

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي و البحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

المديرية العامة للبحث العلمي و التطوير التكنولوجي

Direction Générale de la Recherche Scientifique et du Développement Technologique



Projet national de recherche : rapport final

RESTAURATION DES VIEUX MANUSCRITS ALGÉRIENS

Chef de projet : MEZIANE Abdelkrim

Affiliation : Division Systèmes d'information et systèmes multimédia

Organisme de domiciliation : CERIST

Organisme pilote : CERIST

Programme national de Recherche : 2011-2013



R A P P O R T F I N A L D U P R O J E T

« Restauration des vieux manuscrits algériens »

Dr Abdelkrim MEZIANE

Table des matières

| | |
|---|----|
| 1- Introduction Générale | 3 |
| 2- Rappel des objectifs du projet | 5 |
| 3- Dégradation des images de documents | 7 |
| 3.1 Introduction | 7 |
| 3.2 Dégradations physiques du manuscrit..... | 7 |
| 3.3 Dégradations numériques..... | 8 |
| 3.4 Une classification non-causale des dégradations | 9 |
| 4. Etude faite sur les manuscrits du centre d'Adrar | 10 |
| 5. Résultats obtenus | 11 |
| 5.1 Communications internationales | 11 |
| 5.2 Produits réalisés | 11 |
| 5.3 Rapports | 11 |
| 5.4 Séjours scientifiques | 12 |
| 5.5 Formation | 12 |
| 6. Résultats | 13 |
| 7. BIBLIOGRAPHIE | 42 |
| 8. Bilan financier | 45 |
| 9. Annexes | |

1- Introduction Générale:

Le manuscrit ancien représente une richesse pour les pays, car porteur d'informations inestimables, que ce soit du point de vue du contenu, que du point de vue du contenant, donc le support.

Un manuscrit (du latin « *manu scriptus* ») est, littéralement, un texte « écrit à la main », que ce soit par son auteur (« manuscrit autographe ») ou non.

La production d'un manuscrit constitue un art et fait intervenir plusieurs métiers simultanément. Nous retrouvons, dans un premier lieu, le parcheminier, le copiste ou le scribe, l'enlumineur et le relieur.

Les manuscrits anciens se présentent essentiellement sous deux formats :

- Le volumen ou rouleau de papyrus.
- Le codex, livre de feuillets manuscrits.

Contrairement aux manuscrits anciens, les manuscrits modernes consistent, pour l'essentiel, en un ensemble de manuscrits des écrivains ou de divers documents officiels (registres de mairies, registres de banques,...etc.).

Le propre de tout manuscrit est d'être :

- Un objet unique.
- Qui n'est pas multiplié par des procédés mécaniques de reproduction.
- Défini par son format.
- Son support (papier, parchemin, cuir, etc.)
- Le nombre de lignes par page et de mots par ligne.
- Le style de l'écriture.
- La couleur de l'encre.

Il n'existe donc jamais deux manuscrits identiques, même s'il peut arriver que le texte de deux manuscrits soit identique.

L'assemblage de feuillets de manuscrits forme un codex. Par extension le terme de codex a été employé pour des recueils de lois d'où le nom de code.

Le format du codex dépend de la façon dont est plié le feuillet : en deux, in folio, en quatre, in quarto, en huit, in octavo,...etc.

Le premier support de transcription fut le parchemin, conçu à partir de peaux d'animaux, puis de papyrus. Le papier ne vit le jour qu'au cours du 14^{ème} siècle.

Les manuscrits ne respectent aucunement une forme ou une structure bien définie. Ils sont, généralement, de dimensions inégales et formés d'un nombre de pages très varié. Nous retrouvons certains d'entre eux réduits à une seule page et d'autres allant jusqu'à plus de 500 pages.

Le manuscrit est le véritable témoin de la présence d'un texte à une époque donnée. Il véhicule les connaissances de cette ère. Il serait donc, intéressant de pouvoir accéder au contenu de ces manuscrits. Pour ce faire, il faudrait mettre en place un système de description adéquat des manuscrits.

Un manuscrit est tout d'abord une œuvre produite d'une manière artisanale, utilisant des matériaux rares et coûteux. Il est de ce fait un document archéologique qui peut être approché à travers ses caractéristiques matérielles et son histoire.

Un manuscrit est un document unique, qu'il faut examiner depuis sa fabrication, jusqu'à l'intervention des lecteurs et des possesseurs successifs. Par conséquent, les rédacteurs de notices descriptives de manuscrits doivent relever tous les aspects caractérisant les manuscrits, relatifs à sa composition et à son contenu.

Notre pays recèle un patrimoine très riche en manuscrits, qui n'est pas connu à sa juste valeur car difficile d'accès.

Les trois principaux établissements qui sont chargés de la numérisation du patrimoine culturel sont:

- Le centre algérien de la cinématographie(CAC) pour le patrimoine filmique et documentaire (Cinéma).
- La Bibliothèque Nationale pour les livres et manuscrits (BN).
- Et le Centre National des Manuscrits d'Adrar pour les manuscrits (CNM).

La BN et le CNM ont commencé par faire une opération de repérage et de recensement des manuscrits anciens et documents rares à travers les différentes régions du pays, et ce, en vue de les sauvegarder en prenant en charge leur restauration et leur numérisation. Car, rangés dans des khizanates (khizana, en arabe, armoire ou bibliothèque) traditionnelles, ces manuscrits, vieux livres et documents sont exposés à la poussière, et les changements de températures qui, s'ajoutant à la manipulation, les ont fragilisés. Il faut préciser également que l'encre utilisée pour la rédaction des manuscrits peut, au fil du temps, subir des altérations au contact de l'air, et, suite à des réactions chimiques, s'acidifier. Elle peut ainsi devenir un agent destructeur fragilisant un peu plus le support qui peut s'effriter à la moindre manipulation.

Les manuscrits mènent une existence en dehors des murs d'institutions officielles (Bibliothèque National, Archives...). Beaucoup de personnes possèdent des collections privées de manuscrits. Ce sont généralement des personnes qui les ont hérités et nombre d'entre elles, sont des nomades et ne se déplacent qu'avec leur trésor.

Depuis longtemps inaccessibles, ces documents sont en danger de disparition ; ils sont déjà dans un état de décomposition très avancé et avéré car les conditions de conservation ne sont pas assurées. Cette situation lamentable concourt à la disparition d'une part importante de la mémoire commune et un appauvrissement de l'héritage de toute la nation.

2- Rappel des objectifs du projet

Depuis quelques années, on numérise quotidiennement des quantités importantes de documents issus du patrimoine. Cette révolution technologique a ouvert de nouveaux domaines de recherche comme la restauration des images, la numérisation des documents du patrimoine...

Les images de documents sont des images particulières. Ce sont des images de traits qui forment des structures complexes qui doivent être préservées comme les lettres.

Les informations sont concentrées sur des parties très réduites de l'image. La simple perte d'une partie infinitésimale peut rendre un texte illisible. Il existe une certaine continuité le long des lignes de mêmes intensités. La distribution statistique des couleurs montre des classes rarement séparables linéairement. La proximité colorimétrique entre les différentes couches d'information rend particulièrement difficile leurs séparations quand elles sont mélangées comme la séparation entre le recto et le verso par transparence.

Les dégradations sont aléatoires et donc rarement modélisables et prédictibles. Les méthodes à base d'apprentissage ou de modélisation sont généralement limitées au traitement d'une seule dégradation sur des documents particuliers.

