République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mouloud Mammeri de Tizi- Ouzou

Faculté de Génie Electrique et d'Informatique
Département d'Electronique

MEMOIRE

Présenté en vue de l'obtention du diplôme
de Magister en Electronique

Option : Télédétection

Segmentation contextuelle d'image de documents par analyse de texture en niveau de gris

Présenté par :

M^elle^ Oudjemia Souad

Soutenu le 27/06/2010
Devant le jury composé de :

Président	Mme MAZOUZI Zohra épouse AMEUR	Maître de conférences à l'UMMTO
Rapporteur :	AMEUR Soltane	Professeur à l'UMMTO,
Examinateurs :	ZIANI Rezki	Maître de conférences à l'UMMTO
	LAGHROUCHE Mourad	Maître de conférences à l'UMMTO
	LAHDIR Mourad	Maître de conférences (B) à l'UMMTO

Remerciements

Le travail que nous avons l'honneur de présenter, a été effectué en grande partie au niveau de laboratoire d'Analyse et de Modélisation des Phénomènes Aléatoires (LAMPA) de la faculté de génie électrique et d'informatique de l'Université Mouloud Mammeri de Tizi-ouzou (UMMTO).

J'exprime mes sincères remerciements à mon directeur de mémoire Monsieur AMEUR Soltane, professeur à l'UMMTO pour m'avoir guidée et encouragée tout au long de l'accomplissement de mon travail de magister. Qu'il soit assuré de ma respectueuse reconnaissance.

Je remercie vivement Madame AMEUR Zohra, maître de conférences à l'UMMTO d'avoir accepté de présider le jury de ce mémoire.

Que Monsieur ZIANI Rezki, maître de conférences à l'UMMTO, trouve ici l'expression de mes remerciements les plus respectueux pour l'intérêt qu'il a porté à ce travail en acceptant de participer à ce jury.

J'exprime mes sincères remerciements à Monsieur LAGHROUCHE Mourad, maître de conférences à l'UMMTO pour avoir accepté de faire partie du jury.

Je tiens à remercier Monsieur LAHDIR Mourad, maître de conférences à l'UMMTO pour l'intérêt qu'il a porté à ce travail en acceptant de participer à ce jury. C'est avec un grand plaisir que je le compte parmi le jury chargé d'examiner ce mémoire.

Mes remerciements vont aussi à tous ceux qui ont contribué, de près ou de loin à la réalisation de ce travail, en particulier à Mr HAMOUCHE Kamel. Je remercie également Mr Ait Bachir Youcef, Mr Haddab Salah membres de l'équipe de recherche ‘‘Traitement de signal’’ et Mme Iassamen Alia.

Dédicaces

Je dédie ce travail à :

La mémoire de mon père qui malgré qu'il n'est pas avec moi dans ce monde mais il est toujours dans mon cœur surtout qu'il m'a toujours encouragé dans mes études.

Ma chère mère qui ma toujours soutenu et encouragé.

Mes sœurs et frères surtout ma sœur ''FATIHA''.

Mon cher fiancé bakhlish qui m'a donné la force et le courage pour continuer mes études.

Tous mes amis.

Résumé

Les documents électroniques offrent la facilité de stockage et de recherche d'information, pour cela il est nécessaire de passer du format papier vers un format électronique, cette conversion est souvent réalisée par un système d'analyse et de reconnaissance de documents. Ce travail s'inscrit dans la problématique de la reconnaissance de structure physique de document imprimé complexe. Cette dernière opère sur la reconnaissance des différents constituants (texte, fond, image,….).

Notre travail consiste à analyser la structure physique en utilisant les matrices de cooccurrence basées sur la texture particulière du texte, ainsi que sur l'utilisation des descripteurs en niveau de gris pour rendre compte de l'information statique mais avant cela une diffusion anisotrope est utilisée comme prétraitement.

Cette méthode a été appliquée sur des images tests de type latins et arabes. Nous avons pu atteindre un taux d'erreur de classification de 1.79% grâce au tests effectués pour avoir les paramètres les plus pertinent à ce type d'image et grâce au choix de la taille du bloc qui vaut 32*32. Cette division de l'image nous a permet de réduire le temps de calcule et atteindre la valeur 91.13s pour une taille d'image de 768*1074.Les résultats des tests ont montré aussi que notre méthode est insensible à l'inclinaison des documents contrairement aux autres méthodes qui ont montré leurs limites quand il s'agit des documents inclinés

Mots-clés: matrice de cooccurrence, segmentation contextuelle, descripteur, texture, image document.

SOMMAIRE

Introduction

Introduction

Introduction

Depuis l'avènement de l'écriture, aux environs du III ème millénaire avant Jésus Christ (JC), l'être humain n'a cessé d'améliorer ce moyen de communication. En effet, plusieurs civilisations ont apporté leur savoir faire dans le domaine de l'écriture. Il y a eu d'abord l'écriture des hiéroglyphes des pharaons, l'écriture chinoise, l'écriture grecque, et puis l'écriture arabe et romaine. Certaines de ces écritures ont disparu avec la destruction de leur civilisation alors que d'autres sont encore d'actualité. Malheureusement parmi les 3000 langues dénombrées dans le monde, seul une centaine est dotée d'un système d'écriture. Mais la percée majeure de l'écriture a atteint son apogée au 15ème siècle après JC avec l'invention de l'imprimerie par Gutenberg. Et c'est à partir de cette époque que l'écriture est entrée dans une nouvelle ère, à savoir l'ère du document imprimé.

La notion de document est très générale et il existe une panoplie de définitions. La définition la plus appropriée d'un document est la suivante : « un document est le support physique pour conserver et transmettre de l'information ». Selon le support choisi, un document peut être textuel, graphique, multimédia (sonore, vidéo). Le document imprimé a eu un essor en deux temps grâce à l'introduction de l'imprimerie et de l'informatique. L'avènement de l'ère de l'information n'a pas ralenti ni réduit le nombre de documents circulant partout dans le monde. En effet, l'informatique a permis la sauvegarde de centaines de millions de documents existants en format papier, en les numérisant, et à en extrayant leurs contenus. Et c'est à ce moment que la reconnaissance d'images de documents a vu le jour.

.

La reconnaissance de structures de documents est indispensable pour integrer les documents sur papier dans un système de gestion documentaire.Les connaissances,qu'elles soients techniques,scientifiques,historiques,économoque,juridiques ou médicales etc..sont en majorité mémorisés et véhiculées par des textes. Celles qui ont été publiées récement sont

directement accessibles sous forme électroniques. Par contre, les anciens documents ne sont disponibles que sous forme de document papier. Ainsi , nous sommes confrontés à un besoin énorme de retraitement, dit aussi conversion rétrospective pour passer à un format électronique.Cette conversion est réalisée par un système d'analyse et de reconnaissance de documents.Il ne suffit pas de reconnaître les caractères qui forment le contenu d'un document, pour l'identifier mais il est indispensable de reconnaître également sa structure physique et logique. Ainsi, un document dont on a reconnu la structure et le contenu, pourra aisément être restitué sous un format proche de celui d'origine. Il pourra également être archivé, consulté, mis à jour, transféré, etc.…

Il n'existe pas de méthodes type pour reconnaître la structure d'un document, car la fantaisie de représentation des documents nous oblige à prendre en consédiration en premier lieu les modèles de documents qu'on veut structurer. La mise en page (disposition des colonnes de texte, existence d'image ou graphiques…) nous renseigne sur les méthodes à mettre en œuvre pour faire ressortir en premier lieu la structure physique du document. Elle seule ne suffit pas, car il faut adapter des règles qui permettront d'identifier les objets physiques et attribuer des étiquettes logiques aux portions de base (titre, sous titre, paragraphe, résumé,…).

Notre étude porte sur la reconnaissance de la structure physique de document imprimé de type complexe. Le caractère innovant de ce travail se trouve essentiellement dans la nature de la méthode proposée, qui utilise des traitements sur des images à niveau de gris et non binaire.

Notre mémoire a été structuré de la manière suivante :

Dans le premier chapitre nous présentons des généralités sur l'analyse de document.

Dans le deuxième chapitre nous donnons un état de l'art sur les méthodes utilisées pour l'analyse de la structure physique.

Nous présentons dans le troisième chapitre la solution retenue et les différentes étapes pour sa mise en œuvre.

Dans le quatrième chapitre nous présentons les tests et les résultats obtenus que nous commentons et que nous comparons aux résultats obtenus par d'autres techniques pour évaluer la pertinence de notre méthode.

Nous concluons notre mémoire, en résumant notre contribution et en proposant quelques perspectives à ce travail.

Chapitre I

Généralités sur l'analyse de documents

I. Préambule

L'analyse et la reconnaissance d'images de documents désignent une discipline scientifique qui regroupe un ensemble de techniques informatiques dont le but est de reconstituer le contenu d'un document à partir de son image. Le but ultime de la reconnaissance d'images de documents est de générer une représentation de haut niveau sous la forme de documents structurés, selon une forme adéquate pour l'application visée.

II. Problématique générale

Avant d'entrer dans le vif du sujet et étudier les techniques de reconnaissance, il est utile de décrire plus précisément les objectifs de la reconnaissance d'images de documents. Pour ce faire, nous allons établir une typologie des documents et énumérer les applications visées.

La notion de document est très générale et peut, selon le contexte, évoquer tout objet servant à conserver ou à transmettre de l'information tangible pour l'être humain ; selon cette définition, un document peut être textuel, sonore, graphique ou vidéo. Dans ce travail nous limiterons nos considérations à des documents ayant une représentation statique, c'est-à-dire une image fixe pouvant être reproduite sur papier. Des études de documents du point de vue de leur contenu, de leur complexité et de leur support ont été présentées par R. Ingold [1].

II.1 Contenu

Les documents potentiellement intéressants pour la reconnaissance possèdent une très grande diversité de contenu. Les documents se distinguent dans un premier temps entre ceux à prédominance textuelle et ceux à prédominance graphique. Dans le cas du texte, il convient de distinguer les différentes langues et écritures. Les langues occidentales sont caractérisées par un alphabet restreint et des signes en général bien isolés. L'écriture arabe est une écriture cursive (figure 1.*a)* qui utilise au contraire de nombreuses ligatures, compliquant ainsi la localisation des caractères. Quant aux langues asiatiques, elles peuvent se servir de plusieurs

milliers de pictogrammes formant des structures internes relativement complexes (figure 1.*b)*. Par ailleurs, il est nécessaire de tenir compte de différents types d'écritures. Au niveau des écritures mécaniques, on peut distinguer les textes dactylographiés, et les caractères d'imprimerie ; à cela s'ajoute une diversité des fontes et des styles d'écriture. Dans l'écriture manuscrite, on distingue les écriture en lettres isolées et l'écriture cursive.

Les documents considérés comprennent potentiellement aussi des éléments non textuels, à commencer par des logos ; à cela s'ajoutent des objets plus complexes tels que les formules mathématiques, les formules chimiques, etc. Enfin, les documents peuvent contenir des illustrations .

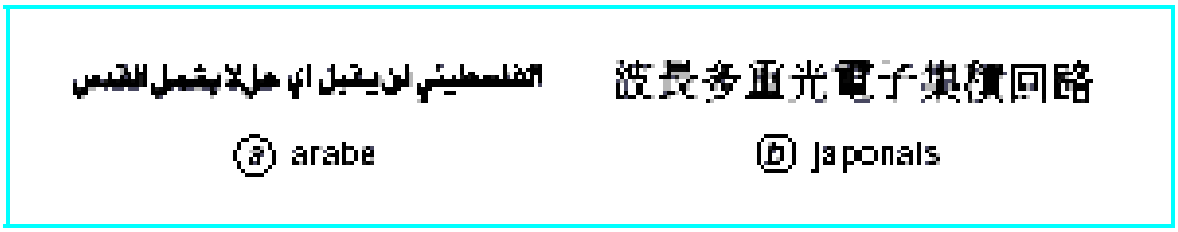


Figure1.1 : Extraits de textes orientaux.

II.2 Complexité

Les documents présentent de nombreuses variations quant à leur organisation. Les plus simples telles que les œuvres littéraires se caractérisent par une structure linéaire alors que les ouvrages scientifiques, les manuels pédagogiques ou les textes de loi présentent une organisation hiérarchique en chapitres, sections, sous-sections, articles, paragraphes et alinéas. Les documents administratifs, tels que des factures, des documents comptables, etc., présentent une structure bidimensionnelle sous forme de tableaux. Ces derniers se caractérisent par une organisation cellulaire plus ou moins régulière, plus ou moins explicite selon la présence ou non de filets séparant les lignes et les colonnes. En plus de ces structures complexes, les formulaires présentent la particularité de posséder des champs fixes (pré imprimés) et des champs variables.

Les magazines et les journaux présentent, quant à eux, des structures de pages particulièrement complexes qui ne suivent pas toujours des règles systématiques. Enfin, les documents techniques regroupent des informations textuelles et graphiques, mais la combinaison des deux formes d'information répond à des règles dépendantes du domaine.

Ainsi, un schéma électrique ne suit pas les mêmes conventions que le dessin d'une pièce mécanique, même si les constituants graphiques élémentaires sont pour la plupart les mêmes. La figure 1.2 illustre la variété des documents pris en considération.

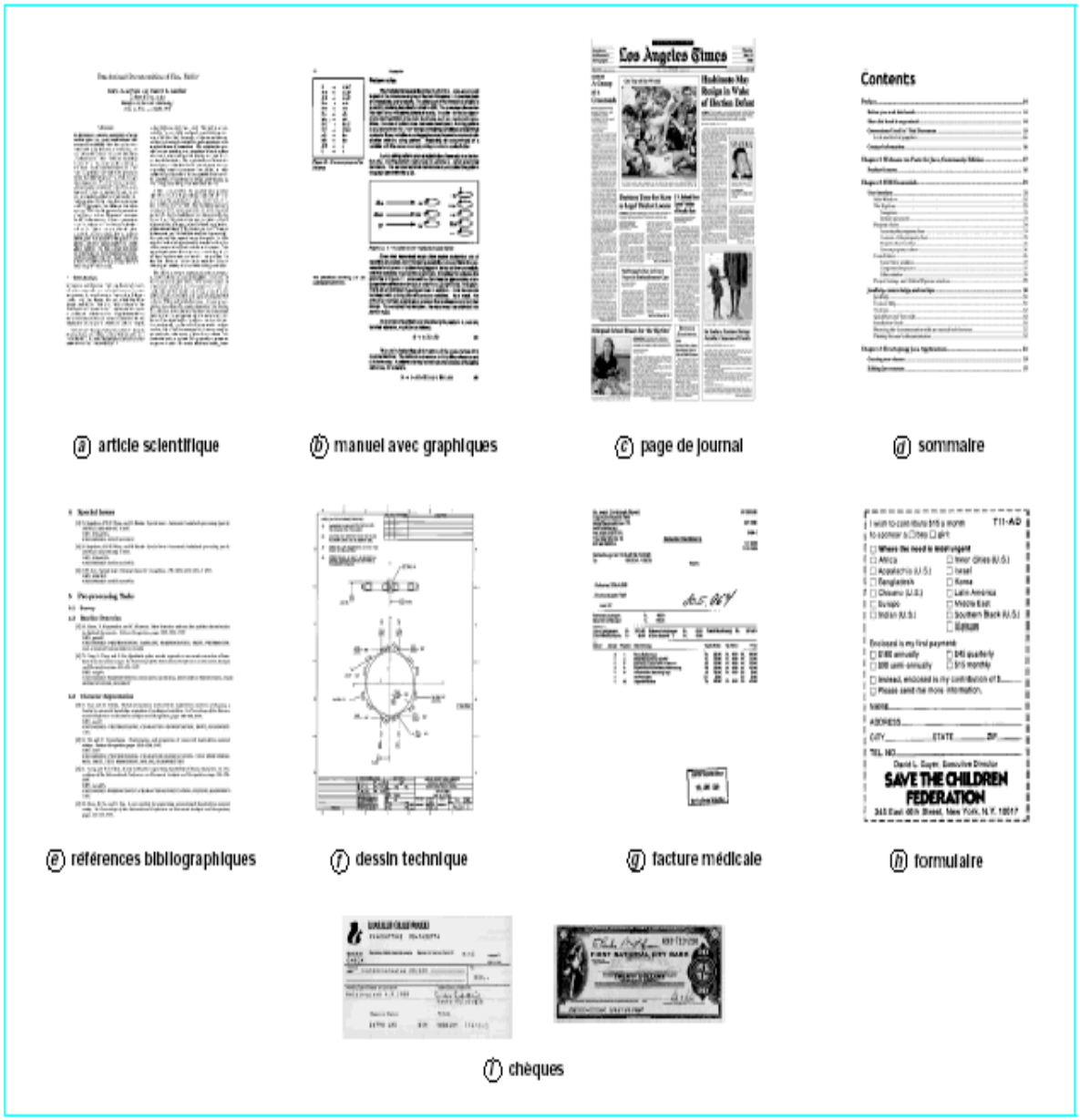


Figure 1.2 : Documents destinés à la reconnaissance.

II.3 Qualité des images

L'origine et l'histoire d'un document peuvent avoir des répercussions importantes sur les caractéristiques de l'image qui en résulte. Ainsi, la transparence du papier, son

jaunissement par la lumière ou sa déformation par l'humidité engendre toutes sortes de difficultés supplémentaires. C'est souvent le cas avec les anciens documents appartenant au patrimoine culturel. Par ailleurs, dans les administrations, on ne travaille souvent pas avec des documents originaux mais avec des photocopies dont la qualité peut être fortement dégradée. Il peut y avoir du « bruit » (de petites taches) occasionné par les poussières. Le texte peut être noirci ou éclairci et les lignes peuvent être déformées, surtout vers les extrémités de la page.

II.4 Domaines d'application

Dans le monde entier, il existe un nombre faramineux de documents papiers. Ces derniers sont de différents types tels que les journaux, les revues, les livres, les encyclopédies, etc.… Un grand volume de ces documents est conservé dans des bibliothèques nationales. Le reste est stocké soit dans les administrations, soit dans des musées, soit dans les bibliothèques universitaires, soit dans les entreprises et, à un degré moindre, dans nos bibliothèques personnelles.

Afin de préserver ces documents de tout genre de détérioration ou d'éventuelle décomposition, qui pourraient survenir, des techniques de conservation sont nécessaires.

La numérisation permet de s'affranchir des problèmes posés par la conservation des documents papier. Le coût de stockage des documents électroniques est inférieur au coût de stockage des mêmes documents au format papier. Il est à noter que le coût de duplication des documents électroniques se trouvant dans un cédérom est bon marché et de cette façon nous assurons une conservation plus longue avec les documents électroniques par rapport aux documents au format papier. Néanmoins la numérisation des documents papiers nécessite des ressources matérielles et humaines qui engendrent des coûts à supporter.

La numérisation seule est insuffisante pour l'extraction d'information du document électronique ; en revanche, elle permet de préparer le terrain à la reconnaissance. La reconnaissance de documents permet l'extraction d'information et porte essentiellement sur les documents papier. Les travaux de recherche portant sur la reconnaissance de documents ont fait des grandes avancées dans ce domaine. Cependant, le domaine de la reconnaissance de documents n'est pas encore un problème résolu.

L'apport de la numérisation et de la reconnaissance de documents est indiscutable et a contribué à étoffer Internet. En effet, la reconnaissance d'images numérisées est importante du point de vue de l'information extraite, puisqu'elle permet d'indexer un grand nombre de documents.

Les applications de la reconnaissance de documents couvrent plusieurs domaines. En effet, nous trouvons des applications de reconnaissance de codes et adresses postales, de reconnaissance de chèques, de reconnaissance de formulaires, d'archivage de documents et de reconnaissance de factures médicales.

Les applications de reconnaissance de codes postaux sont très répandues dans les centres de tri postaux. Elles se basent sur la reconnaissance de l'écriture cursive manuscrite [2] [3]. Parmi ces applications, nous trouvons celles relatives à la reconnaissance manuscrite de chiffres [4] [5].

Une autre application utile est la reconnaissance de formulaires [6] [7]. Les formulaires sont généralement constitués d'une partie fixe et d'une partie variable. La reconnaissance de formulaires repose sur la séparation entre ces deux parties.

Les applications d'archivage de documents traitent assez bien les documents électroniques. Elles se focalisent sur deux axes : il y a celles qui stockent les documents dans leur formats d'origine et celles qui convertissent le document dans un autre format commun.

Le taux de récupération des documents archivés à partir d'une base de données se trouve nettement amélioré si une extraction des structures et une indexation des données du document ont été effectuées. L'extraction des structures permet d'améliorer considérablement la pertinence de la recherche et aussi de faciliter la réédition de documents.

