les images et reconnaître leur contenu.

Les images de documents manuscrits du corpus de notre étude sont fortement bruitées et contiennent des formes qui ne sont pas segmentables par les approches classiques de traitement d'images (approches procédant en général par une binarisation des tracés et une extraction de connexités). La difficulté de la segmentation et l'absence de redondance de forme d'objets connexes rendent impossible de comprimer ces images par les méthodes existantes. Pour remédier à cet obstacle majeur, il faut passer à d'autres approches exploitant cette fois des similarités de forme à des niveaux plus fins.

Le tracé manuscrit est par définition très complexe à modéliser, ceci est dû à la non régularité des formes, et leur importante variabilité que l'on peut constater d'une écriture à l'autre. Chez un même scripteur d'ailleurs on retrouve une hétérogénéité dans la forme des tracés, parfois même sur une même page d'un seul auteur. Tenant compte de cette forte variabilité propre aux tracés manuscrits, nous avons choisi de nous orienter vers une nouvelle méthodologie permettant de trouver localement des similarités partielles sans avoir recours à une segmentation préalable des formes. En nous inspirant de la compression fractale par IFS, et sans séparer les couches d'informations ni segmenter le tracé, nous avons choisi de localiser les différentes zones d'intérêt réparties dans les images de traits. Ces zones d'intérêts vont nous permettre de définir les formes redondantes internes d'une écriture que nous appellerons esquisses de forme dans ce mémoire. Ces formes redondantes ne sont pas nécessairement des mots ni des lettres : elles peuvent être simplement constituées des petites boucles des lettres (les graphèmes) qui se retrouvent localisés à d'autres endroits dans le tracé manuscrit et ceci à une transformation géométrique près (un changement d'échelle, une rotation ou bien une translation). L'échec de la compression fractale à ce stade est lié au mauvais partitionnement initial de l'image. Pour remédier à ce problème, il a fallu trouver un moyen intelligent et efficace de guider le partitionnement et de localiser les zones d'intérêts contenant les informations redondantes du tracé. La localisation de ces zones est un étape primordiale à la mise en place d'une méthode de compression efficace et adaptée aux contenus des images.

III. NOTRE APPROCHE : LOCALISATION **DE ZONES D'INTERET AU NIVEAU DU** **GRAPHEME**

Les images de documents écrits ne sont pas des images naturelles. Elles sont essentiellement constituées d'un ensemble de traits distribués non aléatoirement à la surface d'un support papier généralement homogène. Les méthodes standard de traitement d'images ne sont pas nécessairement adaptées à la localisation et à la caractérisation de ces formes particulières. Il a donc fallu imaginer une « imagerie » spécifique pour traiter les documents, plus spécifiquement encore les documents manuscrits d'ouvrages anciens d'origines historiques très diversifiées sur lesquels nous travaillons. A ce titre, rappelons que les techniques de reconnaissance des caractères actuellement en usage, autour des approches de type OCR (Optical Character Recognition) sont bien adaptées aux polices de caractères modernes et au style d'écriture latine. En revanche, ces logiciels ne sont pas capables de reconnaître les caractères imprimés des siècles passés, moins normalisés et plus bruités. Par conséquent les manuscrits (même très réguliers) ainsi que les diverses abréviations pouvant exister sur des pages de textes manuscrites ne peuvent à ce jour être reconnus que sur des polices très régulières et connues par le système. Mais avant même de procéder à une reconnaissance complète - lorsqu'elle est envisageable - sur cette classe de documents très complexes, il faut pouvoir localiser et caractériser les formes contenues dans les parties de lettres pour pouvoir ensuite éventuellement les identifier. Cette tâche est devenue très difficile dans ces environnements très bruités et impossible à modéliser. A ce titre, nous avons montré dans les chapitres précédents la grande difficulté à segmenter des images de tracés non binarisables avec à l'appui des exemples montrant l'échec de différentes méthodes de compression appliquées à des images de tracés manuscrits. Avec ces résultats, nous avons prouvé la nécessité, très urgente, de nous orienter vers de nouvelles méthodologies permettant de pallier aux différents problèmes rencontrés.

Dans ce chapitre, nous présentons une contribution à la recherche des formes contenues dans les tracés manuscrits de corpus anciens (documents d'écritures arabes, latines ou gothiques allant de l'époque médiévale aux siècles contemporains) sans avoir recours à une segmentation binaire des traits. Notre approche est basée sur une analyse différentielle et

fréquentielle des manuscrits permettant de détecter les zones d'intérêts des images. La première approche, l'analyse différentielle, repose sur la localisation des régions d'intérêt à partir d'une sélection de points guidée par la présence de fortes courbures dans le tracé et caractéristiques de l'existence à ces endroits de similarités visuelles. Il existe des outils de sélection de zones et de points d'intérêts comme le détecteur de Harris que nous allons présenter dans la suite de ce chapitre. Nous avons pu constater à l'observation de ces points que leur distribution est peu stable autour de configurations quasi identiques (contenues dans deux lettres identiques, deux mêmes graphèmes, ou encore deux mêmes portions de traits réparties sur l'image). Cette observation nous a conduit à proposer une autre approche plus stable et plus robuste aux changements de courbure et d'intensité afin de détecter des similarités visuellement saillantes autour de ces zones et pour une même écriture. La seconde approche étudiée est basée sur une analyse fréquentielle impliquant des bancs de filtres de Gabor. Le principe consiste à décomposer l'image originale en cartes directionnelles, utilisant les filtres de Gabor avec un paramétrage adéquat, selon les directions les plus significatives de l'image. Ces directions sont évaluées sur le spectre à partir de l'exploitation d'une fonction d'autocorrélation permettant de construire une rose des directions. La distribution en cartes directionnelles des formes extraites du tracé d'origine permet de constituer des groupements de formes (appelées esquisses) selon des critères de ressemblances dans la direction θ et selon des transformations géométriques affines (translations, rotations, changements d'échelle).

III.1 Première approche : Analyse différentielle des manuscrits

Les tracés manuscrits d'une image peuvent être considérés comme similaires si les fonctions les décrivant sont proches, au moins en certains points. Pour trouver cette similarité entre deux formes semblables, nous avons repéré les points saillants de contours par le détecteur de points d'intérêt Harris. Cette approche est basée sur l'extraction de caractéristiques en certains points saillants de l'image privilégiant les configurations locales jugées intéressantes. En traitant les images en niveaux de gris et sans passer par une recherche systématique de composantes connexes binaires, nous avons implémenté le détecteur de points d'intérêt de Harris localisant précisément les maxima locaux et minima locaux, les pics et les vallées présents sur les hautes fréquences des tracés, ([3], [19], [20], [21], [22]), voir *Figure 7*.

Aucune binarisation, ni recherche systématique de composantes connexes, n'est nécessaire pour extraire ces points saillants. Le détecteur de Harris est appliqué directement aux images en niveaux de gris ou aux images couleurs. Cette méthode différentielle est fondée sur l'analyse de la variation de la luminance au voisinage d'un point : il s'agit en d'autres termes d'une détection de points ayant des configurations locales à fortes variations différentielles que l'on peut distinguer selon leurs propriétés (points pics, points vallées, points marquant une forte concavité ou convexité...).

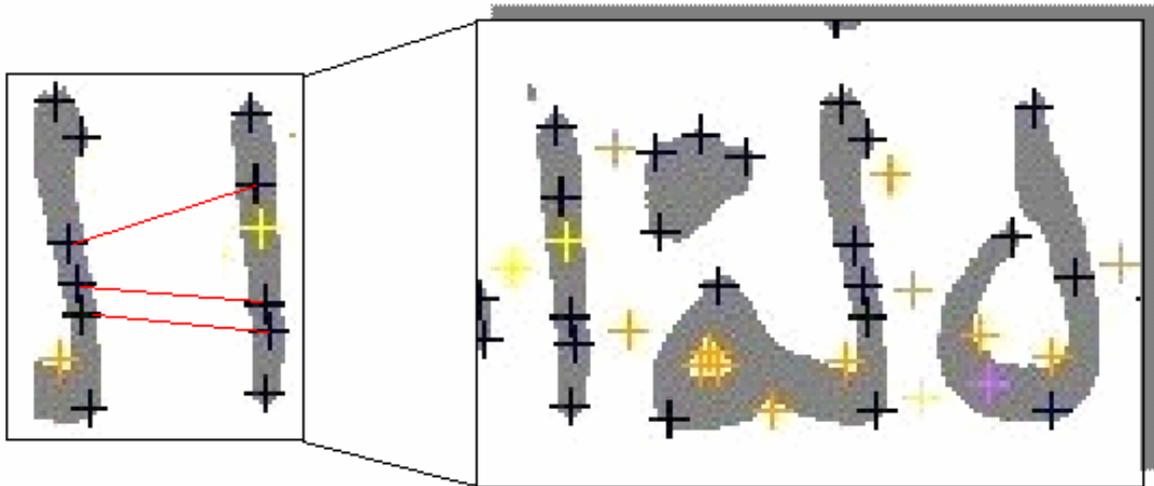
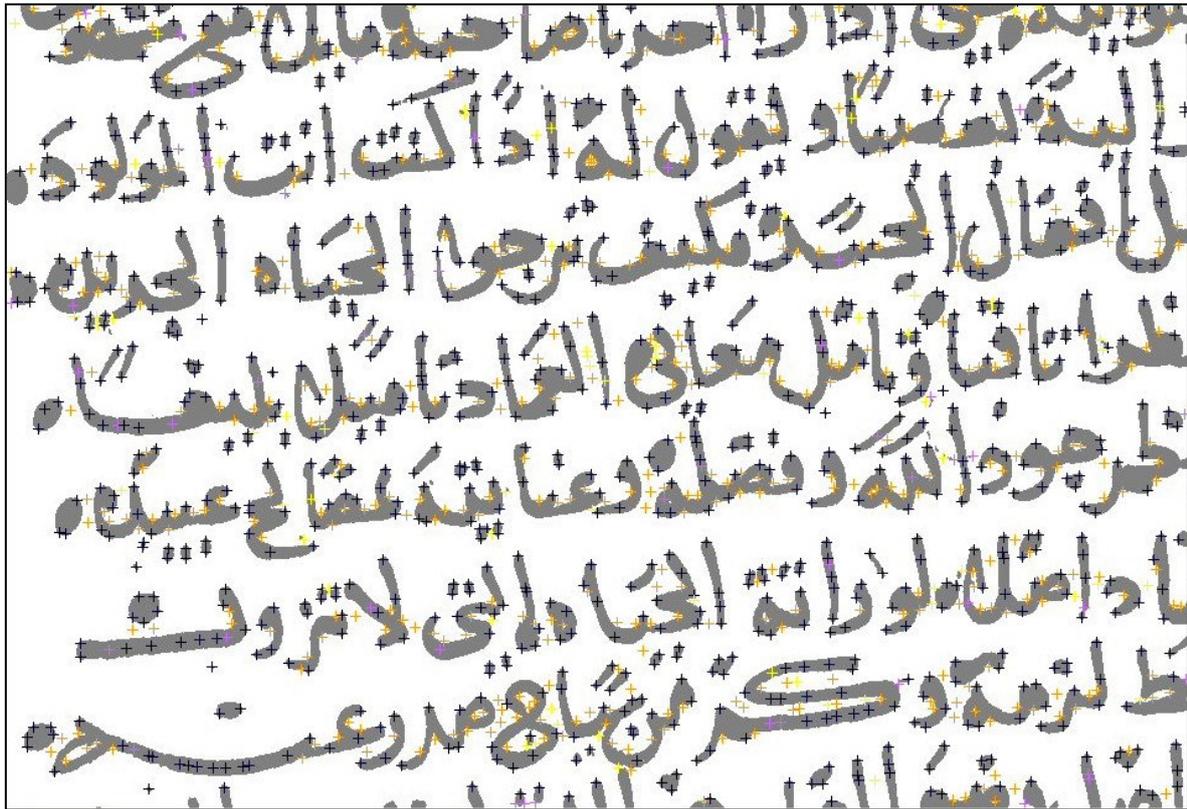