Les dégradations peuvent se superposer de façon non-linéaire rendant très difficiles leurs séparations. Elles peuvent être chimiques (encre), biologiques (insectes, rongeurs) ou physiques d'origine humaine (annotations) ou même numérique (capteur).

En restaurant numériquement ces manuscrits, nous conservons d'une façon pérenne notre mémoire collective, et offrons aux générations futures la possibilité de faire des recherches dessus, et nous pourrons contribuer ainsi à la sauvegarde d'une richesse méconnue, et de la part des chercheurs, et de la part du large public. En effet, de peur d'être encore détériorés, nos manuscrits ne sont accessibles à personne, vu leur état avancé de détérioration ; nous voulions grâce à ce projet les mettre à disposition (une copie) d'un grand public.

Dans ce projet, il est question de chercher les méthodes et outils adéquats qui permettraient de restaurer les vieux documents manuscrits algériens. Ces documents même pas « catalogués » qui représentent la mémoire de notre nation est en voie de dégradation continue. Si nous ne faisons pas l'effort de les sauvegarder, mais aussi de les restaurer, ils seront perdus à jamais.

La restauration, nous entendons ici la restauration numérique, car la restauration physique est un autre problème sur lequel penchent les conservateurs, est un problème qui remonte à quelques décennies déjà ; mais les méthodes utilisées restent limitées, vues les différentes façons avec lesquelles le document se dégrade. En effet, chaque type de dégradation donne un aspect particulier au document. Le cumul de tous ces effets ne fait qu'accentuer la dégradation.

L'analyse d'images de documents est un domaine de recherche actif à la frontière de la reconnaissance des formes et de l'analyse d'images.

L'analyse d'images de documents permet aujourd'hui de retrouver des mots dans les manuscrits imprimés ou bien de reconnaître les mots si on dispose de dictionnaires. La numérisation permet de réaliser une « copie numérique » des documents ; mais, pour un grand nombre de ces documents, la numérisation arrive trop tard et beaucoup de documents ont été numérisés en mauvais état.

En plus de la détérioration physique des documents anciens, les images montrent des dégradations dues au processus même de numérisation.

La restauration qui consiste à essayer d'atténuer ces phénomènes et d'avoir une vue du document fidèle à l'original peut être physique (vieillessement, modification des couleurs, porosité du papier, acidité de l'encre, humidité et variation de la température de stockage, présence d'insectes, sinistres...), numérique (inclinaison du document, courbure apparente le long de la reliure, défauts optiques, défauts électroniques, défauts informatiques) ou virtuelle (complémentaire à la restauration physique, permet par exemple de retrouver la couleur d'origine, d'atténuer le verso quand il traverse le recto, de remplir les trous, d'enlever les taches ou de corriger la courbure après la restauration de la reliure).

3- Dégradation des images de documents

3.1 Introduction :

Les images de documents numérisés laissent apparaître souvent de nombreux défauts. Certains dépendent de l'état de conservation du document (la qualité du papier, l'acidité de l'encre, l'humidité du lieu de stockage, vieillissement du support...) ou des modifications apportées par l'homme (annotations, soulignements, mauvaise restauration physique...). D'autres proviennent du processus même de numérisation (résolution insuffisante, courbure apparente, restitution non fidèle des couleurs...). La plupart des travaux, traitant les images de documents anciens, font référence à une classification des dégradations en fonction de leurs causes.

La notion de dégradation des images tient une place centrale dans le présent PNR. Il convient donc de la définir. On appelle dégradation tous les effets indésirables cumulés nuisant à la lisibilité, au traitement ou à la conservation des manuscrits. Les dégradations ont plusieurs origines et l'accumulation des défauts rend difficile leur séparation.

3.2 Dégradations physiques du manuscrit :

a. Dégradations chimiques :

Le papier est un moyen de stockage fragile et éphémère. Il est constitué par l'essentiel de matériaux naturels ou organiques qui sont chimiquement peu stables, les réactions chimiques, qui peuvent avoir lieu, peuvent provoquer des modifications irréversibles dans la structure des matériaux. Les processus chimiques ont des chances de s'intensifier avec l'accroissement de la température, l'humidité, la luminosité et la pollution atmosphérique. Toutes ces réactions chimiques peuvent se traduire, d'une part, par le jaunissement du papier, et de l'autre, par une décoloration des encres et des pigments. Une destruction complète du papier peut avoir lieu par des réactions photochimiques au contact de l'oxygène. L'utilisation des matériaux tels que la colle ou les feuilles de protection en contact direct avec le support papier aide à propager des substances nuisibles qui peuvent ainsi provoquer des dommages.

b. Dégradations biologiques :

Les dégradations biologiques des documents sont causées par des organismes vivants. Les dommages d'origine biologique diffèrent des précédents quant à leurs effets. Ils peuvent être causés par différents agents :

– **Les micro-organismes** (bactéries, moisissures) provoquent des tâches dont la couleur varie avec l'organisme pathogène. Ces tâches envahissent divers types de matériaux et provoquent des altérations irréversibles. L'humidité et la chaleur favorisent le développement de ces micro-organismes.

– **Les insectes** laissent des traces de morsure et d'excoriations. La concentration de matériaux pouvant servir de nourritures dans les bibliothèques, favorise la prolifération des insectes.

– **Les rongeurs** nuisent en particulier à la partie extérieure des ouvrages. Ils grignotent les couvertures, les reliures et les bords des livres.

c. Dégradations physiques d'origine humaine :

L'homme peut contribuer à la dégradation progressive du patrimoine documentaire. Outre les désastres inévitables comme les accidents ou les guerres, certaines dégradations qui sont imputables aux manques dans les règles élémentaires de conservation, peuvent être évitées. Par manque de moyens financiers, les bibliothécaires ne peuvent pas toujours investir dans des systèmes de climatisation pour réduire l'humidité. Certaines n'ont ni sécurité efficace contre les incendies ni alarmes appropriées. Toute manipulation physique, par les conservateurs eux mêmes, détériore les documents fragiles. Citons à titre d'exemple, le cas d'un livre qui tombe par maladresse d'un chariot en cours de transport, un mauvais rangement des microfilms provoquant des rayures. En outre, les bibliothécaires n'hésitent pas à apposer systématiquement des tampons à toutes les pages et même sur les parties écrites les rendant ainsi illisibles. Les annotations (numéros de pages, dates, explications, traductions...) et les rayures (corrections, censures...) effectuées par les chercheurs et les conservateurs contribuent à «l'histoire» de l'ouvrage mais aussi à sa dégradation. Les erreurs de restauration constituent le cas le plus extrême de dégradations humaines pouvant être facilement évitées. La restauration de la reliure, des iconographies ou la consolidation des supports papier peut être mal réalisée et détériore encore plus le document. Enfin, la consultation physique des ouvrages est aussi une des principales sources de destruction des documents fragiles. C'est la raison pour laquelle, les bibliothécaires limitent l'accès aux ouvrages les plus précieux.

3.3 Dégradations numériques:

La numérisation, même de qualité, n'est qu'une représentation imparfaite, à un moment donné, d'un objet physique. Tout dépend des conditions de la prise d'image (placement, éclairage, réglages...) et de la qualité du numériseur (capteur linéaire ou matriciel, bruit du capteur, précision de l'optique et de la mécanique, algorithmes de prétraitement, codage interne des couleurs...). Les choix techniques effectués (choix des paramètres du matériel, des prétraitements et des formats de fichiers) par les acteurs de la numérisation ont plus d'incidence sur la qualité des images que les problèmes précédemment cités.