Comme application permettant l'extraction des structures, Dengel [8] a développé smartFIX, un système d'analyse et de compréhension de documents. Il permet le traitement des factures médicales. Ces dernières sont des documents composites ; elles renferment des tables, du texte et des annotations manuscrites. SmartFIX a été développé dans le but de faciliter le traitement de ce genre de documents pour une institution d'assurance maladie. Il est à noter que cette dernière reçoit un nombre élevé de factures qui possèdent une grande variabilité d'un médecin à un autre et d'un laboratoire d'analyses médicales à un autre.

III. Composantes d'un Système de reconnaissance de documents

Il existe plusieurs traitements qui doivent être effectués pour aboutir à la reconnaissance complète d'une page d'un document. La figure 1.3 décrit de manière schématique l'architecture d'un système de reconnaissance de documents. Cette décomposition est assez générale mais ne s'applique pas nécessairement à tous les systèmes, dans la mesure où certains traitements peuvent être superflus ou s'articuler différemment en raison de l'application prévue. La première opération consiste à acquérir le document sous forme

d’image numérique. En général, cette étape est complétée par divers prétraitements afin de faciliter le travail des processus ultérieurs. Le plus souvent, le prétraitement s’achève par la binarisation, car, en effet, la reconnaissance proprement dite s’applique généralement sur une image binaire. Elle comprend une phase de segmentation dont le but est d’isoler les régions d’intérêt en distinguant, comme notre diagramme le suggère, le texte, les graphiques et les autres régions. La segmentation d’une page produit ainsi un découpage hiérarchique composé de blocs et de sous blocs.

L’analyse appliquée à chaque région dépend de son contenu. Dans le cas du texte, il s’agit d’effectuer la reconnaissance de caractères, éventuellement complétée par la reconnaissance des fontes afin de disposer d’informations typographiques. Dans le cas d’éléments graphiques, la reconnaissance implique une phase de vectorisation dont le but est d’analyser les traits et de les restituer sous forme de segments. Les formes symboliques restantes doivent faire l’objet d’une identification selon des méthodes qui s’apparentent à la reconnaissance de caractères. Enfin, la fusion de tous ces résultats doit produire une représentation structurée au niveau physique.

L’analyse se termine avec la phase de reconnaissance structurelle dont le rôle est d’effectuer un premier pas vers l’interprétation des résultats sous forme de structures logiques de haut niveau.

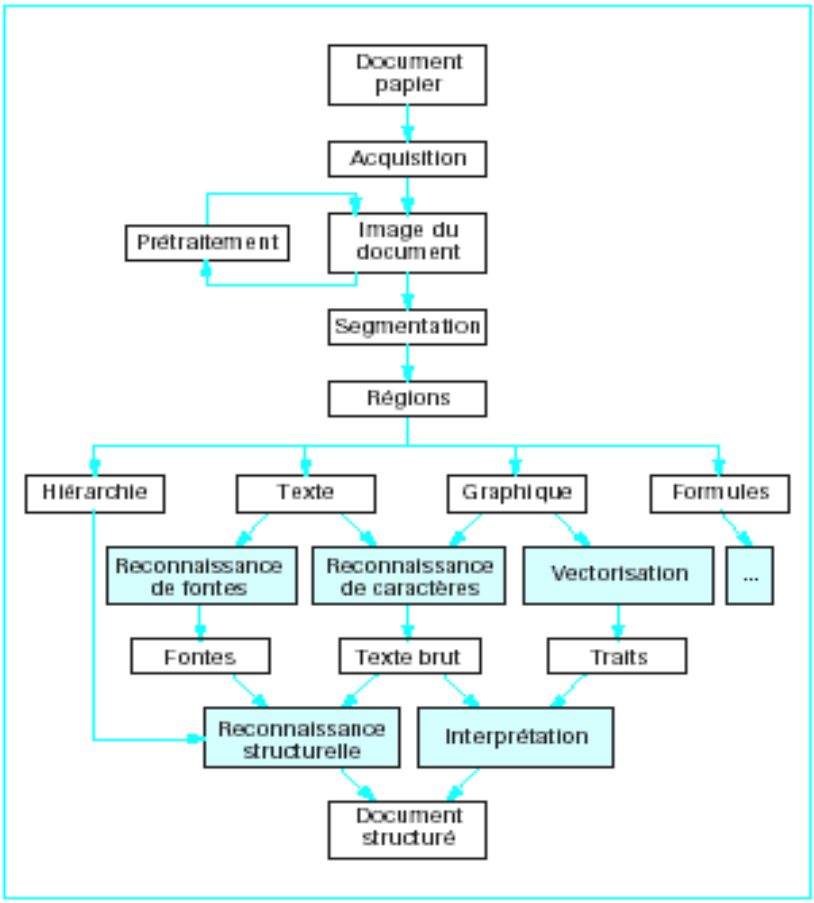


Figure 1.3 : Architecture d'un système de reconnaissance de documents.

III.1 Aquisition des images de documents

L'acquisition d'images de documents se fait le plus généralement au moyen de scanners et, plus rarement, avec des caméras car celles-ci possèdent des inconvénients majeurs comme la sensibilité aux conditions d'éclairage. Le manque de résolution a longtemps constitué un facteur limitatif pour l'analyse de documents ; aujourd'hui, les scanners permettent d'atteindre des résolutions de 1 200 dpi (*dots per inch* ou points par pouce) ou même davantage, en seize millions de couleurs. Pour couvrir les besoins usuels de l'analyse de documents, une résolution de 300 à 400 dpi et 256 niveaux de gris s'avèrent suffisants.

III.2 Prétraitement

La phase de prétraitement opère des transformations au niveau de l'image dans le but d'en faciliter la reconnaissance lors des étapes suivantes. Les principales opérations sont le redressement, le filtrage et la binarisation[9].

- **Redressement :** Lors de l'acquisition de l'image au moyen d'un scanner, pour des raisons mécaniques, le document est rarement parfaitement aligné avec le dispositif de balayage. Il en résulte une image où les lignes ne sont pas parfaitement horizontales. Il s'agit alors de procéder au redressement de l'image au moyen d'une transformation géométrique.
- **Filtrage :** Le filtrage de l'image a pour but de supprimer certaines formes de bruit. On peut utiliser pour cela des techniques de traitement d'image classique, des convolutions ou d'autres types de filtres locaux. Les filtres morphologiques sont, quant à eux, utilisés pour lisser le contour des symboles.
- **Binarisation :** Étant donné que les algorithmes de reconnaissance opèrent en général sur des images en noir et blanc, il est nécessaire de binariser les images. Sur les documents de bonne qualité, de fond blanc, une simple opération de seuillage peut suffire ; par contre, lorsque le papier présente un fond imagé ou texturé ou lorsque le papier a été dégradé par le temps, des techniques plus sophistiquées, fondées sur une analyse locale fine, sont nécessaires. La binarisation entraîne dans certains cas une perte importante d'informations.

III.3 Analyse de la structure physique

III.3.1 Segmentation

La segmentation est une tâche des plus délicates et des plus complexes. Elle a comme objectif de délimiter toutes les régions d'intérêt de l'image, avec un degré de finesse variable. Ainsi, il s'agit de délimiter les régions homogènes appelées blocs (texte, figures, notes en bas de page, etc.), puis, à l'intérieur de ces régions, les lignes de texte, les mots et enfin des caractères.

La détection des filets (lignes de séparation, horizontales ou verticales) et des cadres fait également partie de la segmentation. Une telle étape est surtout utile pour l'analyse de formulaires contenant souvent des grilles et des cases à remplir. Les cadres et les filets ainsi détectés ne servent souvent qu'à des fins de filtrage et de localisation des autres informations, car ils n'ont pas de signification propre. À première vue, il semble que la détection des filets soit facile. Toutefois, si le filet se superpose à un objet de contenu, typiquement du texte, la reconnaissance est confrontée à l'élimination du filet sans détérioration de l'objet superposé. Ce problème a été traité de différentes manières, mais aucune d'entre elles n'est réputée satisfaisante pour traiter tous les cas de figure [1].

III.3.2 Reconnaissance de caractères

Parmi toutes les phases de la reconnaissance de documents, celle de la reconnaissance des caractères constitue sans doute le domaine auquel ont été consacrés le plus de travaux. En ce qui concerne la lecture de caractères imprimés, les systèmes commerciaux sont largement utilisés et ils s'avèrent rentables lorsque les outils sont correctement employés. Des études menées régulièrement par l'université de Las Vegas témoignent d'une progression lente mais constante dans ce domaine [1] .

Les caractères manuscrits isolés sont nettement plus difficiles à reconnaître ; ce n'est que dans des contextes particuliers comme l'analyse de codes postaux que les performances sont suffisantes pour une exploitation commerciale. Les méthodes de reconnaissance de formes reposent sur une étape d'extraction de caractéristiques et une étape de décision. Dans le cas des caractères, l'extraction de caractéristiques produit des mesures (taille, compacité, aire, périmètre, centre de gravité, etc.) et des caractéristiques topologiques (orientation des segments, nombre d'extrémités, etc.). La phase de décision repose sur des classifieurs utilisant des techniques fort différentes.

Il est intéressant de constater que les meilleurs résultats sont obtenus lorsque plusieurs classifieurs différents peuvent être combinés. Dans ce cas, c'est la stratégie de fusion de résultat qu'il convient d'examiner.

III.3.3 Reconnaissance de fontes

La reconnaissance de fontes est un sujet qui a longtemps été négligé par la communauté scientifique. Souvent jugée sans grande importance pratique, cette problématique est pourtant

importante pour au moins deux raisons. La première est que la connaissance *a priori* de la fonte permet d'améliorer la reconnaissance des caractères. En effet, l'identification de caractères dans une fonte connue peut être effectuée avec des fonctions de discrimination plus robustes, étant donné la plus faible variabilité des formes appartenant à chaque classe(voir figure1.4). La seconde raison concerne l'apport d'information pour la reconnaissance des structures logiques. En effet, les différentes entités logiques d'un document se traduisent souvent du point de vue typographique par une fonte, un corps ou un style d'écriture particulier : les citations peuvent être en italique, les titres en gras et les légendes de figures dans un corps plus petit.

La reconnaissance des fontes est étroitement liée à celle des caractères. C'est la raison pour laquelle deux approches complémentaires peuvent être envisagées pour la reconnaissance de fontes : une approche dite *a posteriori*, qui prend en compte le résultat de l'OCR pour comparer les propriétés locales du caractère (par exemple, la forme des sérifs) et une seconde approche, dite *a priori*, qui précède la reconnaissance des caractères et est fondée sur l'analyse des caractéristiques globales d'une ligne ou d'un bloc de texte. La performance de tels systèmes dépend de l'ensemble des fontes considérées. Ainsi, certaines familles de fontes sont si semblables que même des experts en typographie ont de la peine à les discerner.

Figure 1.4 – Extraction de caractéristiques pour la reconnaissance de fontes.

III.3.4 Vectorisation

La vectorisation est la conversion de la partie graphique en une description vectorielle, sous forme de segments de droite, d'arc de cercle et de jonction entre ces primitives géométriques. Exprimé ainsi, le problème semble simple ; pourtant, vu la variabilité des illustrations contenant ce type d'éléments graphiques et les contextes multiples dans lesquels ils apparaissent, ce sujet a donné naissance à une foison de publications traitant de méthodes très différentes les unes des autres. Cette problématique a été approfondie au point que des concours sont organisés pour désigner les meilleures approches appliquées à un contexte universel.

Les traits peuvent survenir comme entité isolée dans des graphiques ou des dessins techniques. Ils peuvent être munis d'attributs comme des flèches par exemple. Parfois, ils sont fusionnés avec d'autres composants tels que des boîtes où servent de primitives pour des surfaces à texture. Enfin, il faut encore tenir compte des lignes discontinues formées de points (lignes pointillées) ou de traits courts (lignes « traitillées »).

Les méthodes utilisées s'appuient en général sur les informations de contours et sur les lignes médianes, aussi appelées squelettes ; ces derniers peuvent être extraits par des opérateurs morphologiques classiques. Une fois le squelette établi, il s'agit de chaîner les pixels les uns aux autres pour former une ligne discrète. L'approximation polygonale peut finalement être construite en découpant récursivement la ligne obtenue de manière à ce que l'ensemble des pixels de la chaîne soit suffisamment proche des segments obtenus. La difficulté réside ensuite dans l'interprétation de ces informations, en particulier lorsque les segments sont discontinus, qu'ils se joignent par les extrémités ou lorsqu'ils se rencontrent en des points de branchement [1].

III.3.5 Reconnaissance de graphiques

L'analyse et la reconnaissance de dessins techniques présente une problématique bien différente et souvent plus complexe que celle rencontrée dans les documents à prédominance textuelle. Une raison à cela provient de la très grande variabilité des documents et de l'absence d'un véritable standard pour représenter les informations. Chaque domaine possède ses propres conventions et même à l'intérieur d'un domaine, les règles changent fortement d'un document à l'autre, en fonction des auteurs et des outils utilisés en production.

D'abord, la segmentation est beaucoup plus complexe que dans les documents textuels. Les informations sont organisées en couches. Ainsi, la segmentation peut pratiquement être qualifiée de tridimensionnelle, dans la mesure où elle dépend à la fois de la disposition planaire des objets et de la couche à laquelle ils appartiennent. Le problème ne peut être résolu que par l'apport d'informations contextuelles caractérisant la catégorie de documents traitée.

Ensuite, les graphiques peuvent contenir des éléments textuels, de présentations variées et orientées dans toutes les directions, ce qui nécessite de redresser l'image avant de procéder à l'identification des caractères.

III.3.6 Reconnaissance de la structure logique

La reconnaissance structurelle de documents, connue également sous la dénomination anglaise de *document understanding*, constitue l'étape finale de l'analyse de documents textuels (ou éventuellement graphiques) ; son objectif est de déterminer l'organisation logique des entités élémentaires ou composées. Il s'agit ainsi :

— de détecter les titres, les légende de figures, les citations, les références bibliographiques, etc.
— de construire l'organisation hiérarchique du document en chapitres, sections, sous-sections
— d'établir les relations transversales, par exemple entre une note ou une figure et sa référence dans le texte.

La ***structure logique*** d'un document n'est pas universelle ; elle dépend à la fois du type de document et de l'application visée. Définie dans un contexte large, la reconnaissance structurelle comprend de nombreux aspects :

— la reconnaissance des macrostructures a pour mission d'étiqueter les blocs de texte et de déterminer les liens entre eux. À ce niveau, on distingue par exemple les titres, les paragraphes, les légendes, les notes de bas de page, etc. ;
— la reconnaissance des microstructures vise à identifier des fragments de texte contenus à l'intérieur d'un bloc. À ce niveau, on identifie par exemple des citations, des définitions ou encore des dates, des références (à la bibliographie, à un autre paragraphe, etc.) ;
— les tableaux constituent des structures bidimensionnelles importantes dans bon nombre de documents. La reconnaissance de tableaux vise à délimiter les cellules et à déterminer les liens entre elles ;
— les formules mathématiques, de par leur organisation bidimensionnelle, sont considérées comme des structures particulièrement complexes.

La reconnaissance des structures de documents n'est pas encore résolue de manière satisfaisante. Seuls quelques systèmes dédiés sont opérationnels et leur degré d'adaptabilité est extrêmement limité. Dans bon nombre d'applications, la reconnaissance structurelle revient à un étiquetage des différents composants. Les structures à établir peuvent être purement linéaires (ensemble de champs) ou hiérarchique. Le langage de structuration de

documents XML (eXtensible Markup Language) semble parfaitement adapté pour représenter ce type d'organisation.

Une difficulté majeure de la reconnaissance structurelle est l'apport de connaissances contextuelles concernant le type de documents et l'application visée. Il est indispensable de connaître les éléments à étiqueter ainsi que les règles de composition autorisées. Des formalismes de type grammaire sont nécessaires, mais la production de ces informations reste problématique ; une solution manuelle s'avère en général fastidieuse, alors que les approches automatiques nécessitent des techniques d'apprentissage encore trop mal maîtrisées.

III.3.7 Classification de documents

Dans bon nombre d'applications, il n'est pas nécessaire de viser la reconnaissance complète du contenu du document et de ses structures. Ainsi, il suffit souvent d'identifier le type de documents afin d'en extraire les informations pertinentes. Du point de vue de la reconnaissance de formes, il s'agit d'un vrai problème de classement. Le principe consiste à extraire du document les caractéristiques pertinentes puis à appliquer un classifieur pour déterminer la classe. Dépendant du contexte, les caractéristiques sont les mêmes que celles décrites dans les paragraphes précédents : des propriétés de l'image, des composantes connexes, des caractéristiques typographiques ou du contenu textuel. Selon l'application visée le classifieur peut être le *k*-plus proche voisin, réseau de neurones artificiels, etc.

IV. Discussion

Tout au long de ce chapitre, nous avons essayé de donner la problématique générale de l'analyse de document qui se manifeste par la diversité de contenu et par les nombreuses variations de leurs organisations. Nous avons montré aussi l'intérêt de la reconnaissance des documents par leurs nombreuses applications dans différents domaines. Ensuite nous avons donné les étapes de la reconnaissance d'un document électronique pour mieux comprendre le travail à effectuer dans les prochains chapitres.

Le domaine d'analyse et de reconnaissance de documents est très vaste. Il existe plusieurs applications de la reconnaissance de documents et par conséquent plusieurs méthodes d'analyse s'imposent tels que la segmentation physique et logique de documents.

Dans le prochain chapitre nous passerons en revue toutes les méthodes qui existent dans la littérature sur la reconnaissance de la structure physique de documents.

Chapitre III

Segmentation d'images de documents par analyse de texture en niveaux de gris

I. Préambule

Les performances des méthodes de reconnaissance d'images de documents ne cessent de s'améliorer. En effet, plusieurs problèmes ont été surmontés grâce à la recherche. Cependant nous sommes encore loin de la perfection et le taux de reconnaissance atteint rarement le 100%.

Dans ce chapitre, nous présentons les différentes méthodes de reconnaissance de la structure physique de documents.

II. Reconnaissance de la structure physique des documents

La structure physique d'un document comporte des entités qui diffèrent d'un document à un autre. En effet, mis à part le texte qui est une entité commune à tous les documents, toutes les autres entités sont des images et des graphiques qui sont généralement insérés au sein des documents soit pour étoffer ces derniers, soit pour expliquer une partie du texte.

Rappelons que le but de la reconnaissance de documents est de retrouver les entités constituant le document. La reconnaissance de structures physiques comprend deux étapes à savoir, la détection et la classification des différentes zones de l'image. La première étape permet de segmenter l'image en zones homogènes et la seconde étape permet d'étiqueter ces zones en tant que texte, images et graphiques. Notons que la décomposition en texte peut être affinée en mots, en lignes et en blocs de texte. Ce choix doit être fait lors de l'étape de segmentation.

Plusieurs articles traitant de la reconnaissance de structures physiques de documents ont été publiés dans la littérature. Certains de ces articles présentent uniquement l'état de l'art pour les méthodes de reconnaissance de structures physiques [10] [11]. D'autres par contre, présentent conjointement l'état de l'art des méthodes de la reconnaissance de structures physiques avec celles relatives aux structures logiques [12] [13] [14] [15] [16].

Les méthodes de reconnaissance de structures physiques, décrites dans la littérature sont des méthodes automatiques et peuvent être classées en quatre catégories à savoir, les méthodes ascendantes, les méthodes descendantes, les méthodes mixtes et les méthodes basées sur la texture.

Les méthodes descendantes opèrent du niveau le plus élevé à savoir, la page et descendent d'un niveau à un autre jusqu'à arriver au niveau des composantes connexes ou au

niveau pixel. Par contre les méthodes ascendantes reposent sur des fusions successives de composantes connexes du plus bas niveau vers le niveau le plus élevé. En effet, la dernière fusion permet la reconstitution de la page une fois les fusions du bas niveau effectuées.

Les méthodes dites « mixtes » sont des techniques qui combinent les approches descendantes et ascendantes pour la reconnaissance de structures physiques.

Les méthodes basées sur la texture regroupent les techniques basées sur les méthodes ascendantes et descendantes auxquelles on rajoute l'information « texture ».

La figure2.1 illustre les approches descendantes et ascendantes.