Figure 7 : Localisation des points d'intérêt sur une image de manuscrit arabe à l'aide du détecteur de Harris sur des images en niveaux de gris.

Le détail de cette approche est présentée dans [3].Voici les principales caractéristiques utilisées pour distinguer les points entre eux et les catégoriser. Sous forme matricielle, le détecteur de Harris se traduit par l'équation suivante :

$$E(x, y) = A.x^2 + 2.C.x.y + B.y^2$$

$$\text{Avec } A = \left(\frac{\partial I}{\partial x} \right)^2 * w \quad B = \left(\frac{\partial I}{\partial y} \right)^2 * w \quad C = \left(\frac{\partial I}{\partial x} \cdot \frac{\partial I}{\partial y} \right) * w$$

Le symbole * désigne l'opérateur convolution. La mesure dépend donc de trois paramètres que l'on peut résumer dans la matrice locale M définie par :

$$M(u, v) = \begin{pmatrix} A & C \\ C & B \end{pmatrix}$$

La matrice M traduit le comportement local de la fonction image. Afin de déterminer si le point (x,y) est caractéristique, Harris a proposé une mesure scalaire :

$$R = Det(M) - k.Trace(M)^2 = (A.B - C^2) - k.(A + B)^2 = (\lambda_1.\lambda_2) - k.(\lambda_1 + \lambda_2)^2$$

Avec λ_1 et λ_2 sont les valeurs propres de M. $k = 0.04$ influence le nombre de points d'intérêt. La caractérisation des points repose sur l'analyse des minima et maxima locaux.

Un maximum local	$\ \nabla f\ = 0, \lambda_1 < 0, \lambda_2 < 0$	Un minimum local	$\ \nabla f\ = 0, \lambda_1 > 0, \lambda_2 > 0$
Un col	$\ \nabla f\ = 0, \lambda_1.\lambda_2 < 0$	Un point plat	$\ \nabla f\ = 0, \lambda_1 = 0, \lambda_2 = 0$
Un minimum local (le ravin)	$\ \nabla f\ \neq 0, \lambda_1 > 0, \nabla f.w_1 = 0$ $\ \nabla f\ \neq 0, \lambda_2 > 0, \nabla f.w_2 = 0$ $\ \nabla f\ = 0, \lambda_1 > 0, \lambda_2 = 0$	Un maximum local (une crête)	$\ \nabla f\ \neq 0, \lambda_1 < 0, \nabla f.w_1 = 0$ $\ \nabla f\ \neq 0, \lambda_2 < 0, \nabla f.w_2 = 0$ $\ \nabla f\ = 0, \lambda_1 < 0, \lambda_2 = 0$

Tableau 1 : caractérisation locale des points

Avec ∇f le vecteur gradient pour le pixel (x,y), et $\|\nabla f\|$ son amplitude.

Le but de l'utilisation de cette approche consiste donc à regrouper les points représentant les mêmes caractéristiques locales, puis à définir, à partir de ces points, des zones d'intérêt (plus étendues) et similaires permettant de trouver une possible répartition des formes de l'image en zones d'intérêt. Cependant, en comparant la distribution des points d'intérêt pour une même lettre (Figure 7), nous avons remarqué une grande variabilité de configurations possibles autour de mêmes zones. Aussi, tenter d'améliorer de tels résultats revenait-il alors à chercher à faire se ressembler des formes sémantiquement identiques (deux même lettres par exemple, ou deux même graphèmes) mais présentant des configurations locales visiblement différentes. Ceci est naturellement lié à la grande variabilité des formes présentes dans un même tracée, qu'une approche typiquement locale ne peut pas contourner. Nous avons donc exploré une approche plus stable et plus robuste pour localiser les zones de forte similarité. Cette seconde approche consiste à orienter les traitements vers une analyse fréquentielle de l'image en exploitant des bancs de filtres de Gabor.

III.2 Notre seconde approche : Analyse fréquentielle des manuscrits et filtrage des données selon les directions émergentes de l'image

La première étape du processus de recherche de similarités de formes au sein d'un même tracé que nous proposons repose sur la décomposition de l'image en plusieurs cartes directionnelles contenant des zones semblables présentes dans une même écriture. En premier

lieu, il nous a fallu exploiter les directions les plus significatives du tracé pour ne pas tenir compte d'un grand ensemble de petites orientations naturellement présentes sur le spectre de l'image (arrière plan de la page par exemple) et ne présentant aucun lien avec le tracé lui-même. Pour cela, nous avons utilisé une approche de détection automatique des directions les plus représentées dans l'image du tracé basée sur la construction d'une « *Rose des Directions* » estimée sur toute la surface initiale de l'image ([5], [6], [11]). Cette méthode permet de représenter dans un diagramme en coordonnées polaires les amplitudes relatives de chacune des orientations présentes dans cette image. Nous allons nous intéresser aux directions les plus significatives, celles pour lesquelles l'amplitude est maximale. Dans ce qui suit, nous présentons l'aspect technique de cette méthode.

III.2.1 Passage par la fonction d'autocorrélation

La construction de la rose des directions repose sur l'exploitation de la fonction d'autocorrélation appliquée à l'image. Cette fonction a souvent été utilisée pour la caractérisation de texture ([13], [14], [15]). En particulier, elle permet de mettre en évidence des périodicités dans une image et traduit de ce fait les orientations principales de certaines textures, [12].

La définition générale de la fonction d'autocorrélation pour un signal bidimensionnel est la suivante, [16].

$$C_{xx}(k, l) = \sum_{k'=-\infty}^{+\infty} \sum_{l'=-\infty}^{+\infty} x(k', l') \cdot x(k'+k, l'+l)$$

Le calcul pratique de cette fonction se fait par l'intermédiaire de la transformée de Fourier. Il faut donc s'intéresser de près au domaine fréquentielle de l'image et à la répartition des hautes fréquences présentes aux frontières des tracés manuscrits.

D'un point de vue pratique, la fonction d'autocorrélation regroupe l'ensemble des valeurs que l'on peut obtenir en faisant la somme de tous les produits des niveaux de gris des points en correspondance après translation de l'image par rapport à elle – même. Ainsi un point C(i,j) de la fonction d'autocorrélation contient la valeur de la somme des produits des niveaux de gris des points en correspondance après une translation de vecteur (i,j). Ces différentes translations permettent d'inspecter l'image selon ses différentes directions. Toutes les translations selon des vecteurs colinéaires donnent des indications selon la direction correspondante. Sur la fonction d'autocorrélation, ces données relatives à une même direction seront situées sur une même droite, ayant aussi cette direction, et passant par l'origine.

III.2.2 Principe de la détection des orientations

Le principe consiste donc à translater les différents points d'un objet dans toutes les directions. Par exemple la translation d'une droite dans sa propre direction va conduire à un fort niveau de correspondance qui se traduit par une valeur importante de la fonction d'autocorrélation dans la direction de la droite. Le même calcul dans la direction orthogonale à la droite conduit à une très faible superposition, et donc par de très faibles valeurs de la fonction d'autocorrélation ([11]). Ceci peut enfin se généraliser à l'ensemble des objets présentant un allongement dans une direction donnée : ce qui est le cas des lignes de texte dans la direction horizontale. Plus particulièrement, la fonction d'autocorrélation est très sensible à la disposition relative des objets les uns par rapport aux autres. Ainsi, les caractères discrets d'une même ligne de texte s'orientent globalement selon les caractères discrets d'une même ligne de texte s'orientent globalement selon la direction horizontale. L'autocorrélation va ainsi mettre en évidence le recouvrement des formes par translation ([5], [6]).

III.2.3 La Rose de Directions : aspects techniques

Le calcul de rose des directions de l'image a été proposé par Bres dans [11]. Il est basé sur le calcul de l'espérance mathématique déterminée à partir de la fonction d'autocorrélation précédente. Soit I une image et (x,y) les coordonnées d'un point de cette image. Supposons que l'orientation préférentielle de l'image soit dans la direction θ . L'espérance mathématique est alors définie par :

$$E_{\theta} = \{I(x,y).I(x+a,y+b)\} \quad \text{pour} \quad \text{Arctg}\left(\frac{b}{a}\right) = \theta$$

E_{θ} représente l'espérance mathématique du produit $I(x,y).I(x+a,y+b)$, calculée pour l'ensemble de points (x,y) et des valeurs a et b qui vérifient $\text{Arctg}\left(\frac{b}{a}\right) = \theta$.

Ainsi, un point $C(a,b)$ de la fonction d'autocorrélation contient la valeur de la somme des produits des niveaux de gris des points en correspondance après une translation de vecteur (a,b) . La fonction d'autocorrélation de l'image donne des valeurs proportionnelles à cette espérance mathématique.