Les dégradations numériques sont les conséquences des choix technologiques délibérés ou accidentels comme le paramétrage du matériel, les prétraitements irréversibles et la compression avec perte des images. Les bibliothécaires ont très souvent privilégié la numérisation rapide de masse sur une numérisation de qualité. Pour réduire les temps d'acquisition et diminuer les coûts de la numérisation, ils ont fait des choix techniques qui auront des conséquences beaucoup plus graves sur la qualité des images. La réduction délibérée de la résolution spatiale ou colorimétrique et l'utilisation de formats de compression avec perte sont encore très largement utilisées pour sauver l'espace disque et réduire les coûts de stockage des images. Les dégradations numériques sont aussi

causées par une mauvaise utilisation du matériel. Citons, à titre d'exemples, l'application d'une méthode de compression d'images couleur sur une image binaire, le changement d'un format de fichier à un autre, l'application de traitements destructifs comme la correction du contraste ou le rehaussement de la netteté. Tous ces traitements réalisés modifient définitivement l'information. Il convient donc d'étudier les conséquences de ces pertes d'information sur les traitements et la lisibilité.

3.4 Une classification non-causale des dégradations :

F.Derrira propose une classification orientée traitements d'image qui regroupe les dégradations décrites précédemment en fonction des méthodes de restauration possibles.

3.4.1 Dégradations uniformes :

Ce sont les dégradations qui affectent tout le document de façon uniforme. Elles englobent les dégradations colorimétriques ou géométriques comme l'inclinaison, les variations uniformes des couleurs ou de la luminosité ; elles regroupent les défauts issus d'une transformation globale de l'image qui peut agir uniformément soit sur la localisation des pixels (inclinaison, courbure) soit sur leurs valeurs (transformation de la couleur ou correction de la luminosité). La restauration revient, dans le cas présent, à définir la transformée inverse qui permettra de se rapprocher de l'information telle qu'elle l'était à l'origine. La restauration sera donc possible avec des méthodes globales qui appliquent les mêmes traitements sur toute l'image.

3.4.2. Dégradations non-uniformes :

Par dégradations non-uniformes on entend toutes altérations qui varient suivant la position dans l'image. Il sera nécessaire d'utiliser des traitements adaptatifs. Nous avons défini deux sous-classes en fonction de la localisation de l'information par rapport au papier ou au texte : dégradations du fond et dégradations des formes.

a. Les dégradations du fond :

Les dégradations du fond se rapportent aux défauts du support papier comme les taches dues à l'humidité, au passage en transparence du verso, aux annotations et aux ajouts indésirables. Ces artefacts représentent une couche supplémentaire de bruit qui vient se superposer sur le texte original. Restaurer revient dans ce cas à séparer les différentes couches pour en extraire celles qui appartiennent au document original. Une dégradation du fond peut aussi contribuer à la dégradation du texte. Par exemple, le passage du verso sur le recto, les trous et les taches du support papier, rendent la lecture du texte très difficile.

b. Les dégradations des formes :

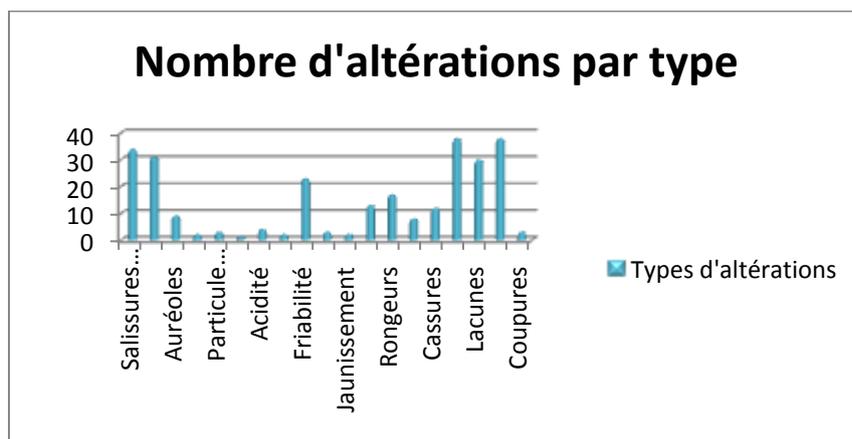
Nous appelons dégradations de formes celles qui modifient la continuité des traits et la topologie des objets. Différentes dégradations peuvent apparaître sur les caractères : rupture des traits, caractères tronqués, vide dans les traits, fusion de caractères, effet d'escalier... En outre, l'apparition des effets de transparence ou de trous risque de rendre certains caractères illisibles. Ces

dégradations sont très pénalisantes pour la lisibilité des textes et pour la reconnaissance automatique des formes.

4. Etude faite sur les manuscrits du centre d'Adrar

Le CNM (partenaire social du projet), nous a fourni les éléments suivants pour ce qui est de la nature des dégradations les plus apparentes sur les manuscrits en leur possession.

| Types d'altérations | Nombre d'altérations |
|----------------------|----------------------|
| Salissures générales | 34 |
| Humidité | 31 |
| Auréoles | 9 |
| Micro-organismes | 2 |
| Particule solides | 3 |
| Gras | 1 |
| Acidité | 4 |
| Décoloration | 2 |
| Friabilité | 23 |
| Oxidation | 3 |
| Jaunissement | 2 |
| Insectes | 13 |
| Rongeurs | 17 |
| Autres... | 8 |
| Cassures | 12 |
| Déchirures | 38 |
| Lacunes | 30 |
| Pliures | 38 |
| Coupures | 3 |



5. Résultats obtenus

5.1 Communications internationales :

Insaf Setitra et Abdelkrim Meziane,

"Old Manuscripts Restoration Using Segmentation and Texture Inpainting". 6th International Conference on Computer and Electrical Engineering (ICCEE 2013). Paris, France. October 12-13, 2013. <http://www.iccee.org>

Insaf Setitra et Abdelkrim Meziane,

"A Restoration Method Based On Texture Inpainting For Algerian Manuscripts". 3d. International Symposium ISKO-Maghreb'2013 Concepts and Tools for Knowledge Management (KM). Marrakech, Maroc. 8 et le 9 Novembre 2013. <http://www.isko-maghreb.org/index.php?headingid=249>

Zineb Hadjadj, Abdelkrim Meziane, Mohammed Cheriet,

« Binarization of Degraded Historical Document Images »

ADDM2013 <http://conf-scoop.org/ADDM/>

ICISIP 2013 <http://www2.ia-engineers.org/icisip2013/>

5.2 Produits réalisés :

-Application pour la restauration de vieux manuscrits par la méthode dite « texture inpainting ».

-Application pour l'élimination des objets superflus sur un manuscrit (tampon, cachet ...) par segmentation.

-Application permettant de supprimer le verso qui apparaît sur une image de document ancien (manuscrit) en utilisant une méthode originale en combinant une technique de seuillage pour la séparation avant/arrière plan avec une technique de classification qui consiste à trouver la couleur dominante de l'avant plan résultant.

-Application permettant de gérer les documents anciens, les indexant, les cataloguant, et permettant de rechercher un document parmi la masse de documents numérisés.

-Application servant à la binarisation des vieux manuscrits arabes, basée sur les contours actifs.

-Application permettant de visualiser, sauvegarder et restaurer par notre méthode de restauration des vieux manuscrits algériens, intitulée « (ADIP) Arab Document Images Processing »

-Un site web permettant de gérer les documents anciens afin de leur donner la chance d'être utilisés par les chercheurs, mais aussi par le public.