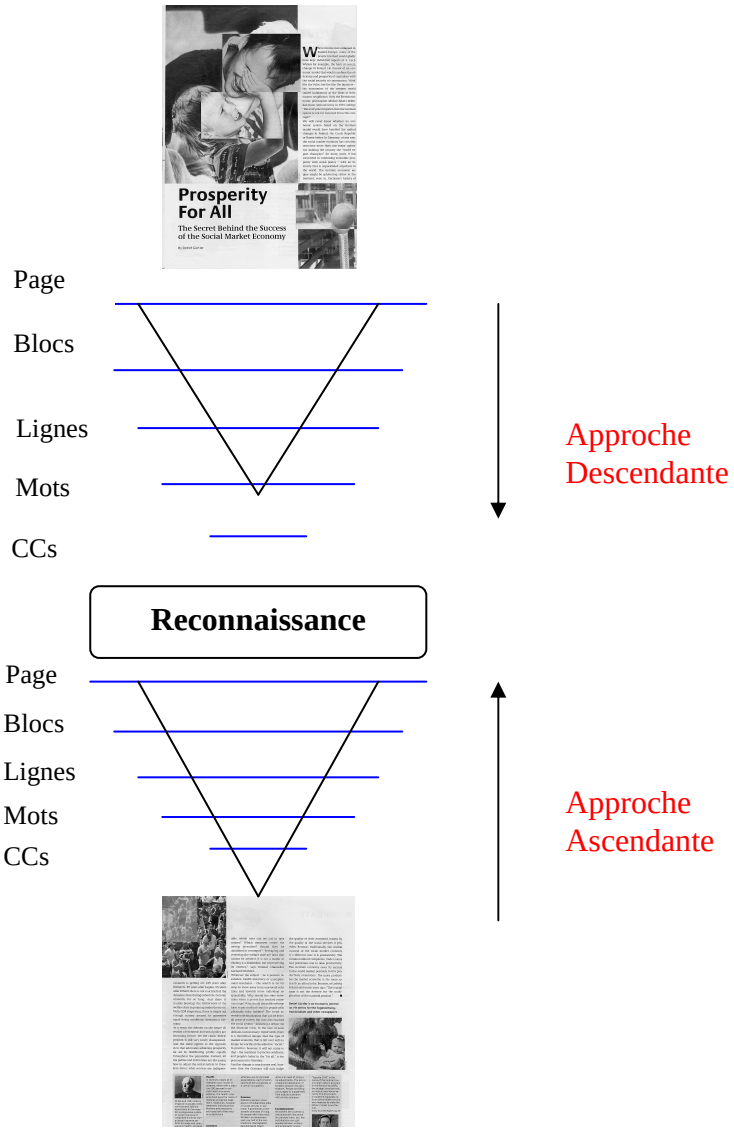


Figure 2.1 : Approches descendante et ascendante

II.1 Méthodes descendantes

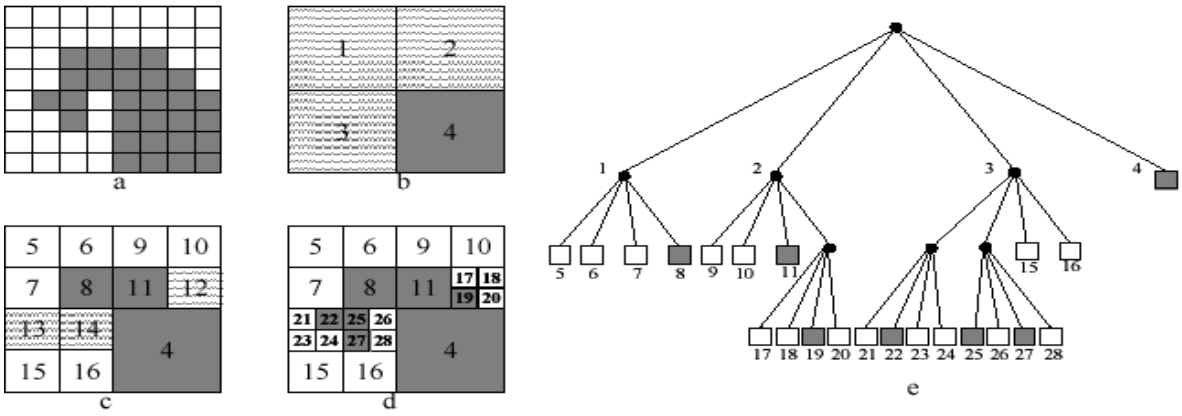
Ces techniques opèrent du niveau le plus élevé à savoir la page jusqu'au niveau des composantes connexes ou au niveau pixel. Nous présentons ci après les algorithmes les plus utilisés.

II.1.1 Algorithme Split and merge

Afin de découper l'image en régions homogènes, la technique du « split and merge » se déroule en deux étapes : une étape de découpage suivie d'une méthode d'agrégation. Le but étant de découper l'image en blocs de plus en plus petits, en fonction de l'homogénéité de l'image

L'étape de découpage consiste à diviser l'image initiale en quatre régions (construction d'un « quadtree ») puis vérifier selon un critère si ces dernières sont homogènes. Les régions inhomogènes sont divisées à leur tour, le processus s'arrête lorsque tous les blocs obtenus sont homogènes. Cette étape produit alors une image sur-segmentée où chaque feuille correspond à une sous-région homogène. L'étape d'agrégation permet ensuite de fusionner les régions homogènes qui auraient été séparées lors de l'étape précédente. Pour réaliser cette fusion, il faut d'abord tenir à jour une liste des contacts entre régions. On obtient ainsi un graphe d'adjacence de régions ou « Region Adjacency Graph ». Ensuite, l'algorithme va marquer toutes les régions comme « non-traitées » et choisir la première région 'R' non traitée disponible. Les régions en contact avec 'R' sont empilées et sont examinées les unes après les autres pour savoir si elles doivent fusionner avec 'R'. Si c'est le cas, la couleur moyenne de 'R' est mise à jour et les régions en contact avec la région fusionnée sont ajoutées à la pile des régions à comparer avec 'R'. La région fusionnée est marquée « traitée ». Une fois la pile vide, l'algorithme choisit la prochaine région marquée « non traitée » et recommence, jusqu'à ce que toutes les régions soient traitées [17]. Un exemple de cet algorithme est illustré par la figure2.2.

Parmi les nombreuses applications de cette technique, citons la segmentation de la vidéo pour retrouver le texte dans l'image [18]. Cette technique a l'avantage de ne pas être sensible à l'inclinaison mais conduit à de mauvais résultats lorsque la mise en page est variable ou lorsque les pages sont mal redressées.



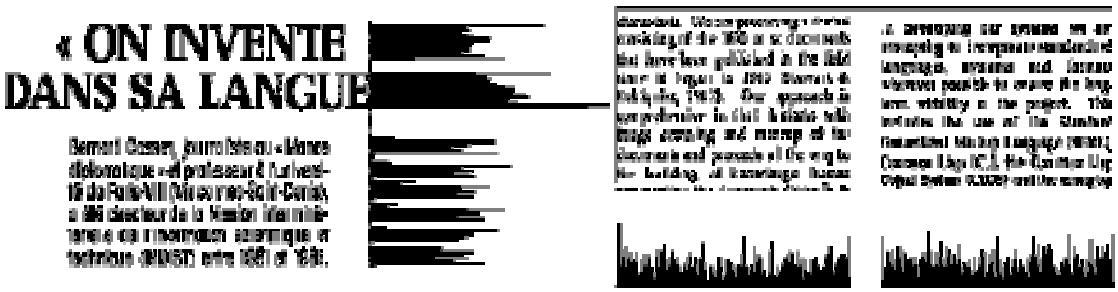
a - image originale
b - premier niveau de séparation
c - deuxième niveau
d - troisième niveau final
e - Quad-Tree correspondant

Figure 2.2 : exemple de l'algorithme split and merge

II.1.2. Algorithme de découpage X-Y ou (X-Y Cut)

Cette technique est la plus connue parmi les méthodes descendantes. Elle utilise les méthodes de profils de projection qui ont été introduites par Nagy [19]. Le principe de la méthode du XY-Cut est d'utiliser une projection horizontale et verticale afin de trouver les espaces interligne (voir Figure 2.3). Une projection verticale (resp. horizontale) est la somme des valeurs des niveaux de gris des pixels sur une ligne (resp. colonne). Ces projections représentent donc, pour une ligne donnée, l'intensité totale des pixels. Une valeur de projection est faible indique alors qu'il y a beaucoup de pixels foncés sur la ligne considérée. La méthode X-Y Cut utilise cette propriété afin de segmenter un document. L'image est dans un premier temps projetée horizontalement et découpée en bandes, là où la projection a les plus faibles valeurs. Sur chacune des bandes, une projection verticale est faite et la bande est découpée en colonnes, en suivant le même principe que précédemment. L'algorithme réitère ce processus jusqu'à ce qu'il n'y ait plus de creux dans les projections ou jusqu'à ce que les blocs aient atteint une taille inférieure à un seuil j. L'avantage de cette technique réside dans

sa facilité d'implémentation. Elle présente néanmoins plusieurs inconvénients dont notamment, le temps de calcul recquis pour sa mise en œuvre, la difficulté du choix du seuil et la sensibilité à l'inclinaison



(a) (b)

(a) histogramme horizontal, (b) histogramme vertical d'un document

Figure 2.3 : Méthode de profil de projection

II.1.3 Algorithme RLSA

Il s'agit d'une méthode itérative basée sur des opérations morphologiques de traitement d'image. Le principe de cette méthode est de noircir toute séquence de pixels blancs comprise entre deux pixels noirs de longueur inférieure à un seuil donné. En pratique l'algorithme est appliqué horizontalement et verticalement sur l'image binaire originale avec des seuils éventuellement différents pour l'horizontale et la verticale, puis une opération «ET logique » est appliquée entre les deux images lissées obtenues (voir figure 2.4). L'extraction des composantes connexes de l'image résultante permet d'obtenir les entités de la structure physique sur un niveau hiérarchique donné. On peut ainsi, en répétant la procédure avec des seuils de lissage horizontal et vertical différents, extraire itérativement les blocs de l'image, puis les lignes de texte et les mots. Ces seuils de lissage sont les seuls paramètres de l'algorithme RLSA [20]. Ils contrôlent la manière dont les composantes sont fusionnées sur un niveau de segmentation prédéterminé. Par exemple pour segmenter des lignes de texte

horizontales, droites et bien espacées, on pourra utiliser un seuil de lissage vertical nul et un seuil de lissage horizontal suffisamment grand pour combler les espaces inter lettres et inter mots. L'algorithme RLSA présente l'avantage d'être simple à mettre en œuvre et très rapide du fait qu'il ne requière qu'un nombre limite d'itérations pour atteindre la segmentation complète du document. Ses inconvénients résident dans la difficulté de réglage des seuils de lissage et sa sensible à l'inclinaison des documents De plus il n'est pas adapté pour séparer les textes des graphiques.

a : Image original b : Lissage horizontal c : Lissage vertical d : Le ET logique entre (b) et (c)

Figure 2.4 : Segmentation RLSA

Une autre méthode basée, elle aussi, sur une approche en composantes a été élaborée par 'Antonacopoulos' [21]. La principale différence est le regroupement en composants des espaces plutôt que des caractères. En effet, ce sont généralement les espaces qui séparent les différents objets d'une image. En détectant les espaces, l'approche permet d'isoler les régions, sans avoir besoin de les caractériser. Cependant, cette méthode reste sensible au bruit du fond et donne de mauvais résultats lorsque les caractères ou les lignes se touchent.

II.1.4 Méthodes utilisant l'analyse du fond de l'image

Les travaux de recherche utilisant les méthodes descendantes s'orientent vers l'analyse du fond de l'image et plus exactement les zones blanches de celle-ci. En effet, Spitz [22] a été le premier à utiliser cette approche. Le principe de cette technique est la recherche des flux blancs dans les deux directions verticale et horizontale, et leur exploitation comme délimiteur générique de structures.

Pavlidis [23] utilise une approche similaire à celle proposée par Spitz tout en présentant une amélioration au niveau de la performance de l'algorithme. Très sommairement cette technique consiste dans un premier temps à calculer le profil de projection vertical puis à rechercher la plus longue plage de valeurs d'intervalles blancs. Dans un second temps, les colonnes d'intervalle sont converties en des colonnes de blocs, tout en fusionnant les petits blocs en des blocs plus grands. Finalement, les blocs sont étiquetés en texte ou non texte en utilisant des caractéristiques.

L'approche proposée par Baird [24] est basée sur la constitution de l'ensemble des rectangles blancs appelés couverture. L'algorithme accepte en entrée un ensemble de rectangles noirs représentant les rectangles englobants des composants connexes et en sortie l'ensemble de rectangles blancs maximaux, résultant de l'union des petits rectangles blancs. Un rectangle blanc maximal est un rectangle qui ne peut pas être étendu tout en restant blanc partout. L'algorithme de l'union des rectangles blancs repose sur une heuristique et donne en sortie les rectangles blancs maximaux. L'heuristique repose sur une règle pour l'arrêt de la fusion des rectangles blancs.

II.2 Méthodes ascendantes

Les méthodes ascendantes commencent par le niveau le plus bas et remontent d'un niveau à un autre jusqu'à compléter la page. En effet, elles se basent sur l'analyse des composantes connexes. Ces dernières sont obtenues en scannant une image pixel par pixel et en regroupant les pixels en des composantes basées sur la connexité des pixels qui peut être en 4 voisins ou en 8 voisins, et en fusionnant les mots en lignes, les lignes en blocs, etc.… jusqu'à ce que la page soit complètement reconstituée.

II.2.1 Utilisation d'heuristiques

Les pixels de l'image de départ, sont regroupés en composantes connexes, ensuite des caractéristiques sont extraites sur ces composantes afin de pouvoir les regrouper en zones homogènes pour former les mots, ces derniers sont regroupés pour former des lignes et des blocs de lignes etc…..jusqu'à ce que la page soit complètement reconstitué.

La méthode de Messelodi prend en compte différents paramètres pour fusionner les composants connexes [25]. Afin d'obtenir les composantes connexes, l'image est d'abord binarisée pour ensuite regrouper les pixels en groupes connexes (voir la figure2.5). Ensuite,

en partant de l'hypothèse que les composantes connexes contiennent tout le texte mixé avec du bruit, des heuristiques permettent de trier les bonnes composantes des mauvaises. La première heuristique sert à détecter le bruit et se base sur le nombre de pixels de la composante. En effet, le bruit produisant de petites taches, il suffit de fixer un seuil au-dessous duquel la composante est classée comme bruit. D'autres critères sont utilisés, comme la densité ou le contraste, pour essayer de supprimer au maximum les composantes qui n'appartiennent pas au texte. Ensuite, les composantes sont regroupées en ligne en utilisant d'autres heuristiques telle que la distance. Cette technique présente l'avantage d'être efficace pour trouver le texte dans l'image. Son inconvénient réside dans la difficulté de réglage de seuil de binarisation.

Bien que cette méthode soit efficace pour trouver le texte dans une image, la binarisation utilisée empêche la détection de texte dans le cas de documents mal éclairés. En 2006, Wang proposa une nouvelle méthode permettant de mieux traiter la segmentation, tout en conservant cette approche basée sur des heuristiques [26]. Cette approche est intéressante pour la segmentation car elle propose de nombreuses heuristiques pour caractériser les boîtes qui englobent les zones d'intérêt.

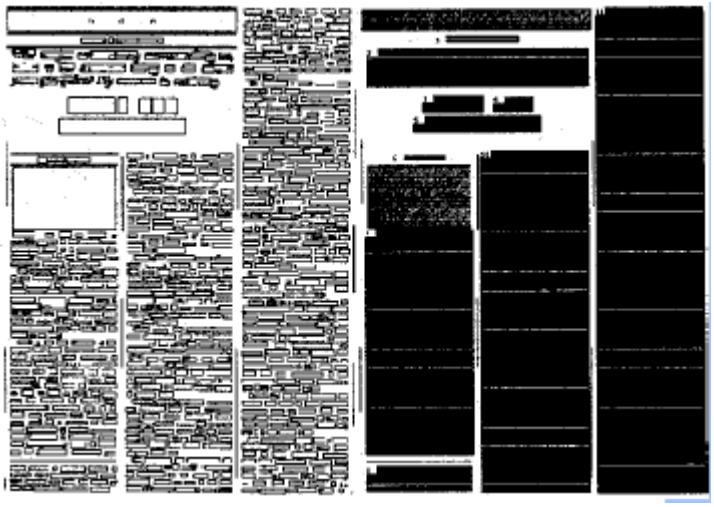


Figure2.5 : segmentation d'une page par regroupement de composantes connexes

.

II.2.2 Champs de Markov

L’approche proposée par Nicolas est une autre application des champs Markoviens au traitement d’images. Le principe de la méthode consiste à retrouver les classes d’origine des régions X en fonction de la valeur des pixels Y. Nicolas[27] propose d’utiliser un apprentissage pour déterminer la loi de probabilité de P(Y/X). Pour ce faire un vecteur de 18 caractéristiques est utilisé pour caractériser chaque site (ou pixel) en utilisant la densité dans une fenêtre entourant le pixel. Ensuite, l’algorithme EM (Espérance Maximisation) permet de définir les paramètres des gaussiennes permettant de caractériser la loi P(X/Y), grâce à des échantillons étiquetés à la main. De même, les potentiels d’interactions Vc sont pris en estimant les fréquences d’apparition de chaque couple d’étiquettes selon les différentes cliques [27]. Cette méthode est parfaitement adaptée aux documents présentant une forte variabilité aussi bien dans la mise en forme que dans la qualité, ce qui permet d’avoir de bons résultats dans le cadre de documents manuscrits. Son inconvénient réside dans l’apprentissage qui rend la méthode peu robuste dans le cas où une même étude porterait sur différents documents n’ayant pas les mêmes caractéristiques.

II.2.3 Méthodes utilisant le filtrage à base de fenêtres

Les méthodes ascendantes, utilisant le filtrage à base de fenêtres, reposent sur un balayage d'une fenêtre d'une certaine taille sur toute l'image du document. Lebourgeois [28] utilise un filtre de 8 x 3 pixels. L'image échantillonnée est dilatée par un élément de structure horizontale pour rassembler les caractères adjacents l'un vers l'autre. Il est à noter que chaque composante connexe est caractérisée par son rectangle englobant et par la moyenne des longueurs de plages de valeurs de pixels noirs. Ensuite, si la composante connexe est à l'intérieur de l'intervalle, celle-ci sera classée en une zone de texte, sinon elle sera classée en zone non texte. Les composantes connexes classées en zone texte sont fusionnées verticalement en blocs selon des règles prenant en considération l'alignement.

II. 2.4 Méthodes utilisant la technique docstrum

O'Gorman [29] introduit la technique "docstrum" qui est une technique d'analyse de structures physiques de page. Basée sur la combinaison de l'analyse ascendante et du clustering, elle fait intervenir le calcul des k plus proches voisins pour chaque composante connexe de la page. Chaque paire de voisins les plus proches possède un angle et une distance associée. En regroupant les composants à travers les caractéristiques citées précédemment, les

régions géométriques de structures physiques de la page peuvent être déterminées (voir figure2.6). La méthode proposée est indépendante du changement de l'orientation de la page mais aussi de l'espacement intertexte. Cependant, la valeur du k est dépendante de la structure de la page

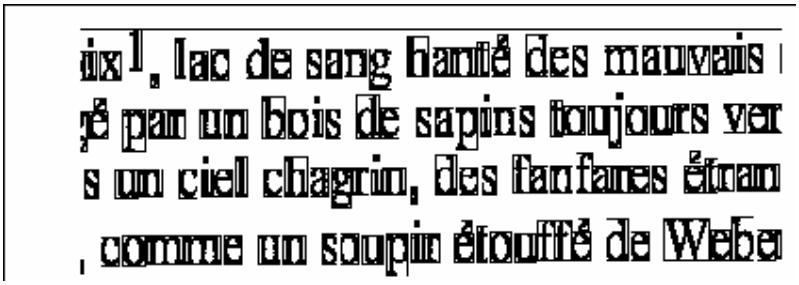


Figure 2.6 : Les composantes connexes

II.3 Méthodes mixtes

Les méthodes descendantes et ascendantes ont leurs limites. En effet, ces méthodes purement descendantes ou ascendantes donnent de bons résultats en présence de classes spécifiques de documents mais restent limitées en présence d'autres types de classes. Devant les insuffisances mutuelles de ces deux types de méthodes, des méthodes dites mixtes ont été mises au point. Ces dernières combinent les méthodes descendantes et ascendantes. Parmi les plus utilisées il y a notamment les méthodes utilisant le principe de découpage et de fusion telles celles de Azokly [30], Hadjar [31], Liu [32] et Pavlidis [33] que nous résumons succinctement ci après.

Pavlidis, utilise une approche basée sur le découpage et la fusion. Cette technique permet de distinguer les régions bitonales et non bitonales tout en permettant la séparation entre texte et graphiques. Cette séparation est réalisée en utilisant la corrélation de chaque ligne balayée dans un bloc avec la ligne de dessous. Le critère de densité de pixel noir est utilisé pour la séparation de texte et graphique.

Azokly utilise un algorithme de découpage hiérarchique, basé sur l'analyse des rectangles blancs qui constituent le fond de l'image. La fusion s'effectue à travers des règles. Celles-ci décrivent les structures à reconnaître.