La rose de directions proposée dans [11] représente pour une direction θ_i donnée, la somme $R(\theta_i)$ des différentes valeurs $C_{II}(i,j)$ (défini en 1) selon les droites de direction θ_i . Ainsi la rose de directions correspond au diagramme polaire sur lequel on reporte pour chaque direction θ_i , la somme $R(\theta_i)$ calculée pour chaque droite D_i d'orientation θ_i , passant par l'origine de la fonction d'autocorrélation. Les points de coordonnées (a,b) appartiennent à cette droite. Ainsi, on a :

$$R(\theta_i) = \sum_{D_i} C_{II}(a,b)$$

A partir de ces mesures, on cherche à ne conserver que les variations relatives des contributions de chaque direction, on utilisera désormais les nouvelles valeurs de $R'(\theta_i)$ suivantes :

$$R'(\theta_i) = \frac{R(\theta_i) - R_{\min}}{R_{\max} - R_{\min}}$$

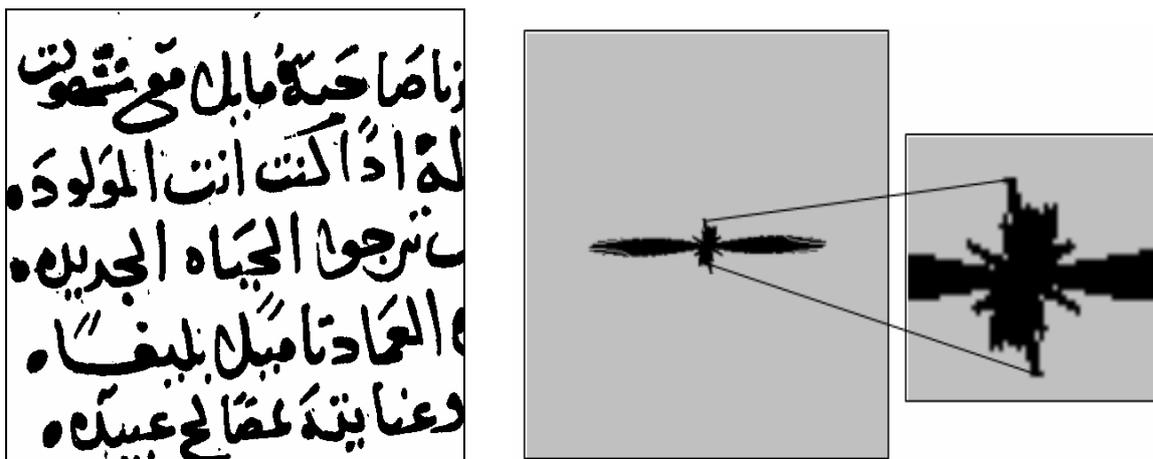


Figure 8 : Rose de direction de la sélection entière et zoom sur les directions principales représentées par des pics locaux.

III.3 Quantification et analyse des esquisses directionnelles

L'approche que nous avons choisie est liée à une décomposition fréquentielle des images de traits et à l'application de filtres directionnels issus des fonctions 2D de Gabor appliqués au domaine spectral de l'image dont les orientations et les bandes passantes fréquentielles sont paramétrables ([5], [6], [12], [17], [18]). La décomposition fréquentielle se fait selon les orientations les plus représentatives détectées par la rose de direction. Les réponses de Gabor dans ces directions initialement détectées à partir de l'exploitation de la rose des directions nous permettent de construire des cartes directionnelles sur lesquelles sont représentées des esquisses des formes du tracé initial. La représentation en esquisses pour chacune des directions significatives θ permet une exploitation simplifiée des formes. Elle est basée sur la mise en place d'indices de similarités favorisant les regroupements entre formes présentant des mêmes caractéristiques. Le chapitre 4 développe la méthode choisie.

III.3.1 Extraction automatique des directions principales

La rose des directions estimée sur une image de traits contient des pics représentant les directions les plus représentées sur la surface complète de l'extrait analysé ([5], [11]). Pour extraire ces valeurs sans tenir compte des pics secondaires sans intérêt, nous avons procédé à une recherche de maxima reposant sur un lissage préalable des valeurs des différentes amplitudes afin d'atténuer les petits pics secondaires. Nous avons choisi de ne conserver que les 6 directions les plus représentées variant entre 0° et 180° . De deux directions proches (dont la différence est inférieure à 10 degrés), on ne conserve que celle dont l'amplitude est maximale. Dans le cas contraire, les filtres de Gabor donnent des réponses trop voisines et donc redondantes. Afin de juger de la pertinence des sélections faites sur la rose des directions, on procède systématiquement à l'analyse du regroupement des réponses binarisées des filtres de Gabor comme cela est présenté à la *Figure 9*.

III.3.2 Le filtre de Gabor : un outil inspiré de la perception visuelle humaine

Le filtrage de Gabor constitue un excellent compromis espace - fréquence pour la représentation du contenu des images. Son paramétrage sur les images de traits que nous analysons ne constitue pas un obstacle à son utilisation ([5], [18]). C'est en règle générale cette étape de paramétrage qui est laborieuse et rarement automatisable. Connaissant grossièrement la répartition fréquentielle des tracés manuscrits (dans les hautes fréquences) par rapport à l'arrière plan (dont la moyenne figure dans les basses fréquences) d'une page numérisée, nous avons réalisé un pavage du plan fréquentiel permettant de paramétrer en fréquences, en échelles et en directions les filtres de Gabor. En privilégiant ainsi les hautes fréquences présentes sur les contours des formes et en filtrant plus massivement les données de fréquences globalement plus faibles représentatives du fond des pages, nous parvenons à un découpage de l'image en plans directionnels multi-échelés sur lesquels repose la recherche des similarités. Ainsi, l'analyse directionnelle que nous avons mise en œuvre se base sur des filtrages successifs de l'image originale faisant varier l'échelle d'analyse et l'orientation : nous procédons ainsi par bancs de filtres successifs en partitionnant le domaine fréquentiel en différents canaux, chacun étant caractérisé par des secteurs angulaires correspondant aux directions principales des tracés.

III.3.3 Paramétrage des filtres de Gabor sur des images de manuscrits en niveaux de gris

Le filtre de Gabor permet de filtrer les fréquences orientées dans une direction et à une fréquence donnée ([17], [18]). Il élimine les formes graphiques et met en valeur les lignes de textes. Ce filtre n'est qu'une fréquence pure modulée par une gaussienne, c'est-à-dire, un filtre passe bande avec une enveloppe gaussienne. Ce filtre est très répandu du fait de sa propriété de résolution optimale conjointe en fréquence et en temps. La fonction du filtre de Gabor est donnée par :

$$G(u, v) = A \left(\exp \left\{ -\frac{1}{2} \left[\frac{(u - u_0)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right] \right\} + \exp \left\{ -\frac{1}{2} \left[\frac{(u + u_0)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right] \right\} \right)$$

Où $\sigma_u = 1/2\pi\sigma_x$, $\sigma_v = 1/2\pi\sigma_y$ et $A = 2\pi\sigma_x\sigma_y$ où σ_x et σ_y sont les déviations standards le long des axes x et y, u_0 est la longueur d'onde sinusoïdale le long de la directions x (à 0° d'orientation).

La *Figure 9* illustre la réponse fréquentielle d'un banc de filtres de Gabor à une impulsion sinusoïdale. Le pavage circulaire de l'espace des fréquences permet une sélection en fréquences croissantes (selon les cercles de diamètres croissants) et en orientation balayant ainsi les directions allant de 0° à 180° .

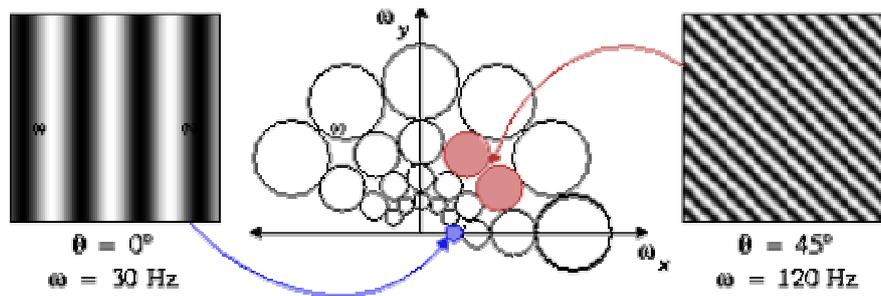


Figure 9 : Spectre du filtre de Gabor traduisant l'amplitude des réponses des filtres $\theta = 0^\circ$ et $\theta = 45^\circ$ avec respectivement une sélection en fréquences $\omega = 30$ Hz et $\omega = 120$ Hz.

Dans notre étude, nous nous sommes intéressés à toutes les directions comprises entre 0° et 180° présentant un pic sur la rose des directions. Le niveau d'échelle retenu permet un filtrage globalement modéré évitant les effets d'empiètement dans le cas d'un filtrage trop important. Les valeurs choisies pour de u_0 permettent de régler ce niveau d'échelle. Pour illustrer des résultats de filtrages à différents niveaux d'échelles et de directions voici sur le

Tableau 2 un exemple de ce qui peut être obtenu sur un tracé manuscrit extrait d'un ouvrage d'auteur du XVIII^{ème} siècle. Le

Tableau 2 montre un filtrage effectué sur plusieurs échelles différentes et avec des orientations de 45° et 135° .

Echelle / Orientation	Gabor dans le domaine fréquentiel	Gabor dans le domaine spatial	Gabor avec binarisation
1 45°			
1 135°			
2 45°			
3 45°			
4 45°			

Tableau 2 : Exemples de bancs de filtres de Gabor et de leurs versions résultantes binarisées.

Les zones blanches présentent sur la représentation fréquentielle des images (à gauche) correspondent aux fréquences non filtrées tandis que les zones noires représentent les fréquences filtrées. Les basses fréquences se situent au centre tandis que les hautes fréquences se trouvent aux bords. Par conséquent, plus les zones filtrées se trouvent près du centre, c'est à dire des basses fréquences, plus le fond est globalement lissé. En revanche, si les hautes fréquences sont peu atténuées, les contours sont alors mis en évidence.