5.3 Rapports

1-Etude bibliographique sur les manuscrits et notamment les manuscrits algériens et recensé les problèmes inhérents à ce patrimoine inestimable qui ne cesse de se dégrader.

2-Un état sur les dégradations subies par les documents en possession du centre des manuscrits d'Adrar. Cet état comprend le nombre de manuscrits ainsi que le type de dégradation subie par chacun d'eux.

3- Etat de l'art sur les systèmes de numérisations adaptés à la restauration des vieux manuscrits.

4- Etat de l'art sur les bruits et dégradations sur les vieux documents, et les différentes approches de segmentations pour remédier à ces dégradations.

5- Rapport sur la restauration des images de documents numérisés.

6- Restauration des images de documents numérisés. Techniques de reconstruction des parties d'information textuelles manquantes.

5.4 Séjours scientifiques :

-Participation de Mr Aouaa au 14e Colloque International sur le Document Electronique : CiDE.14 Rabat, Maroc sur l'Analyse du projet "manuscrits de Stendhal" et retour d'expérience. 7-8 décembre 2011.

-stage de Melle setitra, doctorante, du 25 Mars au 23 Avril 2013. Le stage a été réalisé au niveau du laboratoire « Synchronédia Lab » de l' « Ecole de Technologie Supérieure » Montréal Québec sous la direction du Professeur Cheriet Mohamed. Le séjour a servi de recentrer sa thèse de doctorat afin de nous permettre de profiter de son travail dans les manuscrits.

-stage de Mr Maredj Azze-eddine, 5 Avril au 19 Avril 2013. Le stage a été réalisé au niveau du laboratoire « Synchronédia Lab » de l' « Ecole de Technologie Supérieure » Montréal Québec sous la direction du Professeur Cheriet Mohamed. Le séjour a été l'occasion de découvrir les avancées dans le domaine du manuscrit ainsi que l'ouverture d'une collaboration inter-laboratoires.

5.5 Formation :

Sujets de thèses de doctorat proposés dans le cadre de ce projet (travaux en cours) :

Melle Hadjadj Zineb

« Stratégies de traitement et de rehaussement automatiques des images de manuscrits anciens »

Melle Setitra du CERIST

« classification d'objets mobiles dans une séquence d'images vidéo à basse résolution »

Recadrage de cette thèse avec le projet sur les manuscrits, où un stage au Canada a eu lieu, et a permis à la candidate de recentrer ses travaux sur l'image qui nous intéresse, à savoir celle du manuscrit.

Sujets de doctorat proposés en cours de validation

« Contribution à la reproduction des parties manquantes dans les manuscrits arabes. » Sujet proposé à un candidat de l'USTO Oran avec un co-encadrement.

« Caractéristiques de documents anciens : Espaces de représentation et méthodes de rehaussement et de restauration ». Sujet proposé à un doctorant de l'université de Blida, par Pr Mohamed Cheriet de l'université du Québec.

« Accès en ligne guidé par une ontologie aux manuscrits Arabes anciens numérisés. » Sujet proposé à une candidate du CERIST, Mme Bouzid Soumeia, proposé par Mr Meziane et Mr Soualah.

« Contribution à la reconnaissance des caractères dans les manuscrits arabes. » Sujet proposé à un candidat de l'université de Bejaia par Mr Meziane

Sujets de magister proposés et en phase de soutenance dans le cadre de ce projet (avant juin 2014):

« Restauration de documents anciens : suppression du bruit et segmentation » Proposé à Mr Aouaa Noureddine du CERIST par Mr Meziane.

Sujets de master : (encadrés par Mr Meziane)

« Restauration de documents anciens : Texture inpainting »

« Restauration de documents anciens : suppression du bruit et segmentation »

« Restauration de documents anciens : rehaussement de l'espace des couleurs »

« Catalogage des manuscrits »

6. Résultats

L'objectif de ces travaux s'est focalisé sur la segmentation des images de manuscrits afin d'éliminer le bruit et le contenu non désiré. Une approche basée sur la méthode des K plus proches voisins et sur un calcul de poids a été utilisée. Cette approche même si elle est simple a donné des résultats satisfaisants. Une deuxième partie s'est intéressé aux parties manquantes de ces images. La restauration de ces parties manquantes a été faite par une technique de texture inpainting. Plusieurs approches ont été exploitées et la technique d'inpainting par recherche de similarités « exemplar based » a été choisie pour sa simplicité, son temps de calcul et ses résultats comprenant moins de flou que les autres approches exploitées.

Ensuite, deux méthodes de binarisation seront présentées. En effet, les images de document dans les manuscrits s'y prêtent à cet exercice, puisque la majorité d'entre eux sont formés de textes.

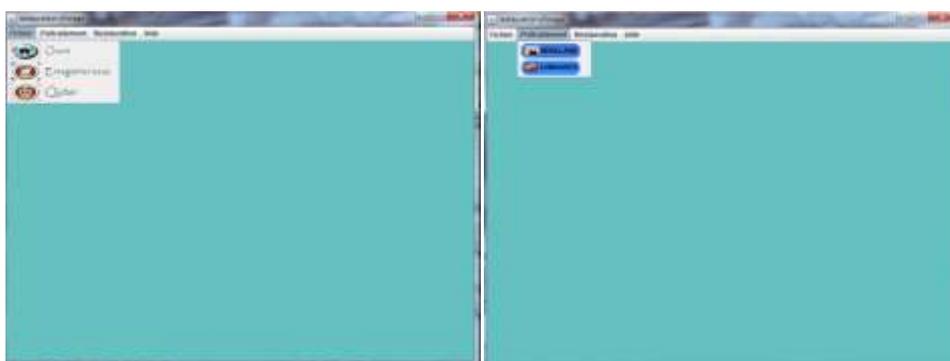
Enfin, un problème majeur dans les manuscrits est l'apparition du verso sur le recto, un problème récurrent dans ce type de documents, auquel nous proposerons une méthode et les résultats obtenus.

Les résultats ont été assez satisfaisants dans leur majorité.

6.1 Application 1

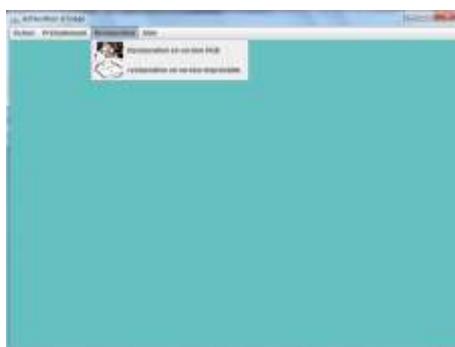
A. Présentation des menus

La fenêtre principale comporte plusieurs menus :



Fenêtre principale/menu Fichier

Fenêtre principale/menu Prétraitement



Fenêtre principale/menu Restauration

B. Résultats :

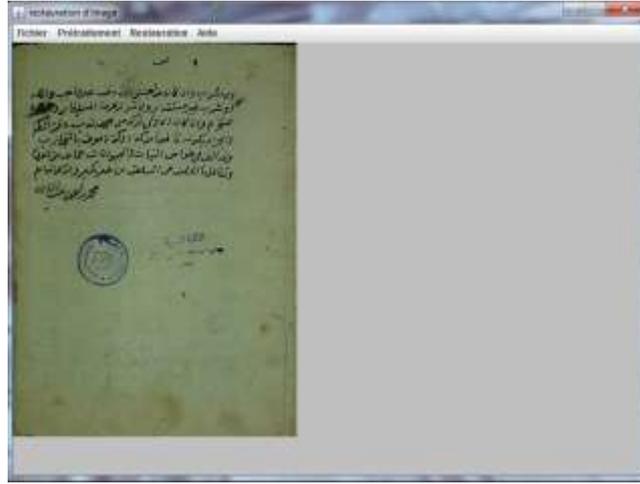


Image originale d'un manuscrit à restaurer

La fenêtre une présente une image dégradée d'un manuscrit que nous allons restaurer à travers l'interface de l'application.