La méthode développée par Liu [32] est basée elle aussi sur une approche de découpage et de fusion. Le processus de découpage repose sur la séparation en des zones non homogènes et la fusion de ces dernières en des zones homogènes. Pour ce faire, un seuil adaptatif est utilisé et permet de calculer les bordures de segmentation. C'est l'algorithme de découpage X-Y qui est utilisé comme méthode descendante. En revanche, la méthode ascendante comporte l'utilisation d'un seuil adaptatif pour calculer les bordures de segmentation.

Hadjar a élaboré une approche similaire à celle proposée par Liu. Il s'est inspiré de l'algorithme de découpage X-Y de Nagy pour le découpage de l'image du document. Ce découpage est effectué après avoir extrait les filets horizontaux et verticaux à partir d'une méthode ascendante. L'image découpée en petites régions est fusionnée pour former des régions plus grandes.

II.4 Méthodes basées sur la texture

La texture joue un rôle très important dans l'analyse d'images. En effet Les principales informations dans l'interprétation du message visuel pour un observateur humain sont les contours et les textures.

Les approches « textures » considèrent le texte et le graphique comme étant deux textures différentes, le problème sera alors de trouver les bonnes caractéristiques qui permettent de bien les séparer.

Ces approches texture sont souvent utilisées pour ajouter de l'information aux techniques ascendantes [34], l'approche par segmentation utilisant la texture regroupe beaucoup de techniques différentes. Le but de ces dernières est de trouver les caractéristiques de texture qui sont propres au texte. De nombreuses techniques sont alors utilisées pour transformer l'image en une représentation mettant en avant ces caractéristiques, les techniques les plus utilisées seront présentés dans la section suivante.

II.4.1 Transformé de Fourier

La transformée de Fourier est une fonction qui transforme une fonction (dans notre cas l'image) en une autre fonction décrivant son spectre de fréquence. Dans le cas d'une image, la transformée sera donc la représentation fréquentielle de l'image. Au centre est concentré l'énergie de basse fréquence (qui correspond à des transitions douces dans l'image,), autour les fréquences moyennes et aux bords les hautes fréquences. Grâce à cette transformation, il

est également possible d'avoir l'orientation des contours principaux de l'image, ce qui peut être directement interprétable pour caractériser le texte. Le calcul de cette transformée dans le domaine discret (donc adapté aux images) est le plus souvent réalisé avec l'algorithme FFT (« Fast Fourier Transform »). L'intérêt de la FFT pour la segmentation est qu'une variante de cette transformée, la DCT (« Discrete Cosine Transform »), est au centre de la compression JPEG et peut être utilisée pour retrouver des zones de texte [35]. En effet, la compression JPEG découpe l'image en blocs de 8 × 8 pixels, ce qui rend locale la transformée DCT. Les blocs dont les coefficients de la transformée DCT présentent une forte intensité verticale (rotation de 90°) sont jugés comme étant potentiellement du texte. Cependant, si cette technique apporte d'un côté une rapidité de calcul, d'un autre côté, elle souffre d'une trop forte sensibilité à la taille des caractères.

II.4.2 Filtres dérivateurs

Un peu à la manière de la transformée de Fourier, les filtres dérivateurs permettent d'avoir une meilleure représentation fréquentielle de l'image. L'avantage est que la transformée contient aussi des informations spatiales. En pratique, la dérivée d'une image est obtenue en convoluant l'image avec un noyau dérivateur. Il existe de nombreux noyaux dérivateurs, et de leurs définitions dépendent les caractéristiques du filtre. En revanche, tous ont la propriété d'avoir la somme des coefficients nulle. Ainsi, il est possible de définir des filtres ayant une plus grande sensibilité aux contours horizontaux, verticaux et diagonaux [36]. Cet aspect, permet de repérer le texte en convoluant l'image avec ces noyaux pour extraire de l'image les formes attendues et repérer ainsi plus facilement le texte. Ces filtres sont sensibles, de la même manière que pour toute analyse par texture, à l'échelle du texte. Pour éviter cet inconvénient, de nombreuses techniques utilisent une approche multi résolution. C'est le cas de la méthode proposé par Wu qui utilise trois filtres dérivateurs à différentes échelles [37]. Tout les filtres sont basés sur la dérivée seconde d'une gaussienne, avec un écart type qui varie, ce qui a pour effet de réaliser une analyse multi-résolution. En effet, un filtrage par une gaussienne revient à réduire l'échelle de l'image, en fonction de l'écart type. Suite à ces filtrages, chaque pixel est associé à un vecteur de neuf dimensions. En utilisant un algorithme de classification classique (les K-moyennes, avec K= 3), il regroupe les pixels entre eux afin d'obtenir trois classes (texte, arrière plan et intermédiaire), pour ensuite réaliser une transformation de morphologie mathématique (dilatation). Cependant, la segmentation n'est pas suffisamment précise pour permettre de s'arrêter là. Les régions

détectées vont servir pour identifier les zones d'intérêt et une étude plus approfondie sera réalisée pour ajuster les frontières. Une étude ascendante est alors réalisée, en partant d'une détection de contours, car les caractères forment généralement des contours bien marqués avec le fond. En prenant en compte toutes ces informations, la méthode conduit à de bons résultats en atteignant un très bon taux de reconnaissance.

Une autre sorte de filtres dérivateurs utilisés sont les filtres de Gabor. Ces derniers sont proposés par Daugman. Ces filtres ont la particularité d'avoir trois paramètres permettant de rendre plus sensible le filtre à certains types de variations.

Datong propose en 2001 de paramétrer ces filtres en réalisant une première étude sur l'image pour détecter les contours, et ensuite d'estimer l'orientation et la taille du texte [38].

II.4.3 Auto corrélation

Nous avons vu que les filtres dérivateurs permettaient de mettre en évidence les contours des textures et permettent ainsi de repérer plus facilement les zones de texte. Il existe une autre transformation, plus coûteuse en calculs, mais donnant une meilleure information sur les orientations générales et les périodicités de la texture. L'autocorrélation, en combinant l'image avec elle-même après une translation, permet de mettre en avant ces informations. Dans le domaine spatial (en pratique, le calcul est réalisé dans le domaine fréquentiel avec la FFT), le calcul de l'autocorrélation se fait en utilisant la formule suivante :

$$C_{xx}(k,l) = \sum_{k'=-\infty}^{+\infty} \sum_{l'=-\infty}^{+\infty} x(k',l').x(k'+k,l'+l) \qquad \textbf{(II.1)}$$

Cette transformation permet alors de créer une rose des directions qui représente les orientations principales de l'image (voir figure 2.7). En découpant l'image d'origine en petits blocs, puis en calculant cette rose sur chacun des blocs, il est possible de déterminer l'organisation de l'image [39]. La rose est en fait un diagramme polaire. Soit (u, v) le point central de l'autocorrélation et θ_i l'orientation étudiée, on calcule alors la droite Di tel que l'ensemble de ses points (a, b) forme un angle θ_i avec l'abscisse. Pour chaque orientation θ_i on calcule ainsi la somme des différentes valeurs de la fonction d'auto corrélation en utilisant la formule suivante :

$$R(\theta_i) = \sum_{D_i} C_{xx}(a,b) \qquad \textbf{(II.2)}$$

Cependant, cette technique présente l'avantage dans la possibilité de connaitre les différentes directions de variation du texte grâce au calcul de la rose. L'inconvénient de cette technique réside dans la difficulté de choisir la taille de la fenêtre d'analyse et dans le temps de calcul élevé.

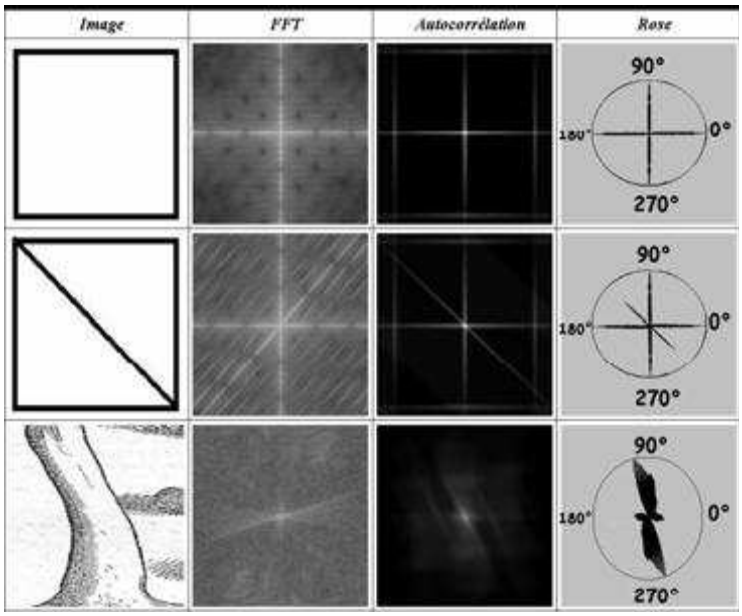


Figure 2.7: Trois transformées d'images sur trois images : FFT, Auto corrélation et Rose des directions.

III. Discussion

Dans ce chapitre, nous avons présenté l'analyse de la structure physique des documents basée sur la segmentation. En effet, la segmentation physique est une étape clé dans la chaîne de numérisation qui a pour but de trouver les différentes classes présentes dans un document.

Il existe plusieurs méthodes de segmentation, nous avons cité les méthodes ascendantes, les méthodes descendantes, les méthodes mixtes et les méthodes utilisant la texture. Les techniques descendantes sont très rapides lorsque la structure du document est connue au préalable. Par contre les techniques ascendantes sont plus robustes et plus exactes mais plus lentes.

Parmi les problèmes de l'analyse de la structure physique, nous citons :

1. L'estimation des seuils pour les deux types de méthodes (ascendantes et descendantes) car elle est nécessaire pour obtenir une segmentation plus fiable.
2. L'estimation de l'angle d'inclinaison est nécessaire pour certaines méthodes de segmentation.

Tenant compte de ces deux principaux problèmes nous avons élaboré une méthode de segmentation de document basée sur les matrices de cooccurrence et la diffusion anisotrope que nous présentons dans le prochain chapitre.

Chapitre III

Segmentation d'images de documents par analyse de texture en niveaux de gris

I. Préambule

Dans le chapitre deux, nous avons passé en revue les différentes techniques de segmentation. L'étude de ces techniques nous a permis de mettre en évidence leurs inconvénients notamment en ce qui concerne le choix des seuils et la sensibilité à l'angle d'inclinaison.

Pour palier ces inconvénients, nous nous sommes orientés vers une méthode basée sur la texture. Ce type de méthodes s'avère plus robuste en ce qui concerne le traitement des documents composites. Dans le présent chapitre nous allons présenter les principales méthodes de caractérisation de la texture des documents ; nous allons détailler en particulier la méthode pour laquelle nous avons optée.

II. Définition de la texture

Le but de l'analyse de texture est de formaliser les descripteurs de la texture par des paramètres mathématiques qui serviraient à l'identifier.

Il n'existe pas de définition précise de la texture. Une définition générale peut caractériser une texture comme un ensemble de primitives arrangées selon des règles particulières de placement.

Coquerez [40] présente deux approches dans la définition de la texture: une première approche est déterministe et correspond à une vision macroscopique de la texture et la considère comme « une répétition spatiale d'un motif de base dans différentes directions ». La deuxième approche est probabiliste et correspond à une vision microscopique, elle cherche à caractériser « l'aspect anarchique et homogène qui ne comprend ni de motif localisable, ni de fréquence de répétition principale »

Gagalowicz [41] propose une synthèse des deux approches en considérant la texture comme "une structure spatiale constituée de l'organisation de primitives (ou motifs de base) ayant chacune un aspect aléatoire ».

On peut distinguer deux grandes classes de textures, qui correspondent à deux niveaux de perception [42] :

A- **Les textures périodiques** qui présentent un aspect régulier, sous formes de motifs répétitifs spatialement placés selon une règle précise (ex: peau de lézard, mur de brique) donc une approche structurelle déterministe. (Figure 3.1)

Figure 3.1 : Exemple des textures périodiques.

B- **les textures aléatoires** présentant des primitives distribuées de manière aléatoire (ex: sable, laine tissée, herbe) (figure3.2) d'ou une approche probabiliste cherchant à caractériser l'aspect anarchique et homogène.

Les textures naturelles sont aléatoires et totalement désordonnées et il est difficile d'isoler un motif de base qui se répète.

Figure3.2 : exemple des textures aléatoires.

III. Texture et niveau de gris

Sur des images à niveaux de gris (voir figure3.3), l'information fournie par la texture est rapidement indispensable pour distinguer les différentes régions d'une image.

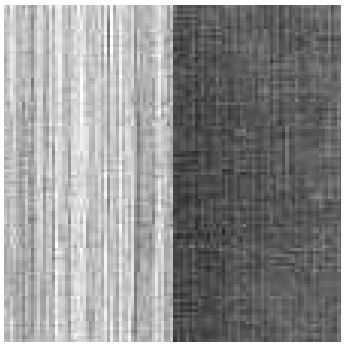


Figure 3.3 : Les niveaux de gris discriminent les régions

Dans l'image 3.3, on parlera de la partie claire et de la partie plus sombre.

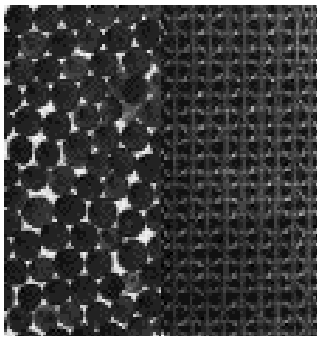


Figure 3.4 : La texture discrimine les régions.

Dans l'image 3.4, on évoquera plutôt les aspects structuraux de chaque partie, ce qui permet alors de considérer la texture comme l'information qui établit une distinction entre ces régions.

IV. Caractérisation d'une texture

Dans une image, la texture est un paramètre très important pour la compréhension et l'interprétation d'une scène. C'est ainsi qu'elle est prise en compte dans plusieurs méthodes d'analyse d'images. Si les méthodes permettant la mesure de textures sont nombreuses, aucune ne peut, aujourd'hui, prétendre généraliser un modèle de texture. En effet, de part sa complexité intrinsèque, l'on n'a pu trouver une définition formelle de ce qu'est la texture. On se contente donc de trouver un modèle adéquat pour l'étude à mener. C'est ainsi une multitude de méthodes et de combinaisons de méthodes ont déjà été proposées dans la littérature et

éprouvées en pratique [43], parmi ces méthodes nous citons les méthodes structurelles et les méthodes statistiques.

IV.1 Méthodes structurelles

Elles tiennent compte de l'information structurelle et contextuelle d'une forme et elles sont particulièrement bien adaptées aux textures périodiques. Les étapes d'analyse sont d'abord l'identification des éléments constitutifs, puis la définition des règles de placement.

Les méthodes structurelles sont généralement peu intéressantes, dans la mesure où elles imposent de travailler sur des textures extrêmement régulières, ce qui n'est pas notre cas (les images de documents possèdent des textures irrégulières).

IV.2 Méthodes statistiques

Elles étudient les relations entre un pixel et ses voisins et définissent des paramètres de la texture en se basant sur des outils statistiques. On distingue les méthodes de premier ordre et les méthodes de second ordre. Comme notre projet concerne la segmentation des documents qui possèdent des formes et structures irrégulières, nous allons nous intéresser dans la section suivante aux méthodes statistiques.

Parmi ces méthodes nous citons les méthodes de premier ordre et les méthodes de deuxième ordre.

IV.2.1 Méthodes statistiques de premier ordre

Les méthodes de premier ordre considèrent l'image comme un processus aléatoire discret et donnent une idée sur l'histogramme. Les attributs les plus utilisés sont la moyenne, la variance, le « skewness », le « kurtosis » et l'auto covariance [40].

IV.2.2 Méthodes statistiques de second ordre

Pour effectuer une analyse large, on fait appel aux méthodes statistiques de second ordre, parmi ces méthodes, celles qui exploitent directement les propriétés statistiques de la texture (les matrices de cooccurrence, appelée aussi « matrice de dépendance spatiale des niveaux de gris », matrice de longueurs de plages, fonction d'auto corrélation, modèle de

Markov, modèles issues de la morphologie mathématique..), soit dans des méthodes qui exploitent les propriétés statistiques a partir d'un plan transformé dans lequel on réécrit l'image de texture (densité spectrale, méthode des extrémas locaux, méthodes de transformation de Fourier, filtres numériques…).

V. Analyse texturale par cooccurrence

La procédure utilisée par Haralick et al. [1973] pour obtenir l'information texturale d'une image est basée sur l'hypothèse suivante : "l'information de texture dans une image est contenue dans les relations spatiales entre les niveaux de gris". Ces relations sont alors représentées numériquement à l'aide des matrices de dépendance spatiale de niveaux de gris, également appelées matrices de cooccurrence. Celles-ci sont longtemps restées inutilisées du fait de leur coût prohibitif en temps de calcul. Mais avec les avancées considérables en informatique et les études menées pour optimiser leur temps de calcul, elles constituent aujourd'hui l'outil de caractérisation texturale le plus populaire. Selon Germain [1997] l'engouement montré par cet outil est sans doute dû au fait qu'il est fondé sur des statistiques d'ordre deux et donc bien adapté à la description des propriétés texturales auxquelles l'œil humain est le plus sensible.

De nombreuses études ont montré l'intérêt des matrices de cooccurrence en traitement d'image, il apparaît donc intéressant de tester cet outil pour la reconnaissance des textures particulières auxquelles nous nous intéressons ici.

V.1 Méthode de la dépendance des niveaux de gris

Les matrices de cooccurrences déterminent la fréquence d'apparition d'un "motif" formé de deux pixels séparés par une certaine distance d dans une direction particulière par rapport à l'horizontale. Afin de limiter le nombre de calculs, on prend généralement comme valeurs 0°, 45°, 90°, 135°, 180° (voir figure3.5) et 1 pour la valeur de d. Les matrices de cooccurrences constituent un outil d'analyse d'images en niveaux de gris et rendent compte des transitions de niveaux de gris dans l'image. La méthode de matrice de cooccurrence est largement utilisée dans le monde du traitement d'image (Haralick *et al*, 1973). Elle présente une grande simplicité de mise en oeuvre et donne de bons résultats sur la plupart des types d'images, c'est ce qui justifie notre choix de la confronter. Ces matrices contiennent une masse très

importante d'informations difficilement manipulable. C'est pour cela qu'elle n'est pas utilisée directement mais à travers des mesures dites indices de texture.

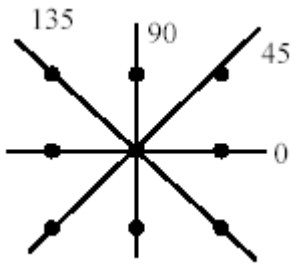


Figure 3.5 : Les directions de matrice de cooccurrence.

Pour une translation t, la matrice de cooccurrence MC_t d'une région R est définie pour tout couple de niveau de gris(i,j) par :

$$MC_t=\text{card.}\{(s,s+t)\,R^2 \backslash I(s)=i, I(s+t)=j\} \quad \textbf{(III.1)}$$

MC_t (i, j) est donc le nombre de couples de sites (s, s+t) de la région considérée, séparés par le vecteur de translation t et tel que s a pour niveau de gris i et s+t pour niveau de gris j. Pour une image I , quantifiée sur Ng niveau de gris, la matrice MC_t est une matrice de Ng*Ng.

On peut aussi travailler à partir des matrices de cooccurrence dont les éléments représentent l'estimation de la probabilité de transition d'un niveau de gris à un autre selon une direction donnée avec une distance inter pixels bien définie selon l'équation suivante :

$$P_{\theta,d}(i,j)=\frac{MC_t(i,j)}{N} \quad \textbf{(III.2)}$$

Avec

$$N=\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} MC_t(i,j) \quad \textbf{(III.3)}$$

La figure (3.6) illustre un exemple de calcul des matrices de cooccurrence.