Comme nous pouvons le constater sur le

Tableau 2, l'échelle permet de faire varier le diamètre de la zone de fréquences non filtrées. C'est pourquoi, plus l'échelle est grande, plus les hautes fréquences sont filtrées, ce qui provoque une image résultante du traitement de plus en plus floue. Nous avons défini les

paramètres de déviation de la fonction de Gabor de telle sorte que la position des zones blanches soit adaptée à la dimension de l'image à traiter ainsi qu'à ses caractéristiques fréquentielles. Grâce à ce paramétrage, les images résultantes des filtres de Gabor seront directement exploitables pour la suite des traitements.

III.3.4 Exploitation des réponses de Gabor et construction des cartes directionnelles

Du fait du paramétrage du filtre de Gabor qui atténue massivement le basses fréquences et privilégie les plus hautes, on peut constater que les images résultant du banc de filtres présentent un fond globalement très sombre (correspondant à une région sans réponse dans la direction recherchée) et des zones très blanches sur les contours du tracé (correspondant aux régions du tracé manuscrit à forte réponse dans la direction recherchée). L'exploitation de ces réponses va nécessiter une étape de binarisation permettant de localiser précisément les régions de l'image répondant de façon maximale aux directions imposées dans le banc de filtres. Le seuil de binarisation est fixé de manière unique pour toutes les images issues d'un même ouvrage.

La réponse du filtre de Gabor selon les orientations détectées par la rose de direction permet ainsi de décomposer l'image en plusieurs cartes directionnelles. La *Figure 10* récapitule le découpage en plans directionnels binarisés pour une image de tracés manuscrits arabes selon les quatre directions les plus représentées dans l'image ($0^\circ, 42^\circ, 89^\circ, 129^\circ$). Le regroupement des plans directionnels nous permet de reconstruire le contour de l'écriture de notre image (voir *Figure 10*) et de confirmer ainsi le recouvrement complet des formes du tracé, autorisant de ce fait à ne pas considérer de directions supplémentaires. Le recouvrement est le résultat de trois opérations *ET logique* entre les quatre images binarisées.

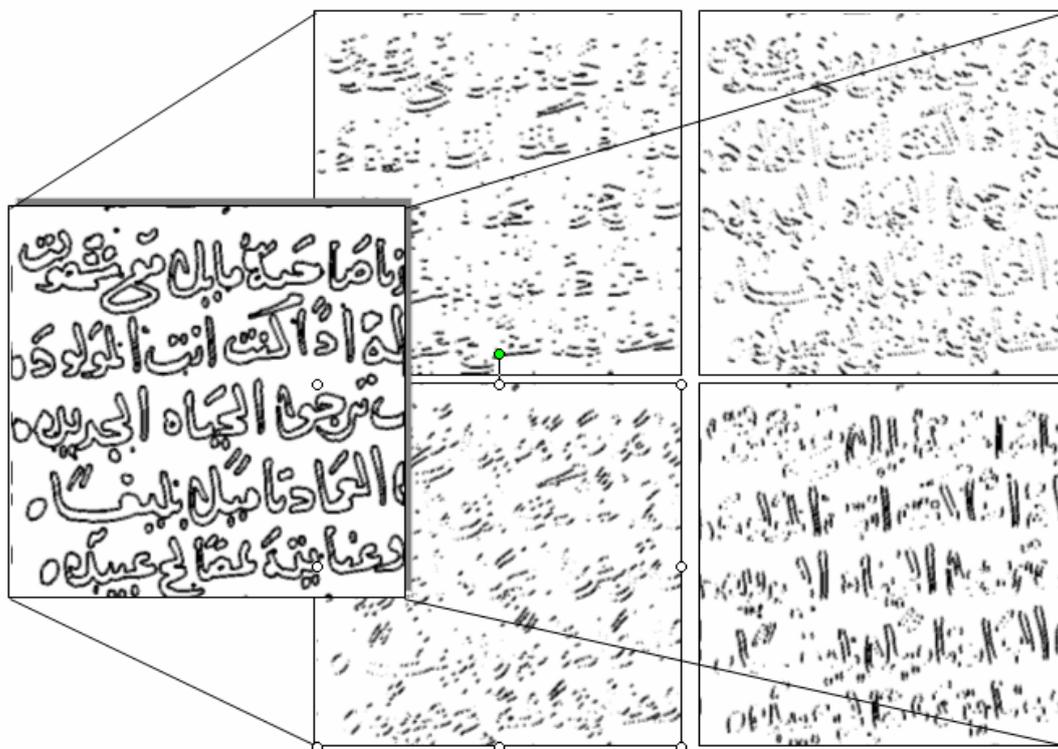


Figure 10 : Filtrage de GABOR, séparation en plans directionnels binarisés.

Représentation du recouvrement finale par opérations logiques.

III.4 Conclusion

Le filtrage de Gabor, selon les orientations choisies par la rose de directions, nous a permis de faire une séparation en plans directionnels binarisés de l'image. En seconde étape à ce processus nous proposons une extraction d'esquisses présentées sur les cartes binaires et à partir d'un traitement adapté des réponses des bancs de filtres de Gabor. Ce traitement revient à faire une élimination des petites surfaces de bruit, de quelques pixels, et à remplir les aspérités présentes dans les formes connexes insuffisamment continues. Après ce traitement, le processus de localisation des esquisses similaires peut être engagé.

IV. EXPLOITATION DES ESQUISSES ET

MESURES DE SIMILARITES

PARTIELLES

Dans ce chapitre, nous présentons, le traitement appliqué aux réponses binaires des filtres directionnels de Gabor permettant de localiser les esquisses de formes et d'estimer les niveaux de similarités existant entre elles. Ce traitement varie d'une image à une autre, vu la variation et la non régularité des formes des tracés dans les pages manuscrites. Il consiste, d'une part, à éliminer les régions non souhaitées mises en évidence par les bancs de filtres de Gabor et qui correspondent la plupart du temps à des zones de bruit du fond de l'image ou à de petites composantes non significatives du tracé. D'autre part, le traitement consiste à remplir les zones de rupture parfois présentes à l'intérieur des formes connexes des réponses des filtres (voir *Figure 11.b*). Nous présentons ensuite la définition du prototype référent appelé « *Esquisse de référence* » estimée pour chaque carte directionnelle et servant de base à la définition des indices de similarité entre esquisses. Nous proposons ensuite une définition d'un taux de similarité entre formes estimé à partir des caractéristiques des esquisses de référence. L'estimation du taux de similarité entre les différentes esquisses permet finalement de localiser précisément les zones de tracés redondantes dans l'image initiale. Nous présentons dans la dernière partie de ce chapitre l'intégration de cette étape de recherche de similarités à un processus complet de compression.

IV.1 Traitement appliqué aux cartes directionnelles

L'épaisseur très fine des esquisses existantes dans les cartes directionnelles rend les opérations de matching entre les différentes formes difficiles à faire. Pour résoudre ce problème, nous avons appliqué aux cartes directionnelles un traitement un peu particulier et adapté aux types d'écritures rencontrées. En premier lieu, une dilatation adaptative est appliquée sur les réponses des filtres de Gabor. Elle consiste à remplir les « trous » existants dans les formes connexes et à élargir l'épaisseur des esquisses. La méthode consiste à repérer les pixels blancs dans une imagerie carré de taille 3*3 et à faire une dilatation adaptative selon la variation de leur voisinage. Une permutation du niveau de couleur de ce pixel se fait à

partir d'un seuillage adaptatif variant selon l'écriture traitée (c'est-à-dire fine ou épaisse). Ce seuil varie entre 1 et 8 et est proportionnel au nombre des voisins objets (pixels noirs). Par exemple pour le manuscrit arabe (*Figure 11.b et 11.d*), nous avons fixé le seuil à 3.

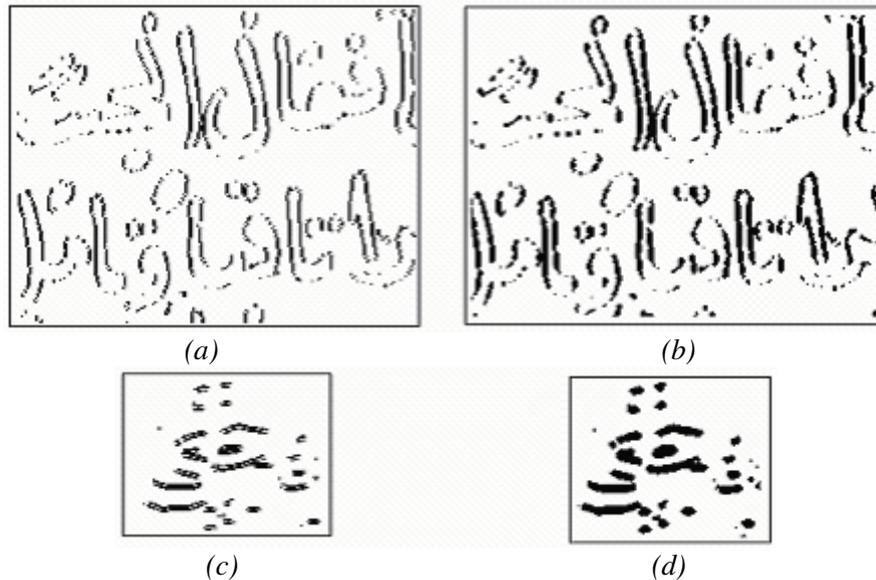


Figure 11 : a) réponse de Gabor selon la direction 90° (et 0° dans (c)) ; b) et d) Dilatation adaptative des formes connexes

L'application de la dilatation adaptative sur les cartes directionnelles permet de mettre en évidence la présence des esquisses les plus significatives de la réponse de Gabor en les épaississant. Cette opération valorise les éléments saillants et également les petites traces considérées comme peu significatives pour l'application (esquisses de trop petites dimensions, ou du bruit de filtrage). Il reste donc à supprimer de cet ensemble ces éléments sans intérêt (de petites surfaces) pour la direction considérée.

L'érosion adaptative consiste ainsi à éliminer les formes connexes ayant une surface inférieure à un seuil choisi. L'estimation de ce seuil dépend de plusieurs critères et varie d'une carte directionnelle à une autre pour une même écriture. Pour fixer ce seuil, nous avons estimé l'amplitude de la variation existante entre la surface maximale et minimale des esquisses présentes sur une carte directionnelle (ce processus est réitéré pour toutes les cartes), voir *Figure 12*.