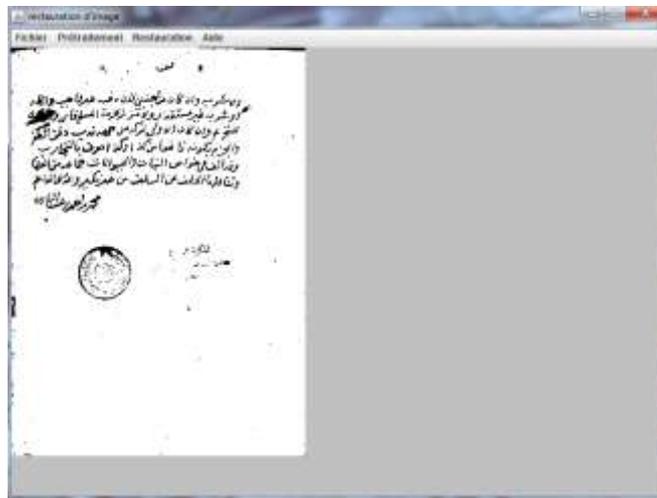


image du manuscrit après seuillage

L'image est binarisée avec moins de dégradation, mais contient toujours des taches et des cachets. Cependant, il est à noter que dans quelques cas, l'image peut être restaurée par simple seuillage, ceci est le cas par exemple de tâche fines, pas très foncées, ou d'un bruit.



image du manuscrit après rehaussement de la luminance



image du manuscrit après restauration RGB

Nous pouvons noter le cachet rond, le cachet sous forme d'annotation et beaucoup de taches qui ont disparu. La couleur du fond quant à elle a été remplacée par une moyenne de la couleur originale afin de présenter une image la plus fidèle à l'image du manuscrit original.

Par ailleurs, des taches à l'intérieur du texte n'ont pas été supprimées. Ceci est dû au fait que l'algorithme des k plus proches voisins traite des éléments voisins et proches pour les segmenter. Les taches confondues à l'écriture sont segmentées comme objets appartenant au texte.

Pour combler cette lacune, des techniques appropriées peuvent être envisagées. Parmi ces techniques la reconstruction de la partie manquante par le text inpainting. Cette piste a été exploitée par le deuxième binôme avec lequel nous avons eu de bons résultats pour compléter l'application ici présente.

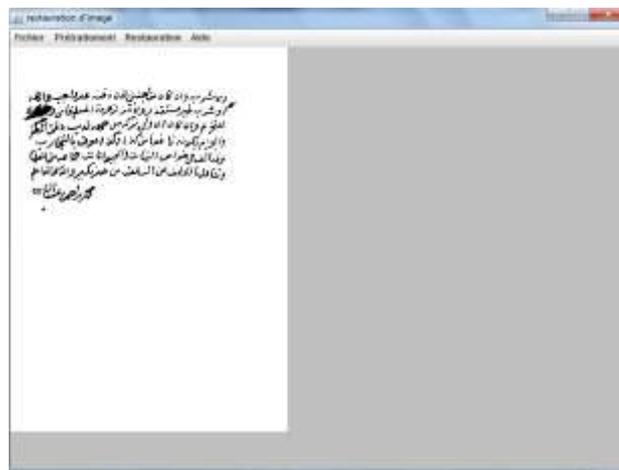


image du manuscrit après restauration pour version imprimable

Disparition des taches, cachets et autre éléments superflus. L'image ici est binaire pour une meilleure impression et une économie d'encre.

C. Conclusion :

Ce travail a permis d'enlever, à partir d'une image de manuscrits, taches, cachets et annotations. Des éléments qui sont souvent considérés comme superflus dans des manuscrits authentiques et qui empêchent leur lecture et exploitation manuelle ou informatique.

Deux résultats sont possibles, le premier étant en couleurs RGB et qui se veut le plus authentique à l'image manuscrite originale. Et un deuxième résultat en binaire (noir et blanc) qui permet la reproduction du contenu du manuscrit et évite d'utiliser beaucoup de couleurs en cas d'impression.

Les résultats de cette phase serviront comme entrée à une phase suivante pour reconstruire des éléments manquants du manuscrit. Cette phase a été faite dans le travail qui suit pour reconstruire les éléments manquants d'un manuscrit à l'aide d'un algorithme d'Inpainting.

6.2 Application 2

A. Présentation des menus

La fenêtre principale comporte plusieurs menus :



Fenêtre principale/menu Fichier



Fenêtre principale/menu Edition



Fenêtre principale/menu Traitement

Le Menu Traitement :

Ce menu contient la commande Inpaint, et de même cette dernière contient les commandes suivantes :

- Run : pour l'inpainting par taille du patch. Dans ce cas, initialement, toute l'image est considérée.
- Fast Inpainting : La taille du patch initialement est inférieure à la taille de l'image c'est-à-dire une partie très petite autour du pixel dégradé est utilisée.

B. Résultats :



Image d'un manuscrit résultat de la première application à restaurer

Cette fenêtre présente une image dégradée d'un manuscrit que nous allons restaurer à travers l'interface de l'application. Cette image est une sortie d'une phase préalable qui consistait à éliminer le bruit, en l'occurrence, les taches, les cachets et les annotations. La partie bleue et la partie sélectionnée manuellement et qui fera l'objet de la reconstruction.

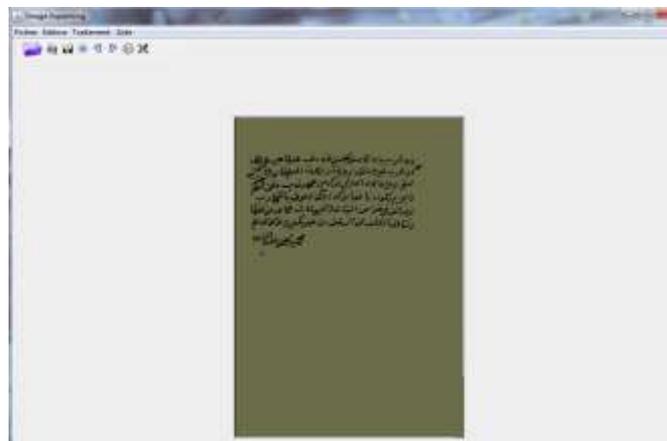


image restaurée par filtre Kernel



image restaurée par filtre Sobel

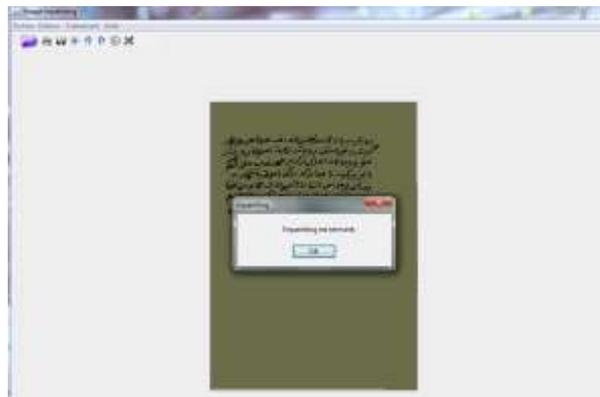


image restaurée par filtre Laplacien

C. Conclusion :

La restauration par filtre Laplacien paraît plus significative que les filtres Kernel et Sobel, le nombre de pixels restaurés est plus grand et les orientations du texte sont mieux respectées. Cependant, nous ne pouvons trancher sur ce contenu, car la reconstruction est faite selon les voisins, ce qui amène à penser à un système de vérification de reconstruction plus intelligent.