-Pour d=1 et θ=0°

3	5	5
1	3	2
6	5	1
5	3	6

M (i, j)=

0°	1	2	3	4	5	6
1	0	0	1	0	0	0
2	0	0	0	0	0	0
3	0	1	0	0	1	1
4	0	0	0	0	0	0
5	1	0	1	0	1	0
6	0	0	0	0	1	0

(a) Image originale f(x, y) (b) Matrice de cooccurrence de l'image originale M (i, j)

-Pour d=1 et θ=45°

M (i, j)=

45°	1	2	3	4	5	6
1	0	0	0	0	1	0
2	0	0	0	0	0	0
3	1	0	1	0	0	0
4	0	0	0	0	0	0
5	0	1	0	0	0	1
6	0	0	1	0	0	0

-Pour d=1 et θ=90°

M (i, j)=

90°	1	2	3	4	5	6
1	0	0	0	0	0	2
2	1	0	0	0	0	0
3	1	0	0	0	1	0
4	0	0	0	0	0	0
5	0	1	2	0	0	0
6	0	0	0	0	1	0

-Pour d=1 et θ=135°

M (i, j)=

135°	1	2	3	4	5	6
1	0	0	1	0	0	0
2	0	0	0	0	1	0
3	0	0	0	0	0	1
4	0	0	0	0	0	0
5	1	0	1	0	1	0
6	0	0	0	0	0	0

Figure 3.6 : calcul des matrices de cooccurrence

V.2 Attributs extraits à partir des matrices de cooccurrence

A partir des matrices de cooccurrence, nous pouvons tirer un certain nombre de renseignements sur la texture de la région considérée. Si la texture est grossière et si d est petit par rapport aux éléments de texture répétitifs, alors les couples de pixels séparés par (d,θ) auront des niveaux de gris voisins. Par contre, pour une texture plus fine, si d est comparable à la taille des éléments de la texture, alors les couples de pixels séparés par (d,θ) auront souvent des niveaux de gris différents, cela se traduit par une dispersion des valeurs de la matrice de cooccurrence.

À partir des matrices de cooccurrence, Haralick et al. [1973] ont défini 14 indices pour caractériser les textures. Les indices sont rappelés en annexe A.

VI. Méthode adoptée pour la de segmentation de documents

L'approche utilisée dans ce travail a pour but de séparer les différentes régions image, texte et fond. Nous avons choisi une méthode de segmentation basée sur la texture. Des prétraitements sont nécessaires pour accentuer les contours et éliminer le bruit qui influe sur les résultats de segmentation. L'un de ces prétraitements est l'application de la diffusion anisotrope. L'architecture générale du système est montrée par la figure (3.7)

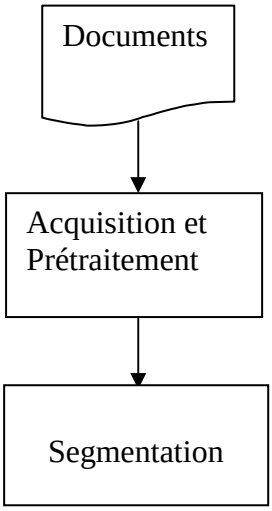


Figure 3.7 : Architecture du système de segmentation de document

La figure (3.8) illustre les différentes étapes de notre système.

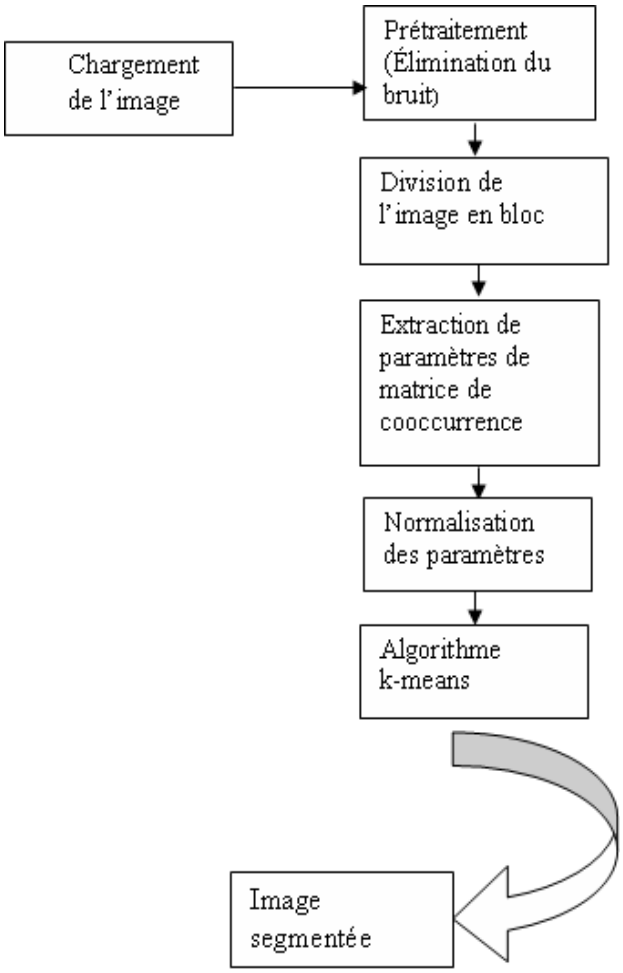


Figure 3.8 : Les étapes de l'algorithme

VI.1 Prétraitement

Avant d'arriver à la phase de segmentation, quelques opérations de prétraitement qui servent à l'élimination du bruit, la transformation de l'image en niveau de gris et le rehaussement de l'image, sont nécessaires.

Nous avons utilisé un filtrage basé sur la diffusion anisotrope pour éliminer et rehausser les contours.

La diffusion permet d'homogénéiser une image de même que la diffusion de température en physique qui sert à homogénéiser la température des objets. Utilisée en traitement d'image la diffusion intervient en prétraitement de façon à supprimer les perturbations locales du signal. Il est alors possible dans un second temps d'effectuer une recherche des contours.

- **Diffusion isotrope**

La diffusion isotrope est un processus itératif qui revient à implanter une approximation discrète de l'équation de diffusion isotrope de la chaleur:

$$\frac{\partial I}{\partial t} = c\Delta I \qquad \textbf{(III.4)}$$

Δ Désignant l'opérateur Laplacien, I le signal, t le temps, et c le cœfficient de diffusion.
Cela revient à faire directement la convolution de l'image avec une gaussienne. En particulier, cette équation a l'inconvénient de rendre les contours de plus en plus flous au cours des différentes itérations.

- **Diffusion anisotrope**

Le problème de la diffusion isotrope est le lissage de toute l'image. Bien que le lieu des contours soit conservé, ces derniers deviennent flous, comme l'illustre la figure (3.9) (en premier, l'image originale bruitée et en second celle qui a été traitée).

Pour éviter d'effacer progressivement les contours, les auteurs ont proposé de modifier l'équation de diffusion isotrope pour faire de la diffusion anisotrope, sur les principes suivants :

- Diffusion et donc homogénéisation maximale loin des contours.
- Diffusion minimale au niveau des contours.

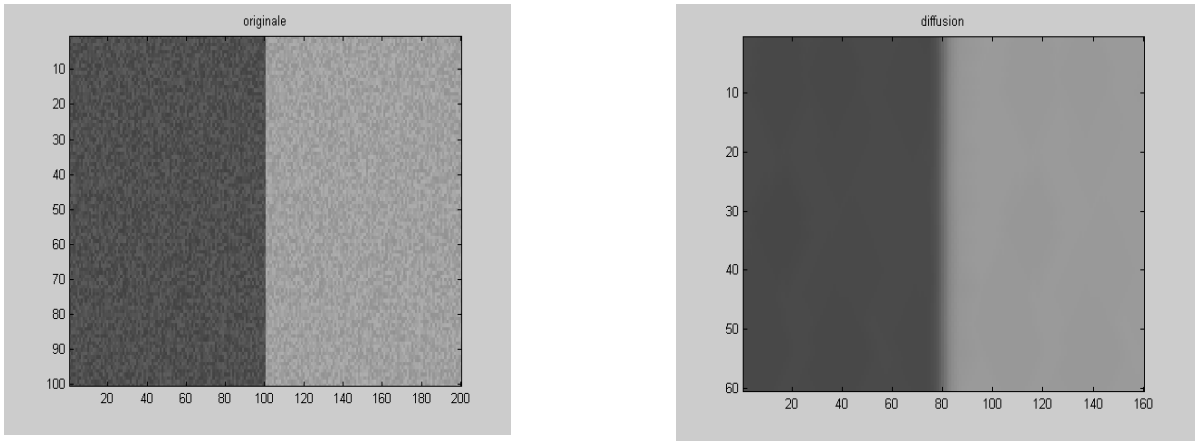


Image originale Image traitée

Figure 3.9 : diffusion isotrope

Il faudrait donc limiter ou interdire la diffusion dans les zones de l'image représentant des contours. Pour cela, le coefficient C contrôlant la diffusion doit varier en fonction de la position dans l'image. On parle alors de diffusion anisotrope.

Dés lors que les cœfficients ne sont plus constants, mais dépendent de la position et du temps, le filtre perd son caractère isotrope et il revient à implanter une approximation discrète d'un processus de diffusion anisotrope, comme l'ont proposé Perona et Malik [44] [45]:

$$\frac{\partial I}{\partial t} = div(c(x,y;t)\nabla I) = c(x,y;t)\Delta I + \nabla c \nabla I \qquad \textbf{(III.5)}$$

∇ Désignant l'opérateur Gradient.

Les filtres basés sur le principe de diffusion anisotrope visent à lisser d'autant plus qu'il s'agissent d'une même région, et de diminuer, voire stopper le lissage lorsqu'on se situe sur

→

une discontinuité importante, relative à un bord significatif. Pour cela, ils consistent à effectuer de façon itérative, une convolution sur un masque (3 *3), où chaque coefficient c^t est une valeur mesurant la continuité du signal en chaque point, à l'aide d'une fonction décroissante $f(d^t(x, y))$ telle que $f(0) = 1$ et $f(d^t(x, y))$ 0 quand $d^t(x, y)$ augmente, $d^t(x, y)$ étant une mesure de la discontinuité du signal au point (x, y). De façon courante, les filtres de diffusion anisotrope utilisent une fonction exponentielle décroissante pour la fonction *f*, et l'amplitude du gradient comme mesure de la discontinuité en chaque point, et le coefficient $c^t(x, y)$ est souvent exprimé d'une manière générale par:

$$c^t(x, y) = e^{-\alpha|\nabla I|^2} \qquad \textbf{(III.6)}$$

Voir ce que donne la diffusion anisotrope sur une image en niveaux de gris sur la figure3.10.

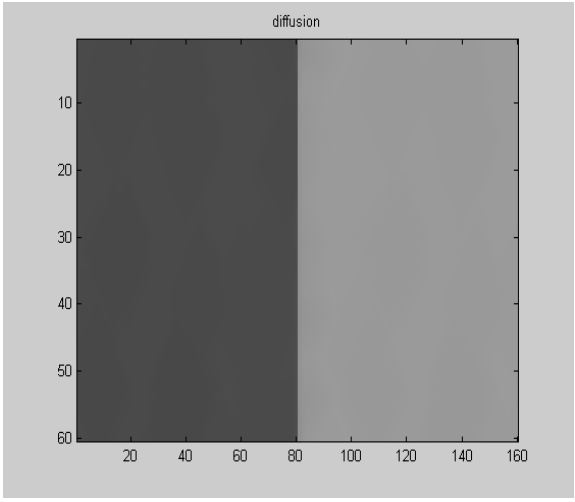


Figure 3.10 : application de la diffusion anisotrope

VI.2 Division en bloc

L'image document est divisée en blocs. On définit une image Img de la manière suivante [46]:

$$Img = \{ P_{ij}\ 0 \leq\ i < H, 0 \leq j < W\} \qquad \textbf{(III.7)}$$

p_{ij} est le pixel de position i, j ; H et W sont respectivement la longueur et la largeur de l'image. La division de l'image en blocs est obtenue comme suit :

$$Img = \{b_{I\,J}, 0 \leq I < \frac{H}{h}, 0 \leq J < \frac{W}{w}\} \qquad \textbf{(III.8)}$$

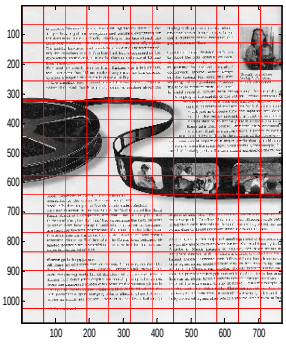
$b_{I\,J}$ représente le bloc de la $I^{ème}$ ligne et de la $J^{ème}$ colonne ; h et w sont respectivement la longueur et la largeur des blocs.

Cette étape présente les avantages suivants :

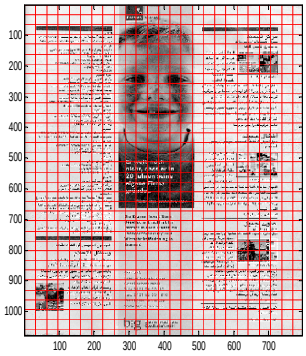
- Segmentation simple
- Pas de redondance d'information
- Moins coûteuse en temps de calcul

La taille du bloc ne doit pas être très grande car il y aura un risque d'avoir plusieurs classe dans le même bloc, ni trop petite pour éviter de ne pas avoir assez d'informations pour le classer, sans oublier le temps de calcul qui sera plus élevé. Par conséquent il faut faire un choix qui optimise les résultas (figure3.11).

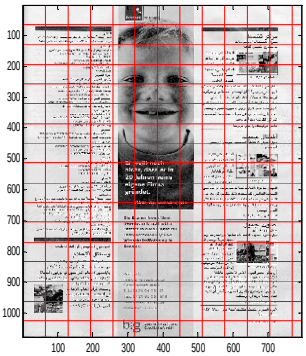
Image1



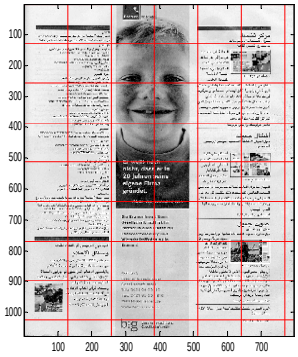
division en bloc de l'image1



(a)



(b)



(c)

(d)

Figure 3.11 : division de l'image en différentes tailles du bloc
(a) 32*32 (b) 64*64 (c) 128*128 (d) 256*256

VI.3 Extraction de paramètres

Les matrices de cooccurrence permettent d'estimer les propriétés des images relatives à des statistiques de second ordre. Cette approche est la plus utilisée pour extraire les caractéristiques de texture [47] [48] [49]. Pour chaque bloc de l'image, cinq paramètres de texture sont calculés dans quatre directions $\{0^{\circ}, 90^{\circ}, 180^{\circ}, 270^{\circ}\}$, ces paramètres sont l'énergie (ENR), l'entropie (ENT), la somme Entropie (SEN), la différence Entropie (DEN) et l'écart type. Leur définition mathématique est donnée par les équations (9 à 13).

$$\mathrm{ENR} = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} P^{2}_{\mathrm{d}}(\mathrm{i}, \mathrm{j}) \qquad \textbf{(III.9)}$$

$$\mathrm{ENT} = -\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} P^{2}_{\mathrm{d}}(\mathrm{i}, \mathrm{j}) \log_2 P_d(\mathrm{i}, \mathrm{j}) \qquad \textbf{(III.10)}$$

$$\mathrm{SEN} = -\sum_{k=0}^{2n-2} P_{x+y}(\mathrm{k}) \log_2 P_{x+y}(\mathrm{k}) \qquad \textbf{(III.11)}$$

$$DEN = -\sum_{k=0}^{n-1} P_{x-y}(k) \log_2 P_{x-y}(k) \quad \textbf{(III.12)}$$

$$STD = \sqrt{\frac{\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}(P_d(i,j)-\mu)^2}{n \times n}} \quad \textbf{(III.13)}$$

Où

$$\mu = \frac{\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} P_d(i,j)}{n \times n} \quad \textbf{(III.14)}$$

$$P_{x+y}(k) = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} P_d(i,j) \quad \textbf{(III.15)}$$

Pour i + j = k, k = 0,1…2n -2

$$P_{x-y}(k) = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} P_d(i,j) \quad \textbf{(III.16)}$$

Pour $|i-j| = k$, k = 0,1…n-1

VI.4 Normalisation des paramètres

Dans ce travail nous avons extrait cinq paramètres de texture définis par les équations 9 à 13 qui permettent de les calculer. Néanmoins, ces calculs peuvent produire diverses gammes de valeurs contenues dans des intervalles différents ce qui fausse la classification d'où l'intérêt de la normalisation pour cadrer les valeurs des différents paramètres dans des intervalles proches. Pour ce faire, nous utilisons la normalisation statique des paramètres [50] [51].

Cette dernière transforme chaque distribution de paramètres en une distribution ayant une moyenne qui vaut '0' et une variance qui vaut '1'. Pour normaliser les valeurs des paramètres, nous opérons comme suit :

Soit p le nombre de paramètres et m la taille de la distribution, la matrice des paramètres Z est définie comme suit :

$$Z = \begin{bmatrix} z_{11} & . & . & . & z_{1p} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ z_{m1} & . & . & . & z_{mp} \end{bmatrix} \qquad \textbf{(III.17)}$$

Où

Z_{ij} est le $j^{ème}$ paramètre du $i^{ème}$ bloc pour

i = 1,2,…m et j = 1,2,…p.

Les nouvelles valeurs des paramètres sont calculées par la relation (18)

$$Z'_{ij} = \frac{(Z_{ij} - \overline{Z_j})}{\sigma_j} \qquad \textbf{(III.18)}$$

Où $\overline{Z_j}$ est la moyenne définie en équation (III.19) et l'écart- type définie en équation (III.20).

$$\overline{Z_j} = \frac{1}{m}\sum_{i=1}^{m} Z_{ij} \qquad \textbf{(III.19)}$$

$$\sigma_j = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(Z_{ij} - \overline{Z_j})^2} \qquad \textbf{(III.20)}$$

VI.5 Classification des blocs par l'algorithme des k-means

Pour mettre en œuvre l'algorithme des k-means [52] [53] [54], nous avons fixé le nombre de classes à trois (k=3) ces dernières représentent le texte, les images et le fond. Pour affecter un bloc à une classe, chaque bloc de l'image est comparé à la valeur moyenne de chaque classe calculée préalablement. Cette comparaison est réalisée en minimisant la distance euclidienne entre les vecteurs paramètres du bloc considéré et ceux des centres de classe. A chaque itération, cet algorithme recalcule le centre des classes. Ce processus sera répété jusqu'à ce que la valeur des centres des classes ne change pas.

Le résultat de classification de l'image1 est donné dans la figure (3.12)

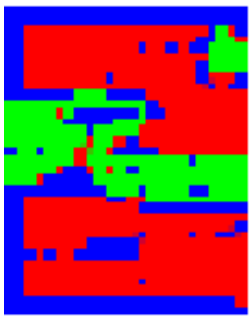


Figure 3.12 : Image segmentée

VII. Discussion

Dans ce chapitre, nous avons présenté notre approche pour la conception d'une méthode de segmentation de document basée sur les matrices de cooccurrence et la diffusion anisotrope.

Pour avoir une bonne idée sur les performances de notre système, des tests ont été effectués, ils seront présentés dans le prochain chapitre.

Chapitre IV

Test et résultats

I. Préambule

Afin d'évaluer les performances de notre méthode, nous présentons dans ce chapitre les différents résultats obtenus en appliquant la technique de segmentation développée dans le chapitre précédent à différents types de documents et nous évaluerons le taux d'erreur de classification.

II. Présentation des données

Pour pouvoir effectuer nos tests, nous avons utilisé des documents latin et arabes de type article. Ces documents ont été scannés à l'aide d'un scanner 'Epson Expression 1600 Pro' avec une résolution de 100 dpi. Les images ont été sauvegardées sous le format bmp. D'autres documents de type journaux ont été scannés à l'aide d'un scanner 'Hewlett Packard ScanJet' avec une résolution de 400 dpi. Les images ont été sauvegardées sous le format TIFF.

L'ensemble de ces documents est caractérisé par des variations de présentation du contenu. Nous avons utilisé treize images scindées en deux types de présentation qui sont :

- Quatre images de tailles différentes représentant des documents composés d'un texte inclus dans les images (voir figure4.1)
- Neuf images de tailles différentes représentant des documents composés d'images inclus dans le texte (figure4.2).



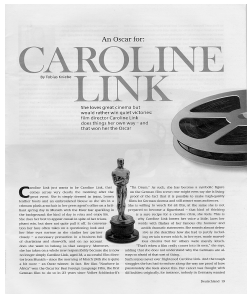

768*1074 pixels 784*1067 pixels 804*1067 pixels 763*1063 pixels

(Incliné de –15°)

Figure4.1 : documents composés d'un texte inclus dans les images

768*1074 pixels

850*1165 pixels



850*1165 pixels

820*1074 pixels

816*1074 pixels

750*1045 pixels

(Inclinée de -15°)

788*1063 pixels

802*1067 pixels

735*1002 pixels

(Inclinée de +30°)

Figure4.2 : documents composés d'images inclus dans le texte

III. Démarche d'expérimentation

Nous avons testé notre système de segmentation sur les documents que nous avons présentés, nous avons effectué une série de tests afin d'évaluer les performances de notre système. Pour ce faire, des tests ont été effectués en faisant varier dans un premier temps le nombre de paramètres texturaux puis la taille de la fenêtre d'analyse.