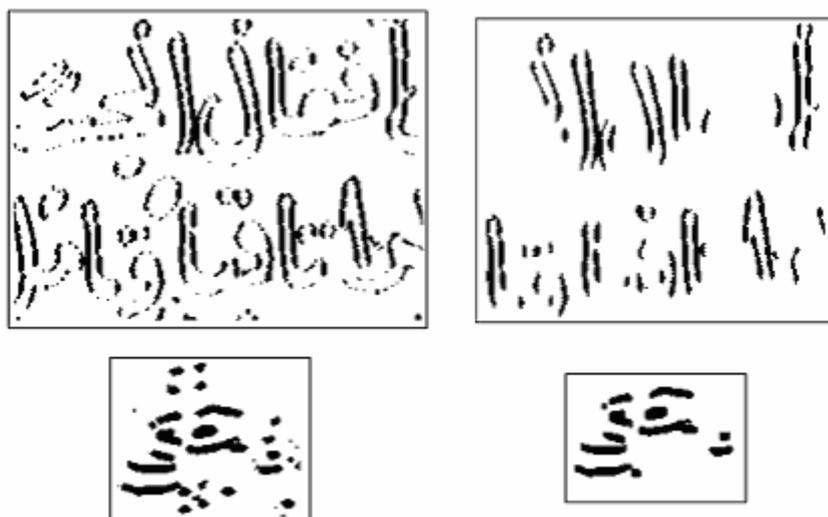


Figure 12 : Erosion adaptative des petites surfaces

Dans le cas de l'exemple traité, nous avons choisi d'éliminer toutes les esquisses ayant une surface S vérifiant la relation $S < \text{Surface minimale} + 10\%$ de l'amplitude de variation (*Figure 12*). Ce seuil garantit un bon compromis entre d'une part la suppression des éléments peu significatifs du tracé (et qui seront rendus visibles dans d'autres directions : aucune information n'est perdue dans ce traitement) et d'autre part la conservation des éléments pertinents de formes pour la direction considérée (et qui constituent les graphèmes très significatifs pour l'orientation considérée, voir *Figure 12*).

IV.2 Formation de connexités par opérations morphologiques et classification des esquisses

Pour détecter les formes connexes dans les cartes directionnelles, nous avons utilisé une approche de la segmentation en régions selon une méthode de croissance de régions (*Region Growing* ou *algorithme de remplissage récursif*). Une alternative à cette approche est la méthode de découpage récursif (*Region Splitting*) que nous n'avons pas exploitée ici. La méthode de croissance de régions consiste à détecter toutes les régions homogènes noires en choisissant initialement N « germes », équivalant à un pixel. Les pixels voisins sont ensuite fusionnés à la région tant que la couleur de la région reste homogène. Nous répétons cette opération pour chaque région jusqu'à ce que les régions n'évoluent plus. La composante connexe est alors intégralement localisée. A ce stade, on peut estimer tous les paramètres de formes, de densités, et autres propriétés de formes nécessaires à la définition de la composante connexe de référence qui est l'esquisse de référence définie pour chaque carte directionnelle et définie en introduction de cette partie.

La détection des formes connexes nous permet de localiser facilement les esquisses dans les cartes directionnelles. Il suffit d'attribuer aux pixels appartenant à une forme connexe le numéro de l'esquisse. Cette indexation simple permet de localiser facilement et de distinguer les formes entre elles.

IV.3 Estimation du taux de similarité entre esquisses

IV.3.1 Taux de similarité et transformations affines au sein d'une même carte directionnelle (*intra-carte*)

Le taux de similarité permet d'estimer le degré de ressemblance entre les esquisses d'une même direction. L'estimation du taux de similarité est obtenue à partir d'une comparaison flexible entre les imageries des esquisses directionnelles en utilisant une opération « AND logique ». Cette opération consiste à comparer les pixels des imageries et à incrémenter le compteur d'un pas lorsque ces deux pixels contiennent le même niveau d'information. Le taux de similarité est égal à la valeur enregistrée par le compteur (après calcul de matching) divisée par la surface maximale des deux esquisses.

$$T_{i,j} = \frac{\sum_{k,l} n(k,l)}{\text{Max}(\text{Surface}_i, \text{Surface}_j)} \quad \text{avec} \quad \begin{cases} n(k,l) = 1 & \text{si } P_i(k,l) = P_j(k,l) = 0 \\ n(k,l) = 0 & \text{ailleurs} \end{cases}$$

Selon ce principe et compte tenu de la différence de taille des fenêtres contenant les esquisses directionnelles, nous avons eu recours à une opération préliminaire de normalisation de la taille des blocs contenant ces esquisses. Parallèlement à ces opérations et toujours dans le but d'augmenter le degré de ressemblance entre esquisses, nous avons appliqué quelques transformations géométriques affines aux esquisses telles que la translation et la rotation. La translation est utilisée au sein d'une même carte directionnelle : elle permet un

repositionnement physique plus adapté d'une esquisse par rapport à l'esquisse à partir de laquelle la similarité est calculée. La translation possible d'une esquisse par rapport à l'autre est appliquée selon les axes horizontaux et verticaux D_x et D_y avec une variation de pixels pouvant varier de -2 à $+2$ pixels, soit une translation totale possible de 5 pixels. Pour les cartes d'une même direction, nous avons également autorisé aux esquisses une possibilité de rotations d'un angle ξ pouvant varier de -10° et $+10^\circ$ permettant d'associer deux esquisses similaires et pas totalement orientées dans la même direction. Ceci est lié au fait que les réponses de Gabor révèlent pour une orientation fixée certains graphèmes dans les directions voisines aux directions du paramétrage. Ces transformations sont envisagées afin d'augmenter le degré de similarité entre esquisses. La similarité par rotation au sein d'une même carte directionnelle est estimée systématiquement pour toutes les esquisses. Elle permet d'associer des formes très similaires qui n'auraient pas pu être mises en correspondance dans un cas général d'estimation statique de la mesure de similarité. Cette estimation flexible de l'indice de similarité entre formes garantit ainsi des taux de similarité plus forts.

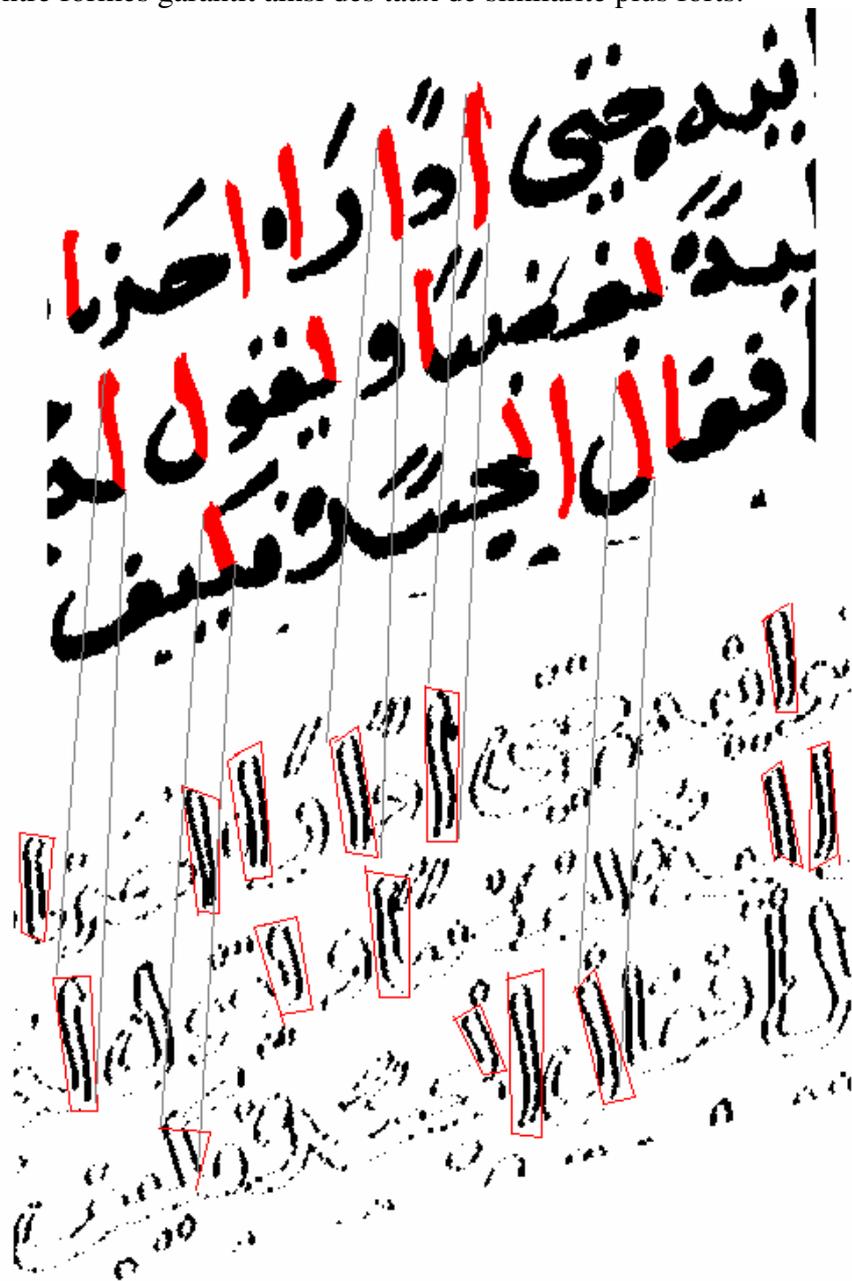


Figure 13 : illustration de quelques traits similaires à une rotation près (pour la direction de 90°)

La *Figure 13* représente une illustration de quelques zones de traits similaires à une rotation près pour la réponse de Gabor à 90° . Une représentation colorée du pavage du texte pour montrer le recouvrement global/partiel des zones de similarités.

IV.3.2 Taux de similarités et transformations affines entre cartes directionnelles (*inter-carte*)

La mesure de similarités peut également s'estimer plus largement sur des esquisses de cartes directionnelles différentes en faisant subir aux esquisses des rotations d'angles plus importants que les petites valeurs estimées sur une même carte. Cette recherche d'invariance par changement de directions entre esquisses de deux cartes directionnelles différentes repose sur l'application d'une opération de rotation définie par la différence des angles de deux cartes directionnelles ($\theta_2 - \theta_1$). Chaque esquisse normalisée subit cette transformation. En réalité, seules les esquisses de référence subissent cette opération afin de limiter les calculs. Nous verrons dans la section suivante à quels niveaux elles interviennent.