L'une des perspectives de ce travail pourrait consister à introduire un système expert pour la vérification du texte reconstruit par cette approche et d'itérer la reconstruction jusqu'à un résultat significatif et satisfaisant.

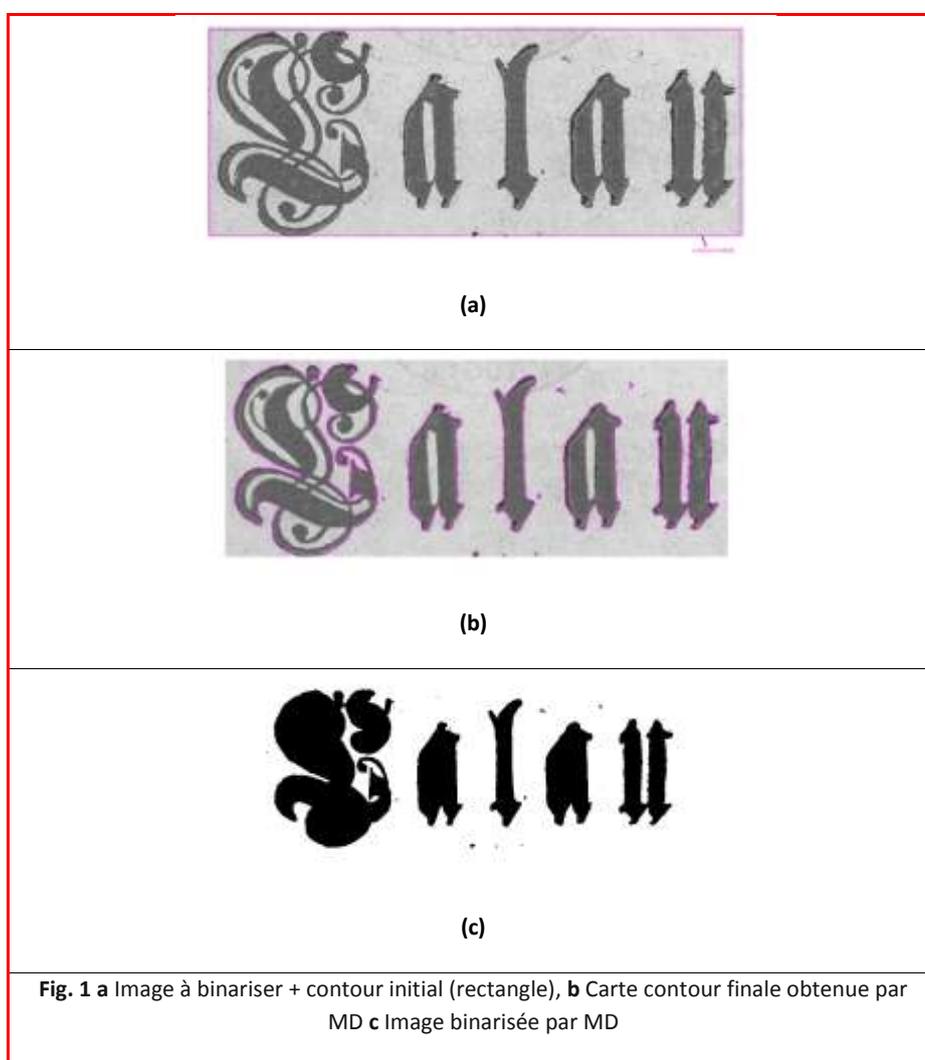
En conclusion, le travail effectué était satisfaisant, et le projet reste ambitieux pour la poursuite des recherches et des développements dans cet axe. L'amélioration des fonctionnalités des systèmes proposés au niveau de notre équipe de recherche est un objectif à atteindre et donnera des résultats puisque plusieurs pistes ont déjà été exploitées.

6.3 Application 3 : Binarisation de documents manuscrits

Bien qu'il existe de nombreuses méthodes de binarisation, aucune ne résout le problème de façon générique, certaines nécessitent un ajustement manuel des paramètres, d'autres ne prennent pas en compte les différences d'éclairage, et d'autres sont très sensibles au bruit. Notre méthode proposée consiste à combiner deux méthodes ayant des propriétés complémentaires afin de binariser des documents dégradés.

Dans ce rapport, on va montrer les deux méthodes choisies et le rôle de chacune, sans détailler leurs principes, et présenter quelques résultats obtenus par ces deux méthodes, ainsi que les résultats de leur combinaison.

Au début, notre méthode consiste à faire une détection de contours de l'image originale par le **modèle déformable (MD) de Ron Kimmel**, qui est une amélioration du modèle de Chan et Vese, pour extraire les caractères manuscrits de l'image (**Figure (1)**) ;



Comme ce modèle présente l'inconvénient de ne pas segmenter les régions intérieures des caractères manuscrits (**Figure (1)**), on a pensé à utiliser une carte contour préalablement calculée par le détecteur de **Canny** comme un contour initial du MD, le résultat final devient : (**Figure (2)**)



(a)



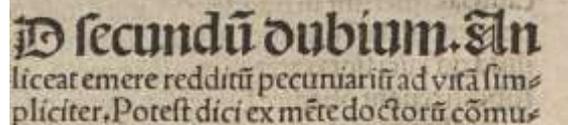
(b)



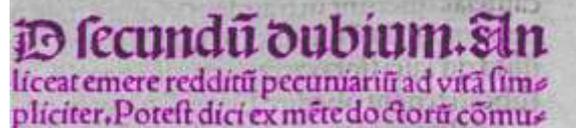
(c)

Fig. 2 a contour initial obtenu par le détecteur de Canny, b Carte contour finale obtenue par MD c Image binarisée par MD

La **Figure (3)** montre un autre exemple de binarisation par MD en utilisant le détecteur de Canny



(a)



(b)



(c)

Fig. 3

a Image à binariser, b Carte contour finale obtenue par MD c Image binarisée par MD

L'utilisation de la carte contour de Canny comme un contour initial a amélioré le résultat final, et la segmentation devient plus rapide, j'ai utilisé la fonction de Reza Farrahi Moghaddam⁽¹⁾ pour évaluer les deux derniers résultats obtenus (Figure (4), Figure (5)) ;



(1) <http://www.mathworks.com/matlabcentral/fileexchange/27652-objective-evaluation-of-binarization-methods-for-document-images>