III.1 choix de paramètres texturaux

La texture joue un rôle très important dans l'analyse d'images. En effet, es principales informations dans l'interprétation du message visuel pour un observateur humain sont les contours et les textures.

Les approches textures considèrent le texte et l'image comme étant deux textures différentes, le problème sera alors de trouver les bonnes caractéristiques qui permettent de bien séparer le texte du fond.

Afin de trouver les bons paramètres caractérisant le texte et l'image, nous avons sélectionné plusieurs caractéristiques préconisées dans la littérature pour analyser la texture et nous avons étudié leur pertinence sur les documents imprimés. Pour cela nous avons effectué plusieurs tests en utilisant les méthodes statistiques de premier ordre (Ecart-type), nous avons utilisé également les méthodes statistiques de second ordre (Entropie, Energie, Différence entropie et Somme entropie).

III.1.1 Cas d'utilisation d'un seul paramètre de texture

Les premiers tests que nous avons effectués sur les documents consistent à utiliser un seul paramètre de texture pour les segmenter et calculer le taux d'erreur de classification. Les résultats de segmentation de l'image1 en utilisant un seul paramètre à la fois sont illustrés par la figure 4.3 et le taux d'erreur associé à chaque résultat est donné par le tableau 4.1.

Notons que pour les tests, L'image segmentée est divisée en M blocs puis chaque bloc est relevé comme étant texte ou non texte. Puis l'image originale est à son tour divisée de la même manière. Ainsi, nous pouvons comparer chaque bloc de l'image segmentée à son équivalent dans l'image originale si le bloc relevé comme texte dans l'image segmentée correspond à un bloc texte donc l'image originale, le bloc est bien classifié sinon il est mal classifié. Le pourcentage d'erreur a été calculé en utilisant la relation suivante :

$$P_e = \frac{\textbf{\textit{nombre de blocs mal classifiés}}}{\textbf{\textit{nombre total de blocs dans l'image}}} * \textbf{\textit{100\%}}$$

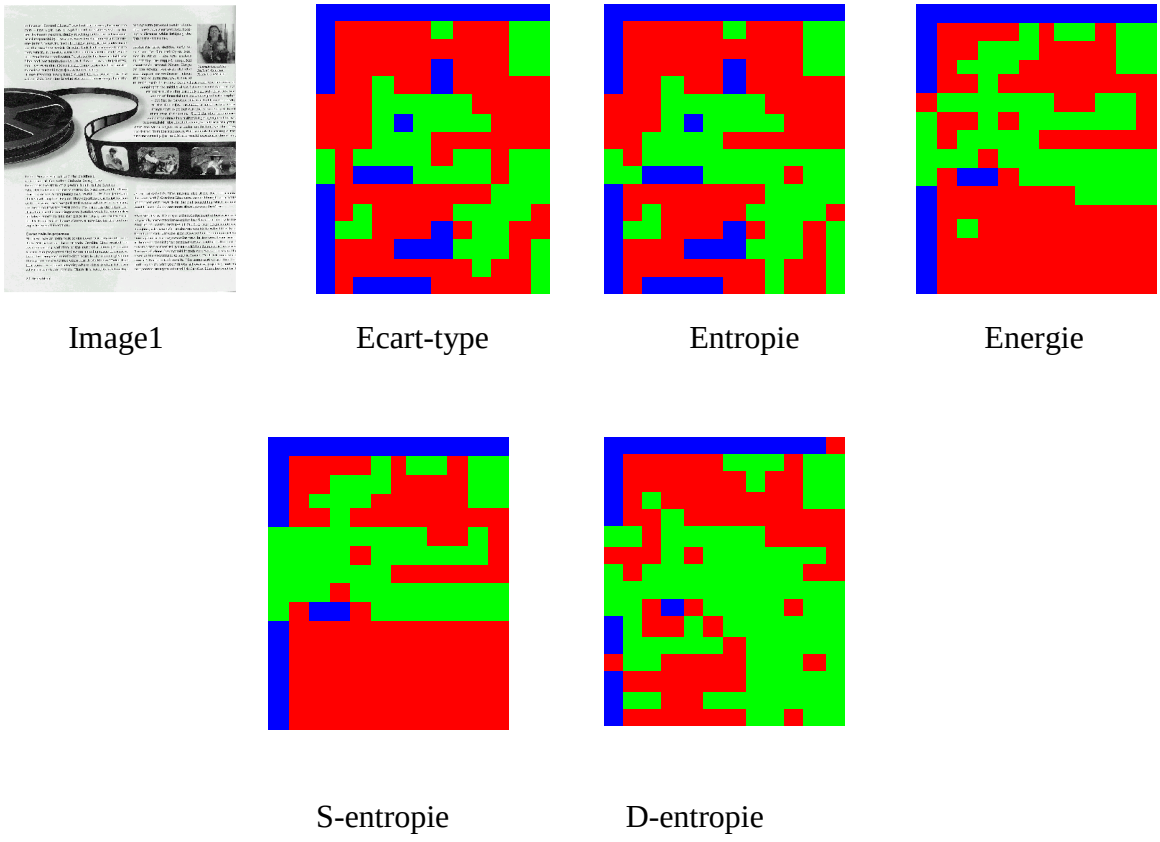


Figure 4.3 : résultats de segmentation en utilisant un seul paramètre.

Taux d'erreur					
	Ecart-type	Entropie	Energie	S-entropie	D-entropie
Img1	18.3%	21.35%	16.15%	14.58%	34.38%
Img2	17.25%	20.65%	16.25%	16.19%	33.70%
Img3	17.03%	20.51%	17.04%	16.09%	31.60%
Img4	19.05%	21.25%	18.41%	16.58%	29.33%
Img5	18.28%	21.34%	16.30%	15.20%	32.72%
Img6	18.03%	22.03%	16.38%	14.41%	30.71%
Img7	17.98%	21.23%	17.03%	14.89%	31.51%
Img8	17.53%	19.88%	16.53%	16.03%	32.69%
Img9	19.41%	20.54%	18.65%	17.02%	33.20%
Img10	18.44%	19.43%	17.38%	15.34%	31.15%
Img11	17.52%	20.76%	18.02%	17.54%	30.12%
Img12	18.42%	21.23%	18.23%	17.98%	31.25%
Img13	18.32%	22.25%	17.58%	17.23%	32.45%

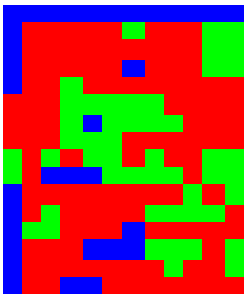
Tableau4.1: le taux d'erreur associé à l'utilisation de chaque paramètre.

III.1.2 Cas d'utilisation de deux paramètres texturaux

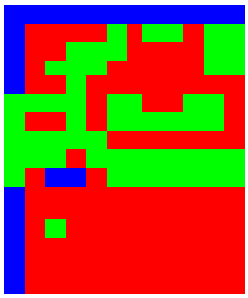
Les résultats des tests ont montré que la segmentation est mauvaise si on utilise seulement un paramètre parmi les suivants: écart-type, entropie, énergie, somme entropie et différence entropie. Afin d'extraire d'autres caractéristiques plus discriminantes, nous avons effectué d'autres tests en utilisant des combinaisons de deux paramètres texturaux.

Les résultats de segmentation de l'image1 en utilisant les combinaisons de paramètres texturaux précédents sont illustrés par la figure 4.4, et le taux d'erreur associé à chaque résultat est donné par le tableau 4.2

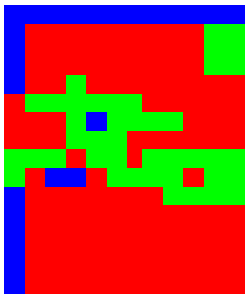
Image1



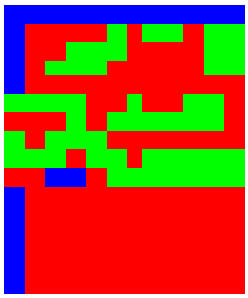
Ecart-type+Entropie



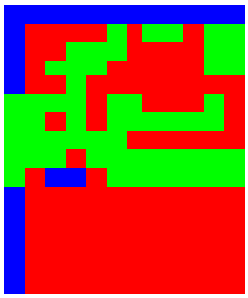
Ecart-type+Energie



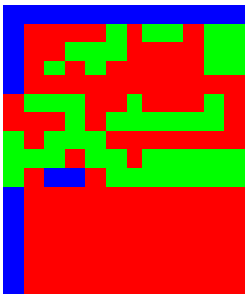
Ecart-type+S-entropie



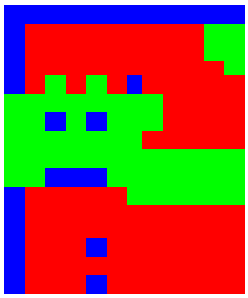
Ecart-type+D-entropie



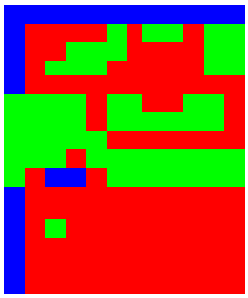
Entropie+Energie



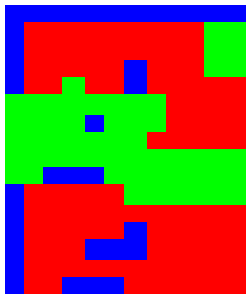
Entropie+S-entropie



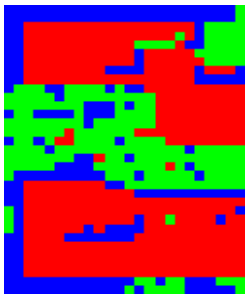
Entropie+D-entropie



Energie+S-entropie



Energie+D-entropie



S-entropie+D-entropie

Figure 4.4 : résultats de segmentation en utilisant deux paramètres texturaux.

	Taux d'erreur									
	Ecart-type +entropie	Ecart-type +Energie	Ecart-type +S-entropie	Ecart-type +D-entropie	Entropie +Energie	Entropie +S-entropie	Entropie +D-entropie	Energie +S-entropie	S-entropie +D-entropie	Energie +D-entropie
Img1	17.19%	15.63%	12.5%	13.54%	14.06%	15.62%	9.38%	13.02%	7.29%	7.81%
Img2	18.20%	14.54%	11.4%	13.25%	14.30%	14.89%	9.27%	13.50%	7.50%	7.65%
Img3	17.46%	15.05%	11.55%	12.89%	13.87%	15.50%	9.50%	12.93%	6.95%	7.91%
Img4	19.91%	17.83%	14.3%	14.22%	14.98%	15.32%	10.04%	12.95%	9.94%	10.04%
Img5	17.02%	15.44%	12.61%	12.99%	13.89%	14.15%	12.10%	11.95%	11.20%	11.89%
Img6	17.25%	15.45%	12.38%	13.41%	14.33%	15.71%	10.95%	13.15%	9.19%	9.25%
Img7	17.39%	16.42%	11.56%	13.23%	14.28%	15.41%	11.65%	12.89%	7.15%	8.29%
Img8	16.95%	15.95%	12.44%	12.98%	13.42%	14.42%	12.30%	12.45%	8.79%	9.13%
Img9	18.03%	16.63%	11.99%	12.87%	13.83%	14.52%	10.17%	13.25%	9.83%	9.89%
Img10	16.75%	15.28%	11.40%	13.52%	14.52%	15.54%	10.39%	12.15%	9.12%	9.25%
Img11	18.24%	16.45%	12.58%	13.95%	13.78%	14.42%	10.24%	13.14%	9.78%	9.85%
Img12	16.23%	15.48%	12.52%	13.54%	14.89%	14.43%	9.87%	12.87%	9.45%	10.02%
Img13	16.45%	15.98%	11.78%	12.98%	13.55%	15.03%	12.75%	12.15%	9.98%	10.13%

Tableau4.2 : le taux d'erreur associé à l'utilisation de chaque combinaison de deux paramètres.

Pour améliorer nos résultats, nous avons effectué des tests en utilisant des combinaisons de trois paramètres texturaux.

Les résultats de segmentation de l'image1 en utilisant les combinaisons de trois paramètres texturaux précédents sont illustrés par la figure 4.5, et le taux d'erreur associé à chaque résultat est donné par le tableau 4.3.

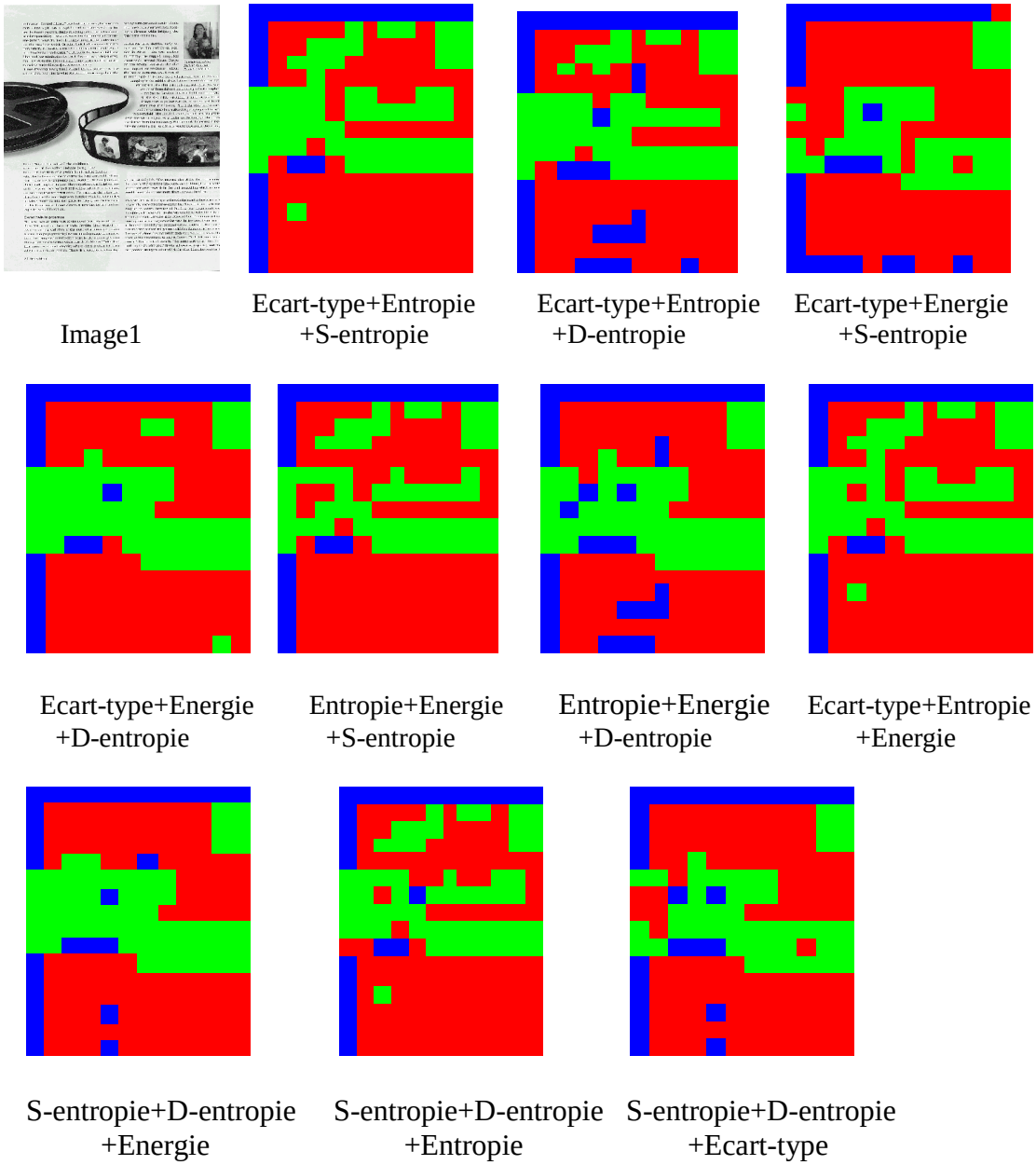


Figure 4.5 : résultats de segmentation en utilisant des combinaisons de trois paramètres texturaux.

	Taux d'erreur									
	Ecart-type +entropie +S-entropie	Ecart-type +Entropie +D-entropie	Ecart-type +S-entropie +Energie	Ecart-type +D-entropie +Energie	Entropie +Energie +S-entropie	Entropie +D-entropie +Energie	Energie +Entropie + Ecart-type	Energie +S-entropie +D-entropie	S-entropie +D-entropie +Entropie	+D-entropie +S-entropie +Ecart-type
Img1	15.10%	10.93%	14.06%	9.38%	14.06%	10.41%	15.10%	9.38%	10.41%	12.50%
Img2	15.60%	11.65%	13.39%	9.20%	14.88%	11.37%	16.73%	9.22%	12.42%	12.33%
Img3	15.29%	12.60%	14.91%	10.19%	13.95%	10.50%	15.23%	10.03%	11.47%	12.47%
Img4	15.81%	10.41%	14.10%	10.25%	13.88%	11.22%	16.50%	10.15%	11.29%	11.98%
Img5	17.20%	11.95%	14.50%	12.41%	13.70%	12.33%	17.03%	11.15%	11.40%	12.02%
Img6	14.18%	12.00%	13.08%	10.25%	13.98%	11.85%	14.20%	10.93%	10.02%	12.41%
Img7	15.41%	12.73%	14.39%	10.11%	13.19%	10.19%	14.25%	10.91%	10.38%	11.91%
Img8	14.93%	11.91%	13.41%	9.27%	14.18%	10.41%	15.27%	9.24%	11.02%	12.44%
Img9	16.02%	11.50%	13.32%	10.63%	13.29%	11.19%	16.73%	11.10%	11.04%	11.51%
Img10	15.97%	10.19%	14.19%	9.32%	13.19%	11.45%	15.39%	11.20%	12.03%	12.00%
Img11	15.24%	10.99%	13.42%	10.32%	14.05%	12.23%	16.32%	10.99%	12.35%	12.47%
Img12	16.89%	11.42%	14.75%	11.25%	15.25%	11.56%	17.89%	11.81%	11.65%	11.47%
Img13	15.97%	12.78%	13.02%	10.98%	15.12%	11.89%	15.23%	11.35%	10.78%	11.24%

Tableau4.3 : le taux d'erreur associé à l'utilisation de chaque combinaison de trois paramètres.

III.1.4 Cas d'utilisation de quatre et cinq paramètres texturaux

Les résultats des tests obtenus en utilisant des combinaisons de trois paramètres ne sont pas meilleurs, pour cela nous avons effectué des tests en utilisant des combinaisons de quatre et cinq paramètres.

Les résultats de segmentation de l'image1 en utilisant les combinaisons de quatre et cinq paramètres texturaux sont illustrés par la figure 4.6, et le taux d'erreur associé à chaque résultat est donné par le tableau 4.4.