Figure 14 : illustration de quelques traits de similarités entre différentes cartes directionnelles à des transformations géométriques près (rotation et changement d'échelle)

L'analyse multidirectionnelle ainsi développée permet de réduire considérablement la taille de la base de référence des esquisses et d'augmenter ainsi les similarités présentes au sein du tracé. Afin de compléter cette étude, il sera abordé dans les perspectives de ces travaux la prise en compte - en plus des opérations précédentes - de transformations liées aux *changements d'échelle* permettant de comparer des esquisses similaires à échelles variables.

IV.4 Critère de recherche de l'esquisse de référence pour chaque carte directionnelle

Etant donné un ensemble d'esquisses directionnelles caractérisées par un certain poids estimés entre chacune d'elle (le taux de similarité), nous allons présenter la méthode développée pour la sélection du prototype de référence (pour chaque carte directionnelle). A partir de cette représentation matricielle du taux de similarité, nous simulons symboliquement une carte directionnelle en un « arbre génétique » dont les pères sont les esquisses de référence de la carte et les fils sont les esquisses reconstruites à partir de leur père. On note

qu'une carte directionnelle peut être représentée par une ou plusieurs esquisses de référence dont le nombre est déterminé automatiquement par notre algorithme.

Le principe de l'estimation de l'esquisse de référence s'inspire de l'algorithme de Dijkstra exploitée dans la théorie des graphes permettant de déterminer le chemin le plus court pour parcourir les nœuds d'un graphe, ([37], [42]). En simulant la carte directionnelle par un graphe, nous avons considéré que les esquisses directionnelles sont les nœuds du graphe et le taux de similarité $T_{i,j}$ est le coût du chemin entre les deux nœuds « i » et « j ». La difficulté de notre problème revient à bien choisir le nœud père le plus représentatif d'un regroupement de nœuds et en même temps présentant le minimum de pertes lors de la reconstruction des esquisses à partir de lui.

Nous allons présenter le principe général de l'algorithme que nous avons développé permettant de définir automatiquement les esquisses de référence dans une carte directionnelle. Le premier pas consiste à parcourir le tableau contenant le taux de similarité entre les différentes esquisses et à trouver la valeur maximale, soit $T_{i,j}$ cette valeur. Si $T_{i,j}$ est plus grand qu'un seuil S_1 défini, par exemple 88%, l'esquisse « i » sera considérée la première esquisse de référence de notre carte directionnelle et l'esquisse « j » sera reconstruite à partir de « i ». On teste ensuite tous les taux de la ligne « i », et selon ce même principe toutes les valeurs $T_{i,k}$ plus grandes que le seuil S_1 seront reconstruites à partir de l'esquisse « i ». Soit K l'ensemble des esquisses reconstruites à partir de l'esquisse de référence « i ». Le second pas de l'algorithme consiste à tester les lignes des esquisses contenues dans l'ensemble K , c'est à dire les lignes des esquisses reconstruites. Par exemple, en testant la ligne « j », toutes les valeurs $T_{j,k}$ plus grandes que le seuil S_2 (avec $S_2 > S_1$) seront reconstruites à partir de « j ». Notons à ce stade qu'il faut choisir une valeur de S_2 élevée (pour garder notre exemple initiale, une valeur plus grande que 85%) pour ne pas augmenter les pertes d'informations à la reconstruction des esquisses.

	1	i	j	k	n
1	∅								
∴									
i			∅		$T_{i,j}$		$T_{i,k}$		
∴									
j			$T_{j,i}$		∅				$T_{j,n}$
∴									
k			$T_{k,i}$		$T_{k,j}$		∅		$T_{k,n}$
∴									
n			$T_{n,i}$		$T_{n,j}$		$T_{n,k}$		∅

Tableau 3 : représentation graphique de la recherche de l'esquisse de référence

La seconde étape de l'algorithme consiste à reprendre récursivement la même démarche décrite précédemment mais cette fois en diminuant le seuil S_1 d'un pas variant de 1 à 5% et en augmentant le seuil S_2 d'un pas variant de 1 ou 2%. Cette fois, on teste toutes les lignes des esquisses non reconstruites. On répète ces itérations dès que S_1 est plus grand que la valeur 75%. En deçà de cette valeur seuil, on commence à perdre beaucoup d'informations à la reconstruction. Le figure 18 en fin de ce mémoire récapitule sur un schéma synoptique le principe général de l'estimation des similarités entre esquisses de formes.

IV.5 Intégration de l'étape de recherche de similarités dans un processus de compression adaptée aux tracés manuscrits

L'exploitation des indices de similarité entre les esquisses manuscrites donne naissance à une nouvelle technique de partitionnement basée sur un découpage intelligent de l'image centré tout autour des zones informantes du tracé. Nous avons montré précédemment dans le chapitre 2, que l'échec des méthodes de compressions existantes est dû en partie au mauvais partitionnement choisi pour localiser les similarités et les redondances entre les zones d'information. Le découpage des zones d'informations selon la répartition des esquisses du tracé manuscrit que nous proposons fournit une base exploitable pour conduire intelligemment un nouveau partitionnement de l'image. Ce partitionnement doit être réalisé sur l'image en niveaux de gris d'origine à partir de la localisation et la distribution des esquisses binaires des formes.

L'étape de recherche de similarités va donc pouvoir s'intégrer à un processus complet de compression où les similarités inter-esquisses serviront de références aux similarités originales contenues dans les régions d'intérêt en niveaux de gris. Le principe retenu et non développé dans le cadre de ce DEA consiste à construire un graphe de similarité reproduisant le comportement du graphe portant sur les esquisses binaires et créant ainsi des liens de similarités effectives entre les graphèmes présents sur l'image d'origine, voir *Figure 15*.

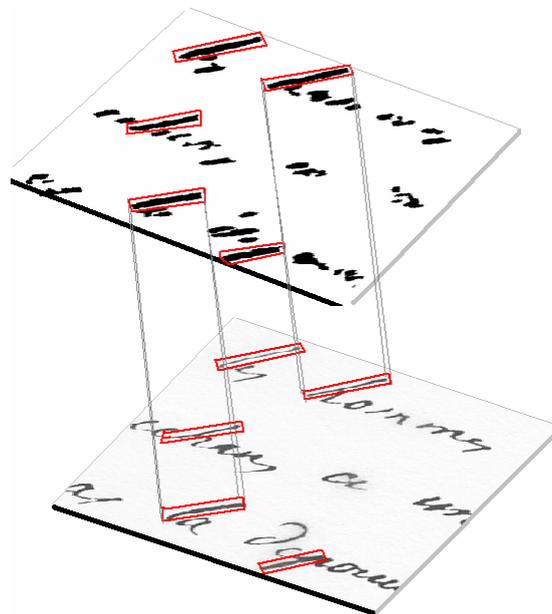


Figure 15 : illustration des liens existant entre esquisses (estimées par filtre de Gabor à 40°) et les graphèmes originaux de l'image en niveaux de gris

L'idée que nous avons retenue et que nous envisageons de développer en perspectives directes de ces travaux consiste à faire une comparaison flexible entre les imagerie des graphèmes isolés à partir de leur représentation en esquisses directionnelles et à ne conserver

que les graphèmes de référence (les esquisses de références) dans un dictionnaire. Cette comparaison a pour finalité de définir un tableau d'adresses permettant de faire pointer chacun des graphèmes fils vers le graphème père issu de « l'esquisse de référence » et d'estimer une mesure permettant de coder leurs différences. A l'aide des informations contenues dans le dictionnaire, le tableau d'adresses et le codage des différences entre graphèmes de références et graphèmes cibles, nous envisageons une reconstruction des formes du tracé de l'image initiale. Ce processus sera reproduit jusqu'à l'exploitation complète de tous les liens existant entre les esquisses initialement étudiées.

A ce stade, nous pouvons relever une différence essentielle et non moindre existant entre cette approche et les techniques de compression adaptée aux tracés imprimés (DjVu ou DEBORA) : les techniques portant sur le texte imprimé reposent sur une compression exhaustive de toutes les images de caractères de l'ouvrage à compresser tandis que notre approche n'envisage qu'une comparaison partielle entre graphèmes à partir de leurs représentants sous forme d'esquisses de formes non segmentables à priori.

Le taux de compression des zones de tracés est naturellement proportionnel aux taux de redondance rencontrée durant la comparaison entre graphèmes relevant des esquisses de références. Il est défini par le ratio existant entre le nombre d'esquisses de référence d'une image (qui peut être ramené au nombre de graphèmes de références), c'est-à-dire entre les esquisses enregistrées dans le dictionnaire N_D et le nombre d'esquisses totales présentes dans la carte directionnelle N_T .

$$\text{Taux de redondance des esquisses} = 100 * (1 - N_D / N_T)$$

Nous ne pouvons pas donner une estimation numérique générale du taux de similarité car le développement de cette idée ne s'inscrit pas dans le cadre de DEA. Néanmoins sur les images de manuscrits arabes traitées nous avons obtenu des taux de redondance intra-carte qui varient entre 58 et 72% selon le choix de seuil de similarité entre esquisses.

La seconde étape à envisager à ce stade concerne le codage de l'arrière plan, qui outre une information moyenne contenant principalement des basses fréquences peut également contenir des portions de tracés non détectés par les bancs de filtres directionnels. Il est donc important de ne pas lui appliquer de techniques de compression avec pertes au risque de créer des zones de rupture préjudiciables à une bonne lecture du manuscrit et a fortiori à une exploitation scientifique pertinente. Finalement, le rendu définitif d'une image décompressée nécessite une estimation d'un plan de compensation faisant une différence entre les données réelles d'origine et les données transmises reconstruites. Ce plan de compensation permet de restituer une image parfaite. Pour reconstruire l'image exacte originale à partir du plan décompressé, il suffit de refaire une différence symétrique entre l'image décomprimée et le plan compensatoire. Le plan composite compensé ainsi obtenu restitue sans erreur l'image d'origine.