Fig. 4 Evaluation par la fonction de Reza Farrahi Moghaddam

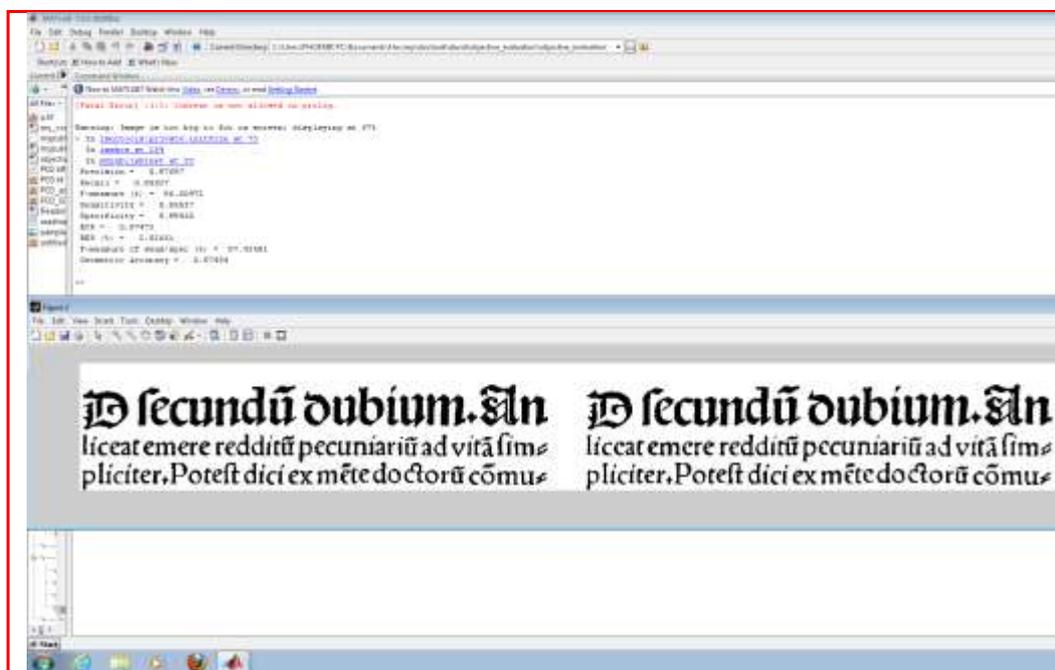
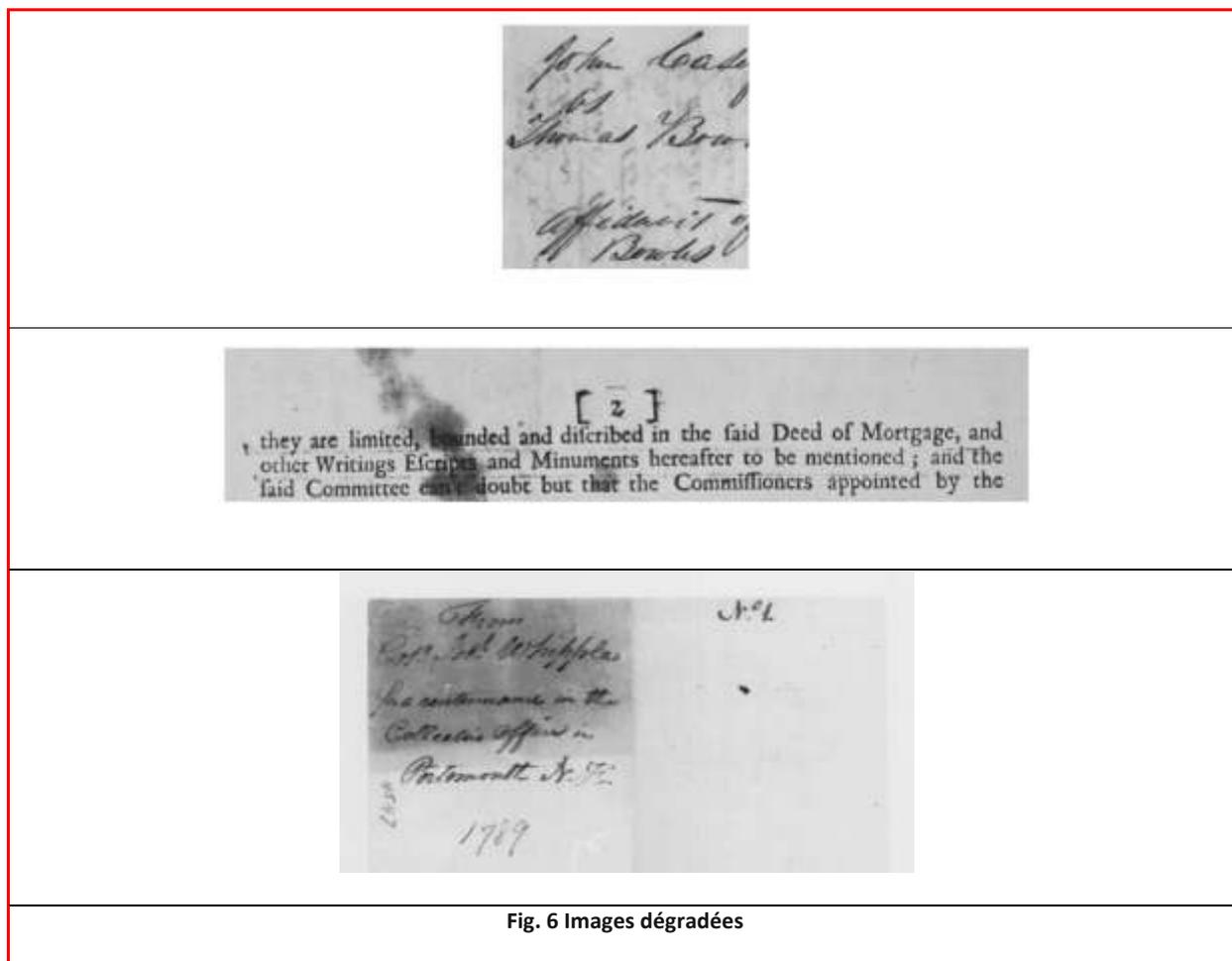


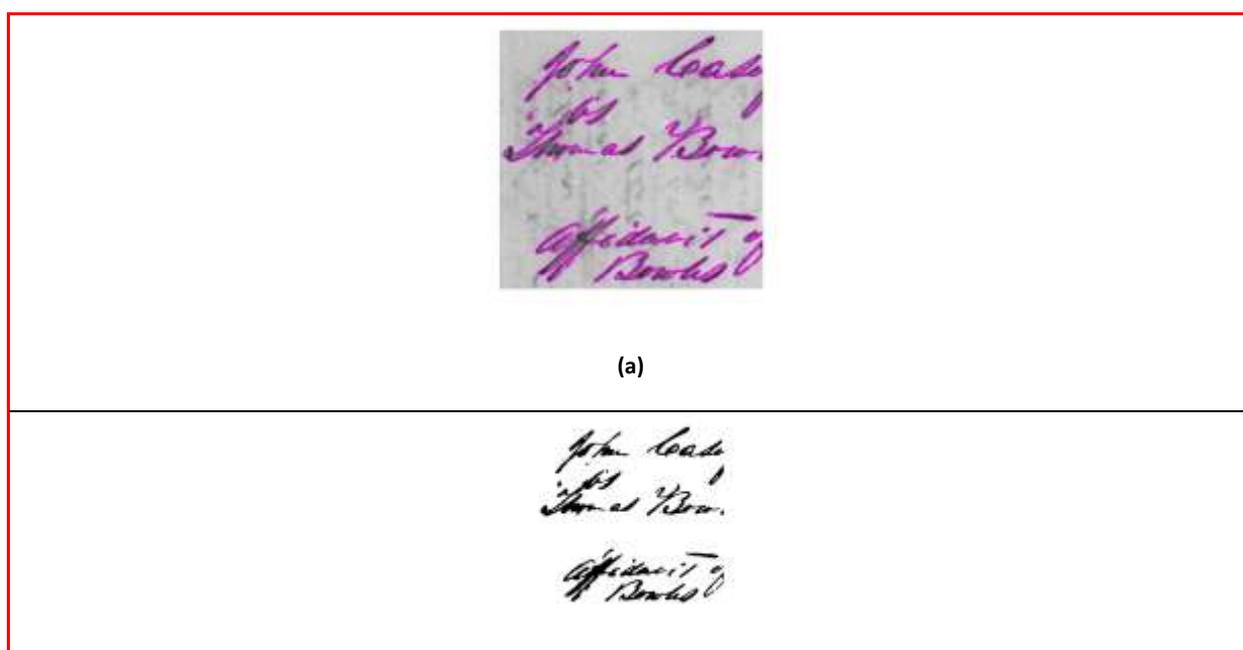
Fig. 5

Evaluation par la fonction de Reza Farrahi Moghaddam

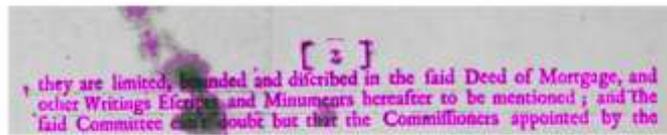
Maintenant, le problème lorsque l'image est dégradée ou fortement dégradée (**Figure (6)**) ;