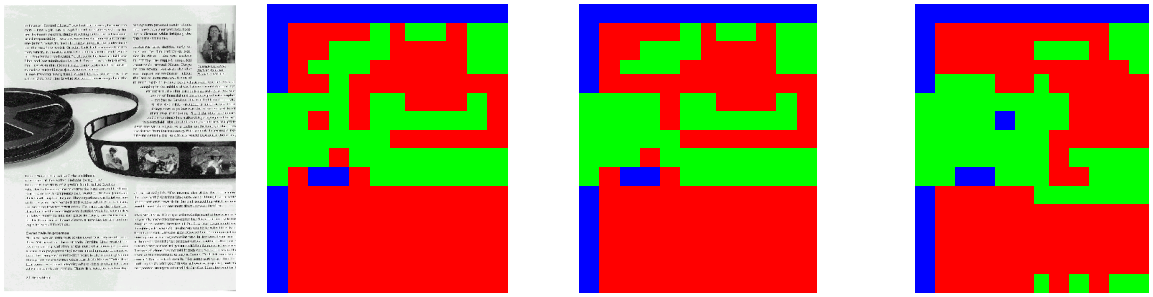
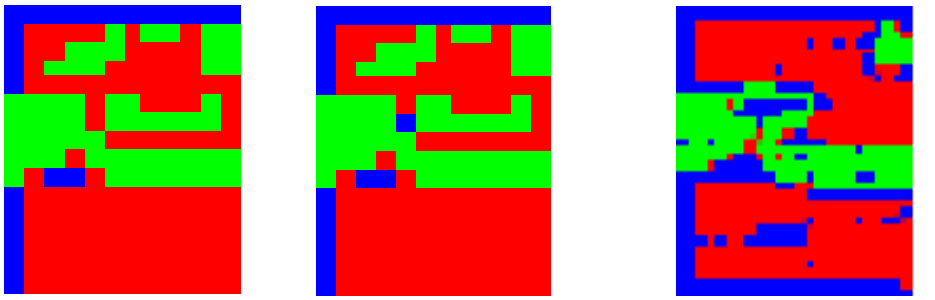


Image1 | Ecart-type+Entropie +Energie+S-entropie | Ecart-type+Energie +S-entropie+D-entropie | Ecart-type+Entropie +S-entropie+D-entropie



Ecart-type+Energie +Entropie+D-entropie | Energie+Entropie +S-entropie+D-entropie | utilisation des cinq paramètres

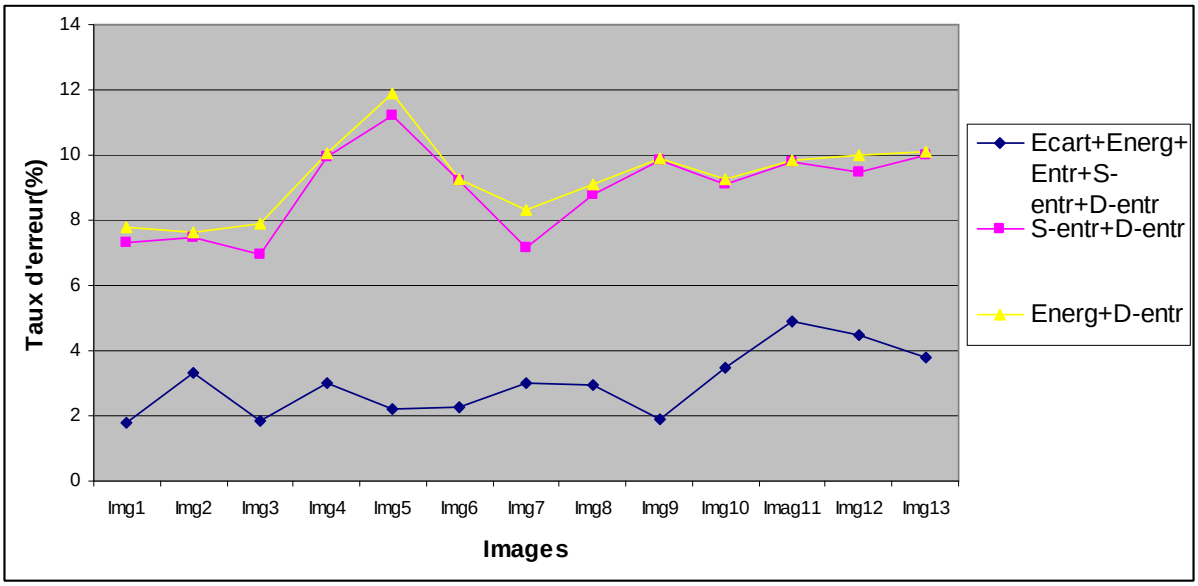
Figure 4.6 : résultats de segmentation en utilisant des combinaisons de quatre et cinq paramètres texturaux.

	Taux d'erreur					
	Ecart-type +S-entropie +Entropie +Energie	Ecart-type +S-entropie +Energie +D-entropie	Ecart-type +S-entropie +Entropie +D-entropie	Ecart-type +Energie +Entropie +D-entropie	Entropie +S-entropie +Energie +D-entropie	Ecart-type +Energie +Entropie +S-entropie +D-entropie
Img1	14.06%	11.46%	10.42%	11.46%	11.45%	1,79
Img2	14.59%	11.23%	10.82%	11.30%	11.13%	3,32
Img3	13.41%	12.29%	10.34%	11.93%	11.88%	1,85
Img4	13.50%	11.30%	10.11%	12.02%	11.90%	3,02
Img5	13.93%	12.31%	12.20%	11.32%	12.40%	2,2
Img6	13.08%	12.29%	10.31%	11.30%	11.50%	2,28
Img7	14.12%	12.20%	11.41%	12.01%	10.97%	2,99
Img8	13.60%	11.41%	10.20%	11.20%	11.46%	2,97
Img9	13.81%	12.30%	10.44%	11.60%	10.85%	1,89
Img10	14.02%	11.41%	10.81%	11.40%	10.73%	3,45
Img11	14.58%	12.45%	12.04%	12.02%	11.12%	4,9
Img12	13.45%	11.98%	11.87%	11.45%	11.42%	4,5
Img13	14.59%	12.87%	12.45%	11.23%	12.02%	3,8

Tableau4.4 : le taux d'erreur associé à l'utilisation de chaque combinaison de quatre et cinq paramètres.

III.1.5 Interprétation des résultats

En analysant les résultats obtenus pour les différentes combinaisons de paramètres on constate que les résultats obtenus en utilisant les deux combinaison de deux paramètres(S-entropie+D-entropie et Energie+D-entropie) sont assez proches et que le meilleur résultat est obtenu en utilisant cinq paramètres texturaux à savoir, Entropie, énergie, différence entropie, somme entropie et écart type. Le graphe 4.1 illustre ces résultats.



Graphe 4.1 : comparaison des résultats de segmentation des trois différentes combinaisons de Paramètres qui donnent les meilleurs résultats.

III.2 Choix de la taille du bloc

Un autre facteur qui influe sur le résultat de la segmentation est la taille du bloc d'analyse. Aussi, après avoir choisi le nombre de paramètres texturaux qui conduisent à la meilleure segmentation nous avons testé différentes tailles de bloc et nous avons évalué le taux d'erreur obtenu pour chaque taille.

Les résultats de segmentation obtenus pour l'image1 et l'image 9 avec différentes tailles de la fenêtre d'analyse sont illustrés par les images des figures 4.7 et 4.8. Les résultats de segmentation des images de la figure4.1 et figure4.2 pour la taille de bloc de 32*32 sont illustrés par les figures 4.9 à 4.21. Les taux d'erreur associés sont donnés par le tableau 4.5.

Image1

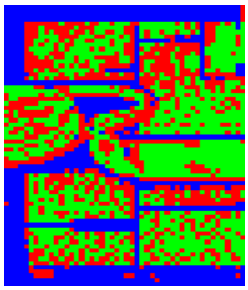


Image segmentée pour la taille 16*16

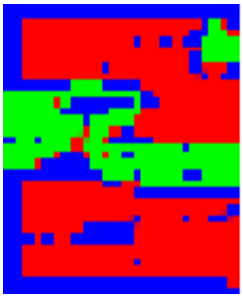


Image segmentée pour la taille 32*32

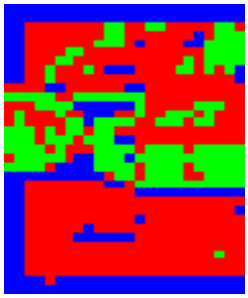


Image segmentée pour la taille 64*64

Figure 4.7 : image1 segmentée en utilisant les trois tailles du bloc.

Image9

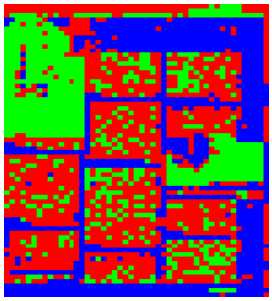


Image segmentée pour la taille 16*16

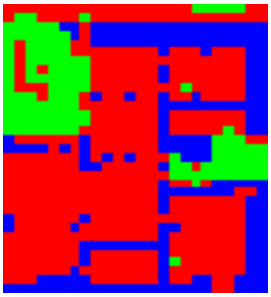


Image segmentée pour la taille 32*32

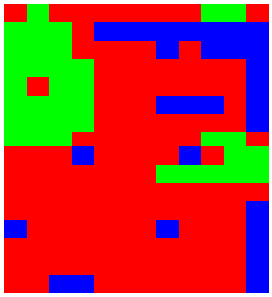
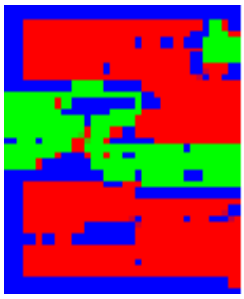


Image segmentée pour la taille 64*64

Figure 4.8 : image9 segmentée en utilisant les 3 tailles du bloc.

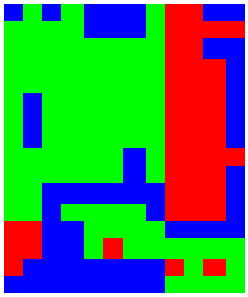
Img1 originale



Imag1 segmentée

Figure4.9 : Exemple d'un document dont l'image est inclus dans le texte avec un fond non uniforme

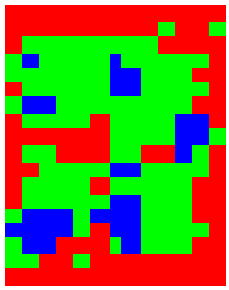
Img2 originale



Img2 segmentée

Figure4.10 : Exemple d'un document dont le texte est inclus dans l'image

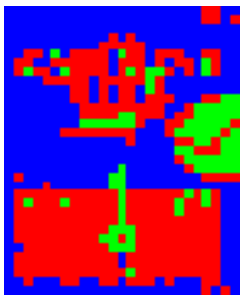
Img3 originale



Img3 segmentée

Figure4.11 : Exemple d'un document dont l'image est inclus dans le texte.

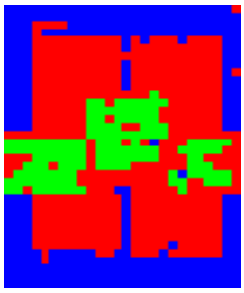
Img4 originale



Img4 segmentée

Figure4.12 : Exemple d'un document dont l'image est inclus dans le texte avec gros caractères

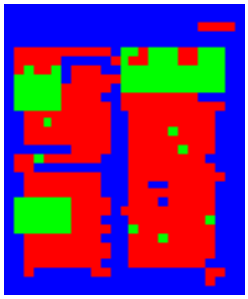
Img5 originale



Img5 segmentée

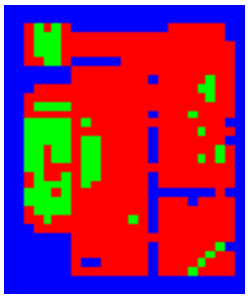
Figure4.13 : Exemple d'un document dont l'image est inclus dans le texte

Img6 originale



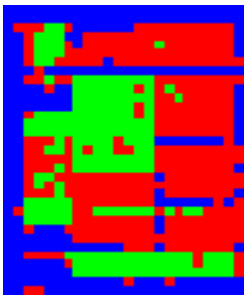
Img6 segmentée

Figure4.14 : Exemple d'un document dont l'image est inclus dans le texte



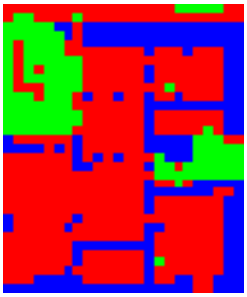
Img7 originale Img7 segmentée

Figure4.15 : Exemple d’un document dont l’image est inclus dans le texte



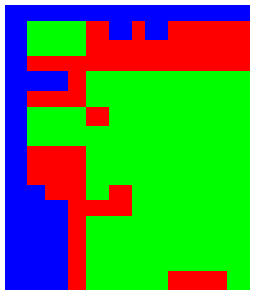
Img8 originale Img8 segmentée

Figure4.116 : Exemple d’un document dont l’image est inclus dans le texte



Img9 originale Img9 segmentée

Figure4.17 : Exemple d’un document dont le texte est inclus dans l’image.



Img10 originale Img10 segmentée

Figure4.18 : Exemple d'un document dont le texte est inclus dans le texte

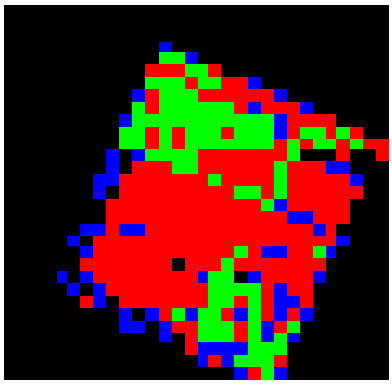


Image11 originale Image11 segmentée

Figure4.19 : Exemple d'un document dont le texte est inclus dans l'image avec une inclinaison de (-15°)

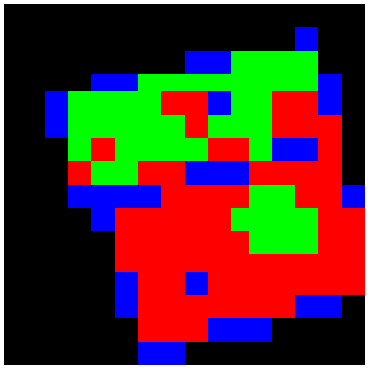


Image12 originale Image12 segmentée

Figure4.20 : Exemple d'un document dont l'image est inclus dans le texte avec une inclinaison de (+30°)

Image13 originale

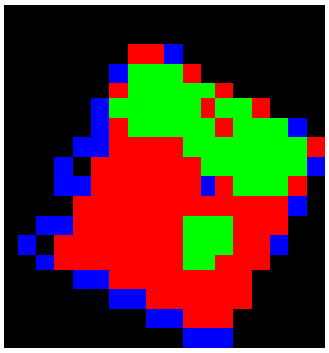


Image13 segmentée

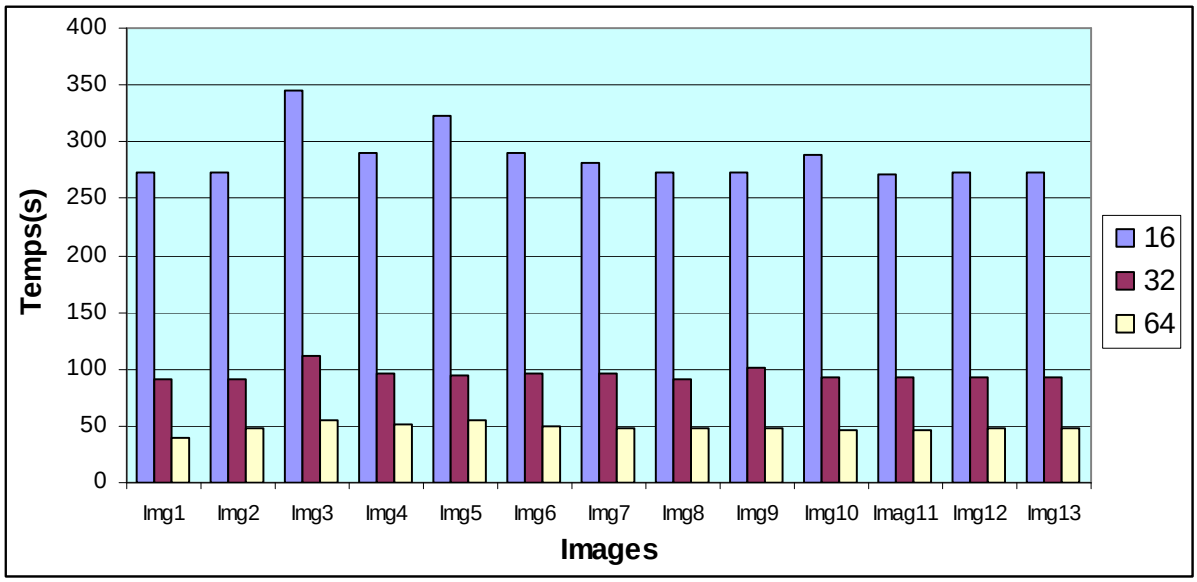
Figure4.21 : Exemple d’un document dont l’image est inclus dans le texte ave une inclinaison de (-15°)

	M=16		M=32		M=64	
	Temps de calcul	Erreur	Temps de calcul	Erreur	Temps de calcul	Erreur
Img1 768*1074	272.45s	25.35%	91.13s	1.79%	40.05s	12.25%
Img2 786*1074	272.61S	25.25%	91.28s	3.32%	48s	18.68%
Img3 850*1165	344.57s	29.25%	112.12s	1.85%	55.54s	24.00%
Img4 820*1074	290.61s	27.36%	96.17s	3.02%	50.76s	26.18%
Img5 850*1165	322.49s	30.13%	94.33s	2.20%	55.52s	14.53%
Img6 816*1074	290.60s	25.21%	96.10s	2.28%	50.60s	18.72%
Img7 802*1067	281.85s	31.15%	96.40s	2.99%	48.34s	25.57%
Img8 788*1063	272.12s	29.22%	91.30s	2.97%	47.80s	16.54%
Img9 784*1067	273.31s	24.05%	100.82s	1.89%	47.94s	17.56%
Img10 804*1067	287.66s	25.32%	91.93s	3.45%	46.79s	18.05%
Img11 763*1063	271.87s	29.45%	93.52s	4.9%	46.80s	17.54%
Img12 750*1045	272.63S	25.89%	92.28s	4.5%	48.25s	18.68%
Img13 735*1002	273.53S	26.89%	91.98s	3.80%	48.85s	19.68%

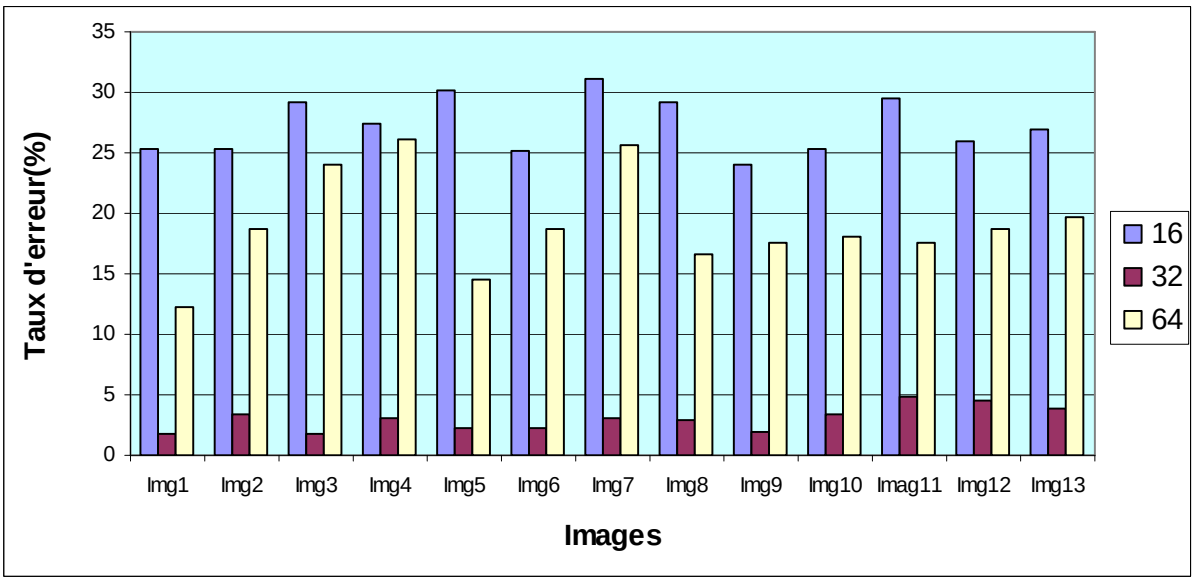
Tableau 4.5 : pourcentage d'erreur de segmentation et temps de calcul en utilisant les différentes tailles du bloc.

III.2.1 Interprétation des résultats

En analysant les résultats obtenus pour les différentes tailles de bloc, on constate que les taux d'erreur en utilisant la taille de bloc 32*32 sont meilleurs par rapport aux résultats obtenus en utilisant les tailles de bloc de 16*16 et 64*64 et le temps de calcul correspondant à cette taille est acceptable. Le graphe 4.2 et 4.3 illustre ces résultats.



Graphe 4.2 : comparaison du temps de calcul pour différentes tailles des blocs



Graphe 4.3 : comparaison du taux d'erreur pour différentes tailles des blocs.

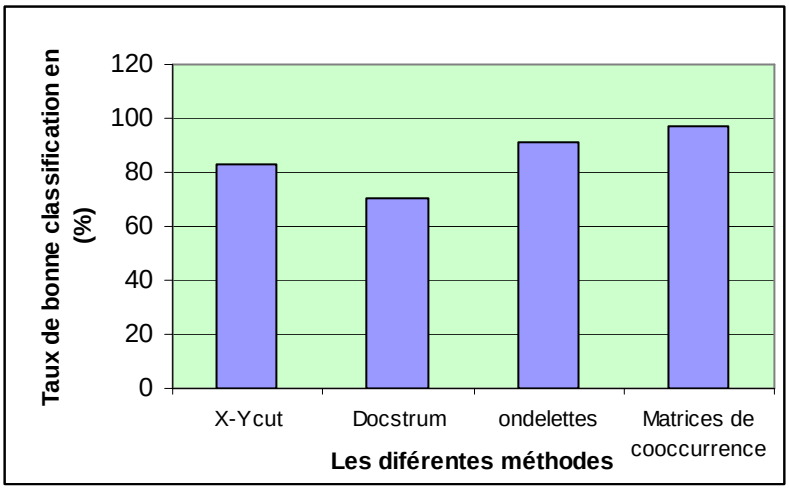
III.3 Evaluation de la méthode

Pour pouvoir évaluer les résultats obtenus par notre méthode, nous les avons comparés à ceux obtenus par les techniques les plus utilisées dans ce domaine à savoir, les ondelettes, les profils de projection et la méthode de docstrum.