IV.6 Discussion et perspectives

IV.6.1 Synthèse et bilan positif des travaux

Ce travail est une contribution originale et innovante à l'extraction et la caractérisation des zones de textes manuscrits pour lesquelles aucune segmentation n'est possible en raison de la mauvaise qualité du support, de la grande variabilité et l'irrégularité des formes présentes. Ce travail s'inscrit dans un objectif de compression des images de tracés pour lesquelles, nous l'avons vu, aucune méthode n'est à ce jour réellement opérante. Nous avons donc proposé une méthodologie complète de recherche et de caractérisation des tracés manuscrits à partir d'une extraction de formes partielles issues des bancs de filtres de Gabor et conduisant à une représentation complète des zones de textes en graphèmes orientés (les

esquisses binaires). Les formes fortement redondantes ont ainsi pu être caractérisées et la construction complète d'un graphe de similarité entre esquisses de formes a été proposée. Notre contribution dans le cadre de ce DEA constitue les briques logicielles qui serviront à la plateforme complète de compression actuellement développée en C++ au laboratoire LIRIS.

IV.6.2 Limitations dans la recherche de similarités

La recherche de similarité à partir du traitement de la réponse de Gabor présente parfois une certaine limitation liée en premier lieu à la nature de l'écriture traitée.

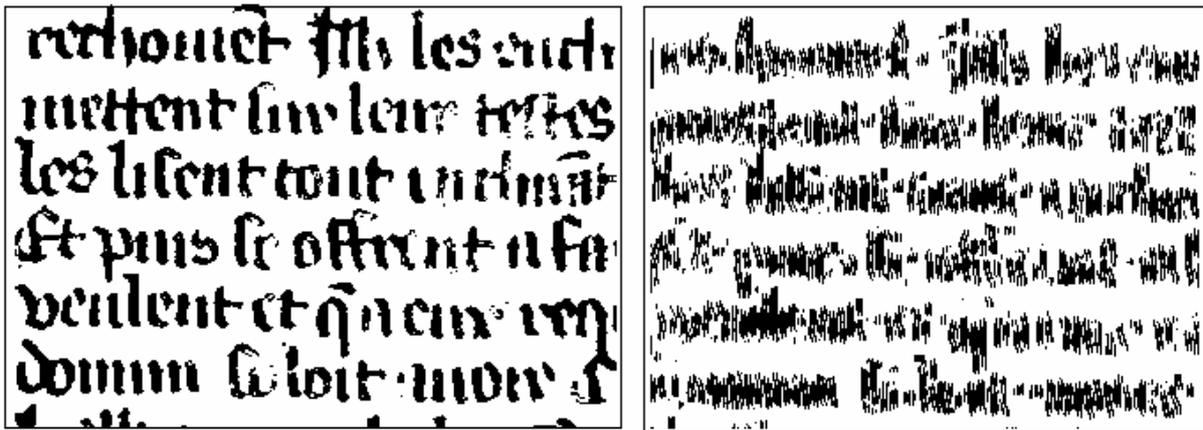


Figure 16 : filtrage d'une écriture latine médiévale dans la direction de 90° par le filtre de Gabor

Par exemple dans la *Figure 16*, on remarque que les traits présents dans la réponse de Gabor après application des opérations morphologiques sont trop peu espacés pour parvenir à segmenter le plan binaire en esquisses séparables. L'estimation de la similarité demeure ici inopérante. La raison de ces empâtements est essentiellement liée au caractère majoritairement vertical des formes du tracé. Les réponses du filtre de Gabor dans les autres directions retenues (0° et 40°) sont quasi inexistantes.

Une autre difficulté a été rencontrée dans le cas des documents présentant des tracés très fins pour lesquels les réponses de Gabor étaient également peu séparables. Dans ce cas, il a fallu envisager une hypothèse supplémentaire autorisant la fusion de deux réponses binarisées au voisinage d'un même contour. A partir de cette hypothèse, l'analyse par esquisses directionnelles a pu être poursuivie sans difficulté, voir *Figure 17*.

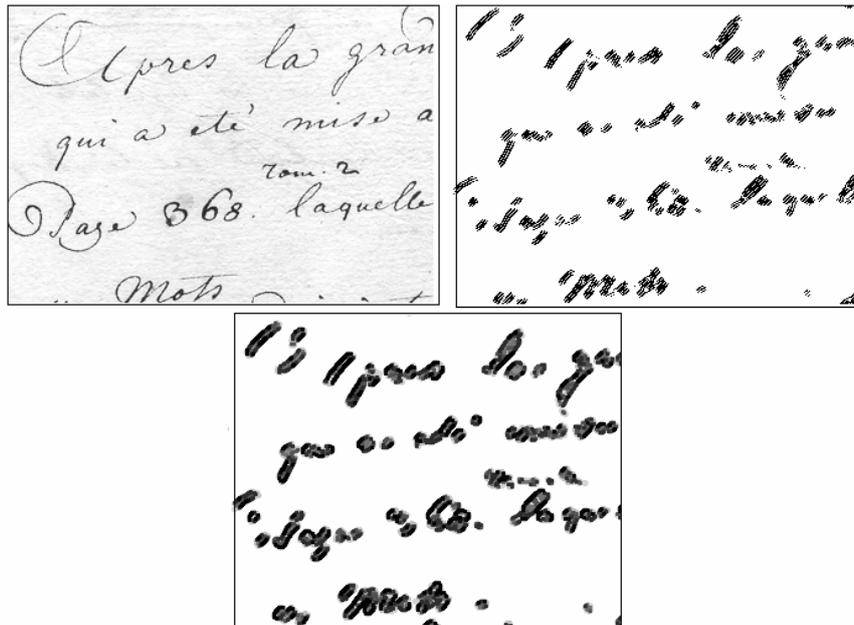


Figure 17 : application du filtre de Gabor sur une écriture de Montesquieu dans la direction de 45° et résultat de la dilatation adaptative aux traits de la réponse

IV.6.3 Perspectives directes et travaux à venir

La première amélioration à apporter à ces travaux concerne la localisation complète des tracés : selon notre approche, certaines zones manuscrites peu contrastées dont les fréquences sont trop faiblement différenciables de celles de l'arrière plan ne peuvent être repérées par les bancs de filtres directionnels. Dans ce cas, la reconstruction du texte ne pourra pas tenir compte de ces zones rejetées. C'est la raison pour laquelle la restitution intégrale et sans perte de l'arrière plan dans le cadre d'une transmission en réseau est impérative. Cependant nous devons envisager de systématiser le traitement qui permet de localiser finement ces fragments de manuscrits oubliés. Une approche complémentaire plus locale doit être envisagée.

Dans un second temps et pour augmenter le taux de similarité entre les esquisses, on pourra introduire d'autres types de transformations géométriques comme le changement d'échelle. Cette opération permet de diminuer le nombre des esquisses de référence du dictionnaire des nœuds pères et d'augmenter la similarité entre esquisses.

Par ailleurs, cette étude de recherche de similarités dans les tracés manuscrits va être étendue à une combinaison entre différentes esquisses voisines de direction différente permettant de retrouver une similarité plus globale au niveau d'un caractère ou même d'une portion de mot, voir éventuellement d'un mot complet. L'amélioration portant sur cette étape de recherche de similarités doit également conduire à une amélioration des temps de compression.

Finalement, ce travail va déboucher sur un projet de mise en place d'un système de compression complet sur des images en niveaux de gris sans perte d'informations. Ces travaux futurs débiteront directement à la suite de ce stage. Les premières briques logicielles sont actuellement en cours de développement.

IV.7 Conclusion générale

Durant ce stage de DEA, nous avons développé une approche générique adoptée pour pallier aux difficultés de traiter des images de textes manuscrits du patrimoine fortement bruitées, dont les formes ne sont pas segmentables par des approches classiques développées en analyse d'images de documents. Cette approche repose sur une représentation robuste des tracés basée sur la recherche de similarités entre formes (esquisses de formes) tenant compte d'hypothèses non restrictives sur les changements d'échelle, les possibilités de rotation et de translation. Pour trouver cette similarité entre deux formes semblables, nous avons repéré les points saillants de contours par le détecteur de Harris et les régions de saillance directionnelle par les bancs de filtres de Gabor dans les directions privilégiées des tracés exploitant la fonction d'autocorrélation appliquée sur l'image spectrale des traits. La recherche de similarités entre les zones d'images dépend des certaines transformations géométriques affines comme la translation et la rotation inter et intra carte directionnelle. Cette étape de recherche de similarités a été introduite pour pallier aux lacunes des méthodes actuelles de compression d'images non adaptées aux tracés manuscrits. Nous proposons dans une section le principe d'intégration de cette étape dans un processus complet de codage et de compression des régions d'intérêt de ces images.

Ce travail constitue une première démarche innovante dans le domaine de la caractérisation et la recherche de similarités dans les zones de tracés manuscrits : les pistes de développements futurs sont donc nombreuses.

REFERENCES BIBLIOGRAPHIQUES

- [1] Hubert EMPTOZ, Jean-Pierre DALBERA, Bertrand COUASNON, « *Numérisation et patrimoine* », Document numérique, vol. 7, n°3-4, Hermes, 2003, 188p
- [2] Isabelle WESTEEL, Martine AUBRY, « *La numérisation des textes et des images : Technique et Réalisations* », Presses de l'université Charles de Gaulle - Lille III, 2003, ISBN 2-84467-050-4, 190 p.
- [3] Stéphane BRES, Jean-Michel JOLION, Frank LEBOURGEOIS, « *Traitement et analyse des images numériques* », Hermes, 2003, 412 p.
- [4] Projet européen « DEBORA », livre en ligne : <http://rfv6.insa-lyon.fr/debora/client.htm>, juin 2000.
- [5] Véronique EGLIN, « *Contributions à la structuration fonctionnelle des documents imprimés* », Thèse de doctorat, Lyon, France, INSA de Lyon, 1998, 251p.
- [6] Véronique EGLIN et Catherine VOLPILHAC-AUGER, « *Caractérisation multi-échelle des tracés manuscrits en vue de la catégorisation de scripteurs* », actes de CIFED04, Colloque International Francophone sur l'Écrit et le Document, La Rochelle, 21-25 juin 2004, pp. 106-114.
- [7] Audrey SEROPIAN, Michel GRIMALDI, Nicole VINCENT, « *Différenciation entre alphabets dans des textes manuscrits* », actes de CIFED04, Colloque International Francophone sur l'Écrit et le Document, La Rochelle, 21-25 juin 2004.
- [8] Ying ZHANG, Lai-Man PO, « *Fractal Color Image Compression using vector distortion measure* », Proceeding of IEEE International Conference on Image Processing, Washington, III – 276 to III – 279, D.C, U.S.A, October 1995.
- [9] Geoffrey M. DAVIS, « *A wavelet-based analysis of fractal image compression* », IEEE Transactions on Image Processing, 7(2):141-154, February 1998.
- [10] Abdel BELAID « *Reconnaissance automatique de l'écriture et du document* », LORIA , Document en ligne : www.colisciences.net/pdf/reconnaissance.pdf*
- [11] Stéphane BRES « *Contributions à la quantification des critères de transparence et d'anisotropie par une approche globale. Applications au contrôle de qualité de matériaux composites* ». Thèse de doctorat, Lyon, France, INSA de Lyon, 1994, 231p.
- [12] BRUCE, GREEN. « *La perception Visuelle : Physiologie, psychologie et écologie* ». Presse universitaire de Grenoble, 1993, 431p.
- [13] ROSENFELD « *A digital straight line segments* ». IEEE Transactions on Computers, 1974, vol.23, pp.1264-1269.

- [14] PRATT, FAUGERAS, GAGALOWICZ, « *Visual discrimination of stochastic texture field* », IEEE Transactions on systems, Man and Cybernetics, 1978, vol. SMC-8, N°11, pp.796-804.
- [15] PRATT, « *Digital Image Processing* », 2nd edition New – York: Wiley, 1991, 230p.
- [16] KUNT, « *Traitement numériques des signaux* », Lausanne : Editions Presses Polytechniques Romandes, Paris : DUNOD, 1981, 402p.
- [17] Thomas P. WELDON, William E. HIGGINS, « *Algorithm for designing multiple Gabor filters for segmenting multi-textured images* », IEEE International Conference on Image Processing, Chicago, Illinois, 4-7, October 1998.
- [18] Fen YANG, Roly LISHMAN, « *Land Cover Change Detection Using Gabor Filter Texture* », Proceedings of the 3rd international workshop on texture analysis and synthesis, 17 Oct 2003, co-located and in conjunction with ICCV2003 Nice, France.
- [19] Krystian MIKOLAJCZYK, Cordelia SCHMID, « *Indexing based on scale invariant interest points* », Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada pages 525 – 531, 2001.
- [20] Krystian MIKOLAJCZYK, Cordelia SCHMID, « *A performance evaluation of local descriptors* », 2003 Conference on Computer Vision and Pattern Recognition (CVPR '03) – Volume II, June 18 - 20, 2003, Madison, Wisconsin.
- [21] Krystian MIKOLAJCZYK, Cordelia SCHMID, « *An affine invariant interest point detectors* », Proceedings of the 7th European Conference on Computer Vision-Part I, p.128 – 142, May 28 – 31, 2002.
- [22] Cordelia SCHMID, Roger MOHR, Christian BAUCKHAGE, « *Evaluation of Interest Point Detectors* », International Journal of Computer Vision, v.37 n.2, p.151-172, June 2000.
- [23] Abdurrahman CARKACIOGLU, Fato-Yarman VURAL, « *Learning similarity space* », IEEE 2002 International Conference on Image Processing (ICIP 2002), Rochester, New York, September 22-25 2002, Paper: MA – P6.
- [24] Y. Y. TANG, C. D. YAN, M. CHERIET, C. Y. SUEN, « *Handbook of pattern recognition and computer vision* », ch. 3.6 : Automatic analysis and understanding of documents, World Scientific Pub., Singapore, 1993, pp. 625-654.
- [25] J. HA, R. M. HARALICK, I. T. PHILLIPS, « *Document Page Decomposition by the Bounding-Box Projection Technique* », 3rd ICDAR: International Conference on Document Analysis and Recognition (Montréal, Canada), August 1995, vol. 2, pp. 1119 – 1121.
- [26] A. K. JAIN, B. YU, « *Document representation and its application to page decomposition* », PAMI: Pattern Analysis and Machine Intelligence, 1998, vol. 20, n°3, pp. 294 – 308.
- [27] Frank LEBOURGEOIS, « *Localisation de textes dans une image à niveaux de gris* », CNED'96 : 4^{ième} Colloque National sur l'Écrit et le Document (Nantes, France), Juillet 1996, pp. 207 – 214.
- [28] Frank LEBOURGEOIS, Hubert EMPTOZ, « *Documents Analysis in Gray Level and Typography Extraction Using Character Pattern Redundancies* », 5th ICDAR: International Conference on Document Analysis and Recognition (Bangalore, India), September 1999, pp. 177 – 180.

- [29] F. BAPST, R. BRUGGER, A. ZRAMDINI, R. INGOLD, « *Integrated Multi – agent Architecture for Assisted Document Recognition* », DAS: Document Analysis Systems (Malvern, Pennsylvania), October 1996, pp. 172 – 188.
- [30] D. G. ELLIMAN, « *A review of segmentation and contextual analysis techniques for text recognition* », Pattern Recognition, 1990, vol. 23, n°3, pp. 337 – 346.
- [31] F. WAHL, G. CASEY, K. WONG, « *Block segmentation and text extraction in mixed text / image documents* », Computer Graphics Image Processing, 1982, n° 20, pp. 375 – 390.
- [32] http://developpeur.journaldunet.com/tutoriel/gra/011026gra_fractales.shtml*
- [33] <http://www.irisa.fr/temics/Accord/CDROM/timc/compression/intro.html>*
- [34] <http://donut.99.free.fr/En-vrac/tipe/fractale.html>*
- [35] <http://www-ensimag.imag.fr/elevés/perso/Simon.Derr/tipe/node4.html>*
- [36] http://www.mines.u-nancy.fr/~tisseran/I33_Reseaux/compression.images/fractale.compression.html*
- [37] http://fr.wikipedia.org/wiki/Th%C3%A9orie_des_graphes*
- [38] <http://www.geog.umontreal.ca/Geotrans/fr/ch2fr/meth2fr/ch2m1fr.html>*
- [39] <http://www.eduscol.education.fr/index.php?./D0015/graphes.htm>*
- [40] http://www.csaffluents.qc.ca/wjbm/matieres/mat514/pages/Math_514/theo_514/t_graph.htm*
- [41] <http://www.aromath.net/Page.php?IDP=624&IDD=0>*
- [42] <http://www.nimbustier.net/publications/dijkstra>*
- [43] http://fr.wikipedia.org/wiki/Algorithme_de_Dijkstra*
- [44] <http://www.irem.univ-mrs.fr/activites/lycee/lygrap08.php>*
- [45] <http://wims.auto.upsud.fr/wims/wims.cgi?lang=fr&+module=U1%2Fgraph%2Fdijkstra.fr>*
- [46] http://rfv6.insa-lyon.fr/flebourg/Option/cours2_fichiers/frame.htm*

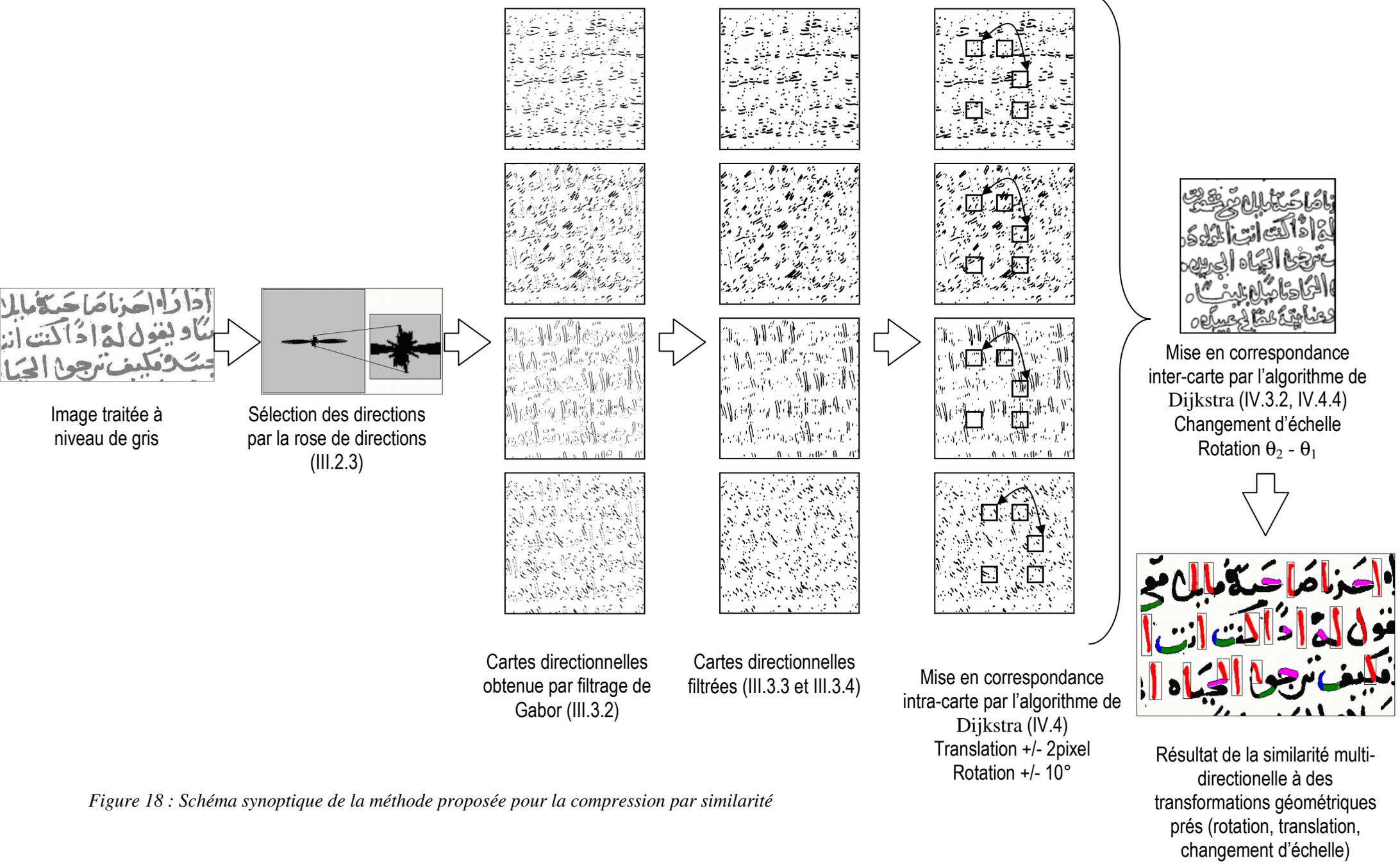


Figure 18 : Schéma synoptique de la méthode proposée pour la compression par similarité