Les résultats de ces méthodes sont donnés dans le tableau4.6.

	Taux de bonne classification	**Temps de calcul**
X-Ycut	82.94%	6.37s
Docstrum	70.11%	15.43s
ondelettes	91.00%	3.54s
Matrices de cooccurrence	97.00%	101.65s

Tableau4.6 : Comparaison des résultats des différentes méthodes de segmentation.



Graphe 4.4 : comparaison du taux de bonne classification entre les différentes méthodes.

IV. Discussion

La méthode que nous avons développée a donné de bons résultats sur les deux types de présentation du contenu des documents traités à savoir, les documents dont le texte est inclus dans les images (img2, img9, img10, img11) et dans le cas des documents contenant des

images inclus dans le texte (img1, img3, img4, img5, img6, img7, img8, img12, img13).Le choix des paramètres texturaux caractérisant la texture de documents est très important pour avoir une bonne segmentation comme le montrent les résultats des tests effectués (tableau4.1,tableau4.2,tableau4.3 et tableau4.4).

Le résultat de la figure4.9 montre bien que les trois classes ont bien été identifiées. Dans cette image, la couleur verte représente la classe 'graphique', le rouge représente le 'texte' et le 'bleu' représente le fond. Malgré le fond non uniforme et les dispositions complexe de la page, les paramètres texturaux des matrices de cooccurrence ont permis de caractériser les différentes textures présentent dans l'image, et nous avons pu atteindre un taux de bonne classification de 98,21% soit un taux d'erreur de 1.79% pour une taille de bloc de 32*32.

Les figures 4.10, 4.17, 4.18 et figure 4.19 montrent aussi que les trois classes de l'image à savoir, le texte inclus dans l'image, le fond et l'image ont été correctement segmentée. En effet nous avons pu atteindre pour les images traitées des taux d'erreur de classification qui sont respectivement de 3.32%, 1.89%,3.45% et de 4.9%.

Les figures 4.11, 4.12, 4.13, 4.14, 4.15, 4.16, 4.20 et figure 4.21 montrent aussi que les trois classes ont bien été misent en évidence et nous avons pu atteindre pour ces images traitées des taux d'erreur de classification qui sont respectivement de 1.85% 3.02%, 2.2%, 2.28%, 2.99%, 2.97%, 4.5%, et 3.8%.

La méthode basée sur les matrices de cooccurrence que nous avons conçues a permis de résoudre les problèmes posés par la segmentation des documents inclinés ou la majorité des méthodes de segmentation ne peuvent s'appliquer qu'après une étape préalable de redressement des documents (voir la figure4.19, figure4.20 et figure4.21).

Les résultats des tests (tableau 4.5) montrent que le meilleur résultat est obtenu pour une taille de bloc correspondant à 32*32. Dans ce cas, le temps de calcul nécessaire à l'exécution est plus réduit par rapport aux méthodes utilisant les matrices de cooccurrence sans la division en bloc.

L'ensemble de ces résultats montre que la technique que nous avons développé est bien adaptée aux documents à structure composite et améliore de façon significative le taux de bonne classification comparée aux méthodes usuelles basées sur les ondelettes [55], X-Y cut et l'algorithme docstrum [56] qui donnent des taux de bonne classification de l'ordre de 91%,82.94% et 94.11% respectivement (voir le graphe 4.4).

La qualité des résultats de segmentation des documents obtenus est une conséquence du choix de l'approche de segmentation, des paramètres et de la taille du bloc.

Conclusion

Conclusion

L'une des problématiques majeurs liées à l'analyse des documents est de trouver une méthode qui permet de segmenter les documents à structure complèxes qui possèdent une typographie riche et qui ne sont pas composés uniquement de texte mais d'une combinaison de texte et d'images.

En effet, les differents algorithmes existant tendent à segmenter les documents à structure linéaire en utilisant des images binaires, toutefois, ces algorithmes perdent en précision lorsque la structure du document est complèxe.

Pour palier à ces inconvénients, nous avons développé et mis en œuvre une méthode basée sur l'analyse de texture en utilisant les matrices de cooccurrences. Pour ce faire, après une phase de prétraitement basée sur la diffusion anisotrope, des matrices de cooccurrences sont calculées pour chaque bloc de l'image et des paramètres texturaux sont extraits pour chaque bloc et sont utilisés pour classifier l'image par l'algorithme des k-means.

Les différents tests effectués, ont montré d'une part que les meilleurs résultats sont obtenus pour des tailles de blocs de 32pixels*32 pixels.Nous avons ainsi constaté que le choix de la taille est très important, En effet, elle ne doit être ni très petite au point qu'elle ne contienne pas assez d'informations pour classifier le bloc, ni très grande , au point de contenir toute les régions de l'image.
D'autre part, nous avons montré à travers les résultats, l'importance du choix des paramètres texturaux qui vont caractériser l'image. En effet la combinaison des cinq paramètres retenus a permis d'avoir la meilleure segmentation. De plus, les tests ont montré aussi l'importance de la division en bloc de l'image, qui a permis de réduire le temps de calcul, surtout quand on utilise les matrices de cooccurrences qui sont très Coûteuse en temps.

Comme perspective à notre travail, il serait intéressant de rajouter une technique pour la reconnaissance de la structure logique et un système pour l'extraction des différentes régions de l'image.

Bibliographie

[1] R.Ingold « Analyse et Reconnaissance d'image Documents » /H7020, techniques-ingenieur.

[2] C. I. Tomai, K. M. Allen and S. N. Srihari. “Recognition of Handwritten Foreign Mail”. Proceedings of the 6th International Conference on Document Analysis and Recognition, Seattle, USA, September 2001, *pp*. 882-886.

[3] J.Zhou, C. Y. Suen and K. Liu. “A feedback-based Approach for Segmenting Handwritten Legal Amounts on Bank Cheques”. Proceedings of the 6th International Conference on Document Analysis and Recognition, Seattle, USA, September 2001, *pp*. 887-891.

[4] M. Pfister, S. Behnke and R. Rojas. “Recognition of Handwritten ZIP Codes in a Real World Non-Standard-Letter Sorting System”. Journal of Applied Intelligence, 2, 1998, *pp*.1-25.

[5] Q. Xu, L. Lam and C. Suen. “Automatic Segmentation and Recognition System for Handwritten Dates on Canadian Bank Cheques”. Proceedings of the 7th International Conference on Document Analysis and Recognition, Edinburgh, Scotland, August 2003, *pp*.704-708

[6] H. E. Nielson and W. A. Barrett. “Consensus-Based Table Form Recognition”. Proceedings of the 7th International Conference on Document Analysis and Recognition, Edinburgh, Scotland, August 2003, *pp*. 906-910.

[7] H. Sako, N. Furukawa, M. Fujio and S. Watanabe. “Document-Form Identification Using Constellation Matching of Keywords Abstracted by Character Recognition”, Proceedings of the 5th International Workshop on Document Analysis Systems, DAS2002, Princeton, USA, August 2002, *pp*. 261-271.

[8] A. Dengel and B. Klein. “smartFIX: A Requirements-Driven System for Document Analysis and Understanding”, Proceedings of the 5th International Workshop on Document Analysis Systems, DAS2002, Princeton, USA, August 2002, *pp*. 433-444.

[9] F. Chang, K. Liang, T. Tan and W. Hwang. “Binarization of Document Images using Hadamard Multiresolution Analysis”. Proceedings of the 5th International Conference on Document Analysis and Recognition, Bangalore, India, September1999, *pp*. 374-377.

[10] M. Nadler. “A Survey of Document Segmentation and Coding Techniques”. Computer Vision, Graphics, and Image Processing, vol. 28, 1984. *pp*. 240-262.

[11] S. N. Srihari and G. W. Zack. “Document Image Analysis”. Proceedings of the 8th International Conference on Pattern Recognition, Paris, France, October 1986, *pp*. 434-436.

[12] R. Cattoni, T. Coianiz, S. Messelodi and C. M. Modena. “Geometric Layout Analysis Techniques for Document Image Understanding: a Review”. Technical Report, IRST, Trento, Italy, 1998.

[13] R. M. Haralick. “Document Image Understanding: Geometric and Logical Layout”. Proceedings of the International Conference on Computer Vision and Pattern Recognition, 1994, *pp*. 385-390.

[14] A. K. Jain and B. Yu. “Document Representation and its Application to Page Decomposition”. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 3, March 1998, *pp*. 294-308.

[15] S. Mao, A. Rosenfeld and T. Kanungo. “Document Structure Analysis Algorithms: A Literature Survey”. Proc. SPIE Electronic Imaging, Santa Clara, California, USA, January 2003, *pp*. 197-207.

[16] Y. Y. Tang, M. Cheriet, J. Liu, J. N. Said, C. Y. Suen. “Document Analysis and Recognition by Computers”. Handbook of Pattern Recognition and Computer Vision.

[17] S. L. Horowitz and T. Pavlidis. ‘’Picture segmentation by a traversal algorithm. Comput’’. Graphics Image Process, vol. 1, 1972, *pp*.360–372.

[18] R. Lienhart and F. Stuber. ‘’Automatic text recognition in digital videos’’. Technical report, Department for Mathematics and Computer Science, University of Mannheim, 1996.

[19] G. Nagy and S. Seth. “Hierarchical representation of optically scanned documents”. Proceedings of ICPR, 1984, *pp*. 347-349.

[20] K.Y. Wong, R.G. Casey, and F.M. Wahl. Document analysis system. IBM Journal of Research and Development, vol. 26, no. 6, 1982, *pp*.647–656.

[21] A. Antonacopoulos. ‘’Page segmentation using the description of the background’’. Computer Vision and Image Understanding, vol. 70, no. 3, 1998, *pp*.350–369.

[22] A. L. Spitz. “Recognition Processing for Multilingual Documents”, In R. Furuta(ed.), Proceedings of the 1990 International Conference on Electronic Publishing, Document Manipulation and Typography, Gaithersburg, Maryland, USA, September1990, *pp*. 193-205.

[23] T. Pavlidis and J. Zhou. «Page Segmentation by White Streams”, Proceedings of the 1st International Conference on Document Analysis and Recognition, St-Malo, France, September 1991, *pp*. 945-953.

[24] H. S. Baird. “Background structure in document images”. In H. Bunke, P. S. P. Wang, & H. S. Baird (Eds.), Document Image Analysis, World Scientific, Singapore, 1994, pp. 17-34.

[25] S. Messelodi and C.M. Modena. ‘’Automatic identification and skew estimation of text lines in real scene images’’. Pattern Recognition, vol. 32, no. 5, 1999, *pp*. 791–810.

[26] Y.Wang, I. T. Phillips, and R. M. Haralick.’’ Document zone content classification and its performance evaluation’’. Pattern Recognition, vol. 39, no. 1, 2006, *pp*. 57–73.

[27] S. Nicolas, T. Paquet, and L. Heutte. ‘’Extraction de la structure de documents manuscrits complexes à l’aide de champs markoviens’’. In Actes du 9ème Colloque International Francophone sur l’Ecrit et le Document, 2006, *pp*. 13–18.

[28] F. Lebourgeois, Z. Bublinski and H. Emptoz. “A Fast and Efficient Method for Extracting Text Paragraphs and Graphics from Unconstrained Documents”. Proceedings of the 11th International Conference on Pattern Recognition, the Hague, 1992, *pp*. 272-276.

[29] L. O'Gorman. “The Document Spectrum for Page Layout Analysis”. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 11, November 1993, pp. 1162-1173.

[30] A. Azokly. “Une approche générique pour la reconnaissance de la structure physique de documents composites”. PhD thesis, IIUF-University of Fribourg, 1995.

[31] K. Hadjar, O. Hitz and R. Ingold. “Newspaper Page Decomposition using a Split and Merge Approach”. Proceedings of the 6th International Conference on Document Analysis and Recognition, Seattle, USA, September 2001, *pp*. 1186-1189.

[32] J. Liu, Y. Tang, Q. He and C. Suen. “Adaptive Document Segmentation and Geometric Relation Labelling: Algorithms and Experimental Results”. Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria, 1996, *pp*. 763-767.

[33] T. Pavlidis and J. Yhou. “Page Segmentation and Classification”. CVGIP vol. 54, No. 6, 1992, *pp*. 482-469.

[34] Y. Zhong, K. Karu, and A. K. Jain. ‘’Locating text in complex color images’’. Pattern Recognition, vol. 28, no. 10, 1995, pp. 1523–1535.

[35] Y. Zhong, H. Zhang, and A. K. Jain. ‘’Automatic caption localisation in compressed video’’. IEEE Trans. Pattern Anal. Mach. Intell. vol. 22, no. 4, 2000, *pp.385* – 392.

[36] T. Sato, T. Kanade, E. K. Hughes, and M. A. Smith. ‘’Video OCR for digital news archive’’. In International Workshop on Content-Based Access of Image and Video Databases (CAIVD ’98), 1998, *pp*. 52–60.

[37] V.Wu, R. Manmatha, and E. M. Riseman. Text finder: An automatic system to detect and recognize text in images. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 11, 1999, *pp*. 1224–1229.

[38] C. Datong, K. Shearer, and H. Bourlard. ‘’Text enhancement with asymmetric filter for video OCR’’. In ICIAP ’01: Proceedings of the 11th International Conference on Image Analysis and Processing, page 192, Washington, DC, USA, 2001. IEEE Computer Society.

[39] N. Journet. Analyse d’images de documents anciens.’’ Catégorisation de contenus par approche texture’’. PhD thesis, Université de La Rochelle, 2006.

[40] J-P.Cocquerez et S.Philipp : ‘’Analyse d’images : filtrage et segmentation’’, Masson ; 1995.

[41] A.Gagalowicz, "Vers un modèle de textures ", Thèse de doctorat université de Pierre et Marie Curie, Paris V, mai 1983.

[42] K.Kpalma :''Analyse fractale de texture naturelles dans un contexte multi résolution'' : Application à la segmentation d'images multi résolution, 1992.

[43] A.K Jain and M.Tuceryon:''Texture analysis'', 1992.

[44] P.Perona, J.Malik "Scale-Space and Edge Detection Using Anisotropic Diffusion". IEEEPattern Anal. Machine Intell, vol. PAMI-12, no. 7, *pp* 629-639, 1990.

[45] B.Jahne "Digital Image Processing".Edition Springer Verlag 2002.

[46] M-W Lin, J-R Tapamo, B.Ndovie "A Texture-based Method For document Segmentation and Classification". School of Computer Science, University of KwaZulu-Natal, Durban 4041, South Africa, 2006.

[47] R. M. Haralick, K. Shanmugam and I. Dinstein. "Textural features for image classification". IEEE Trans-actions on Systems, Man and Cybernetics, vol. SMC-3, no. 6, November 1973, *pp*. 610–621.

[48] R.M. Haralick. "Statistical and structural approaches to texture". Proceedings of the IEEE, vol. 67, no. 5, May 1979, *pp*. 786–804.

[49] M. Tuceryan and A. Jain. "Texture Analysis". In L. P.C.H. Chen and P. Wang (editors), The Handbook of Pattern Recognition and Computer Vision (2nd Edition), World Scientific Publishing, 1998, *pp*. 207–248.

[50] S. Teeuwsen. "Feature selection for small-signal stability assessment". In Proceedings of the Dresdner Kreis 2002. Werningerode, Germany, March 2002.

[51] D. Frandkin and I. Muchnik. "A Study of K-means clustering for Improving Classification Accuracy of Multi-Class SVM". Tech. Rep. TR: 2004-02, DIMACS, April 2004.

[52] J. MacQueen. "Some methods for classification and analysis of multivariate observations". In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. 1967.

[53] A. Belaid and Y. Belaid: ''Methods and Applications''. Pattern Recogition, Dunod Informatique, 1992.

[54] A. K. Jain, M. N. Murty and P. J. Flynn. "Data clustering: a review". ACM Computing Surveys, vol. 31, no. 3, September 1999, *pp*. 264–323.

Annexe A

Indices d'Haralick

Dans Haralick et al. [1973], 14 indices ont été définis à partir de la matrice de cooccurrence pour caractériser les textures. Ils sont rappelés dans cette annexe avec les notations suivantes :

- Ng : Nombre de niveaux de gris distincts dans l'image
- $\sum_i , \sum_j : \sum_{i=1}^{N_g} , \sum_{j=1}^{N_g}$ respectivement
- $\sum_{ij} : \sum_i \sum_j$
- P_{ij} : entrée (i, j) de la matrice de cooccurrence normalisée
- $P_x(i) = \sum_j P_{ij}$: entrée i de la matrice de probabilité obtenue en sommant sur les lignes de P (i, j)
- $P_y(j) = \sum_i P_{ij}$
- $P_{x+y}(k) = \sum_{i+j=k} P_{ij}$, k = 2, 3, . . ., 2Ng (somme sur les diagonales principales).
- $P_{x-y}(k) = \sum_{|x-y=k|} P_{ij}$, k = 0, 1, . . ., Ng−1 (somme sur les diagonales secondaires)

Remarque : La traduction en Français de ces indices étant variable selon les auteurs, je conserve les noms anglais donnés par Haralick.

1. Angular Second Moment : Ce paramètre mesure l'homogénéité de l'image ; il est d'autant plus grand que l'image a des transitions de niveaux de gris dominantes.

$$f_1 = \sum_{ij} P_{ij}^{\ 2} \qquad \text{(A.1)}$$

2. Contrast: il mesure le contraste d'une image ; il est élevé lorsque l'on passe d'un pixel très clair à un pixel très foncé ou inversement.

$$f_2 = \sum_{n=0}^{N_g-1} n^2 P_{x-y}(n) \qquad \text{(A.2)}$$

3. Correlation: il décrit la corrélation des niveaux de gris d'un pixel avec un autre distant de d dans la direction θ

$$f_3 = \frac{1}{\sigma_x \sigma_y}(\sum_{ij}(ijP_{ij}) - \mu_x \mu_y) \qquad \text{(A.3)}$$

Où μ_x , μ_y , σ_x et σ_y sont les moyennes et les écart-types de P_x et P_y .

4. Sum of Squares: cet indicateur décrit la dispersion des transitions entre niveaux de gris

$$f_4 = \sum_{ij}(i-\mu)^2 P_{ij} \qquad \text{(A.4)}$$

5. Inverse Different Moment : ce paramètre renseigne sur l'importance des transitions entre niveaux de gris proches.

$$f_5 = \sum_{ij}\frac{1}{1+(i-j)^2}\ P_{ij} \qquad \text{(A.5)}$$

6. Sum Average:

$$f_6 = \sum_{i=2}^{2N_g} iP_{x+y}(i) \qquad \text{(A.6)}$$

7. Sum Variance:

$$f_7 = \sum_{i=2}^{2N_g}(i - f_8)^2 P_{x+y}(i) \qquad \text{(A.7)}$$

8. Sum Entropy:

$$f_8 = -\sum_{i=2}^{2N_g} P_{x+y}(i)\log(P_{x+y)}(i)) \qquad \text{(A.8)}$$

Puisque la probabilité $P_{x+y}(i)$ peut être nulle et que log (0) n'est pas défini, il est recommandé d'utiliser le terme $\log(P+\varepsilon)$ où ε est une constante arbitraire, petite et positive.

9. Entropy : elle fournis une indication sur le désordre que peut présenter une texture.

$$f_9 = -\sum_{ij} P_{ij} \log(P_{ij}) \qquad \text{(A.9)}$$

10. Difference Variance:

$$f_{10} = \sigma^2(P_{x-y}) \qquad \text{(A.10)}$$

11. Difference Entropy:

$$f_{11} = -\sum_{i} P_{x-y}(i) \log(P_{x-y}(i)) \qquad \text{(A.11)}$$

12. Information Measures of Correlation:

$$f_{12} = \frac{HXY - HXY1}{\max(HX, HY)} \qquad \text{(A.12)}$$

13. Information Measures of Correlation:

$$f_{13} = (1 - \exp[-2(HXY2 - HXY)])^{1/2} \qquad \text{(A.13)}$$

$$HXY = -\sum_{ij} P_{ij} \log(P_{ij})$$

Où HX et HY sont les entropies de p_X et P_y

$$HXY1 = -\sum_{ij} P_{ij} \log(P_x(i)P_y(j))$$

$$HXY2 = -\sum_{ij} P_x(i)P_y(j) \log(P_x(i)P_y(j))$$

14. Maximal Corrélation Coefficient :

$$f_{14} = (\text{second largest eigenvalue of Q})^{1/2} \qquad \text{(A.14)}$$

Où

$$Q_{ij} = \sum_k \frac{P_{ik} P_{jk}}{P_x(i)P_y(k)}$$