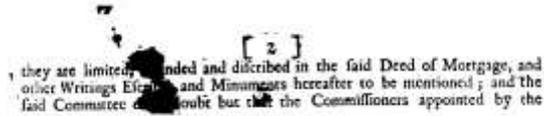
Dans ce cas, le résultat du modèle déformable devient : (**Figure (7)**)



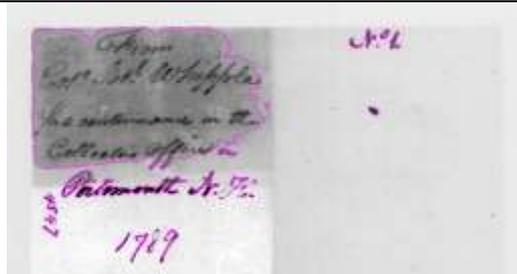
(b)



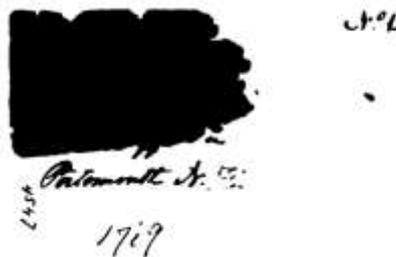
(c)



(d)



(e)



(f)

Fig. 7

a, c et e Cartes contour finales obtenues par MD, b, d et f Images binarisées par MD

Pour résoudre le problème de propagation des contours dans les régions dégradées qui entourent les caractères manuscrits de l'image à binariser, on a proposé de faire coopérer ce modèle avec d'autre méthode pour casser cette propagation, pour cela, on a appliqué à l'image originale un masque obtenu par un average thresholding (moyen) (**Figure (8)**) pour la segmenter par la suite par MD et obtenir l'image binarisée finale (**Figure (9)**);

John Lead
Thomas Bow
Affidavit
Bowles

[2]
they are limited, bounded and discribed in the said Deed of Mortgage, and
other Writings Escrites, and Minuments hereafter to be mentioned ; and the
said Committee can doubt but that the Commissioners appointed by the

From
Col^o Jos^o Whipple
In a continuation in the
Collectors Office in
Portsmouth N. H.
1789
A. 01

Fig. 8

Images obtenues après un average thresholding

On remarque d'après la (**Figure (8)**), que ce type de seuillage a détecté les caractères manuscrits et des parties discontinues du background,

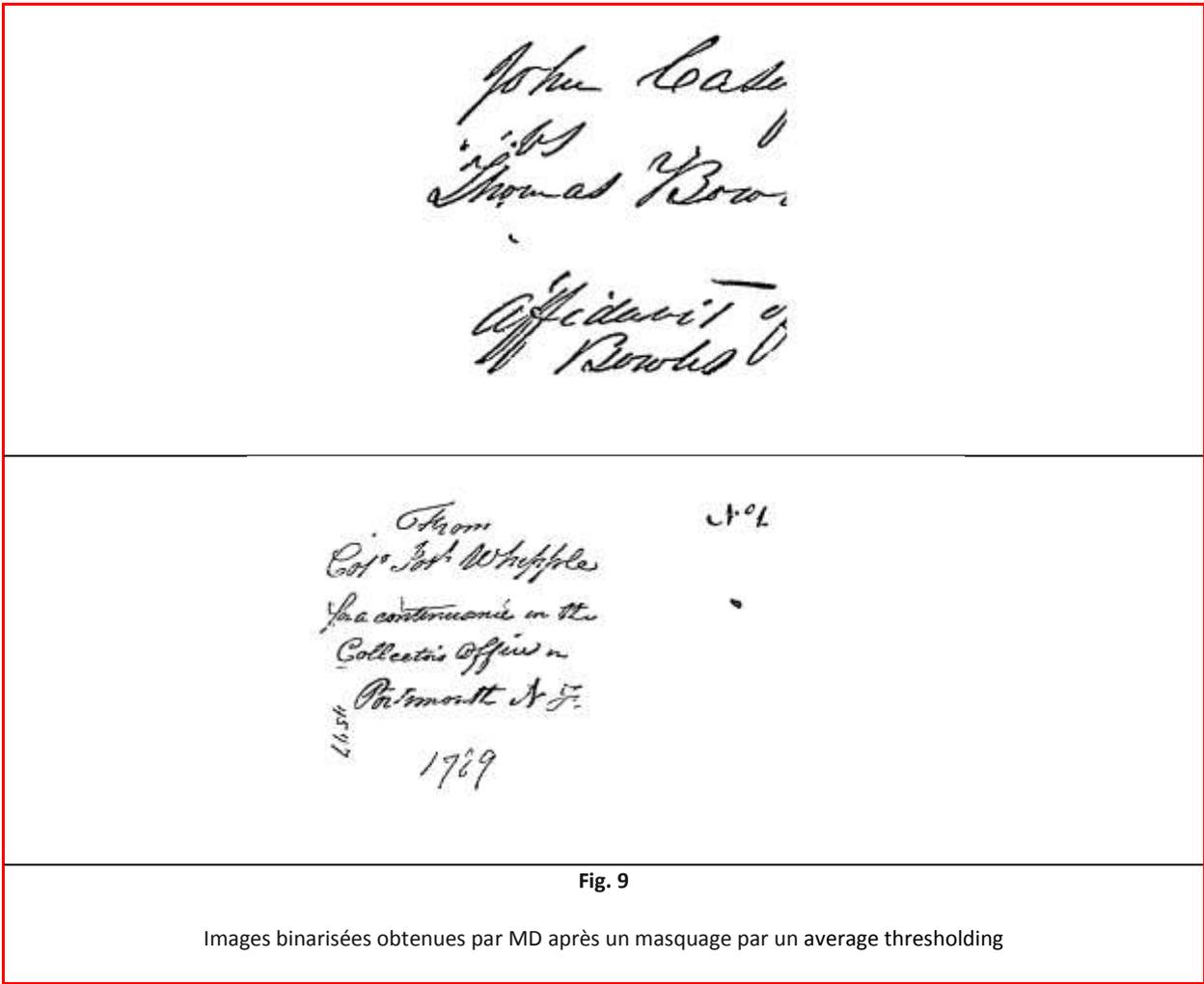


Fig. 9

Images binarisées obtenues par MD après un masquage par un average thresholding

Dans certains cas, on a utilisé un ajustement pour augmenter le contraste et améliorer la qualité globale de l'input image avant de la binariser, (**Figure (10)**) ;

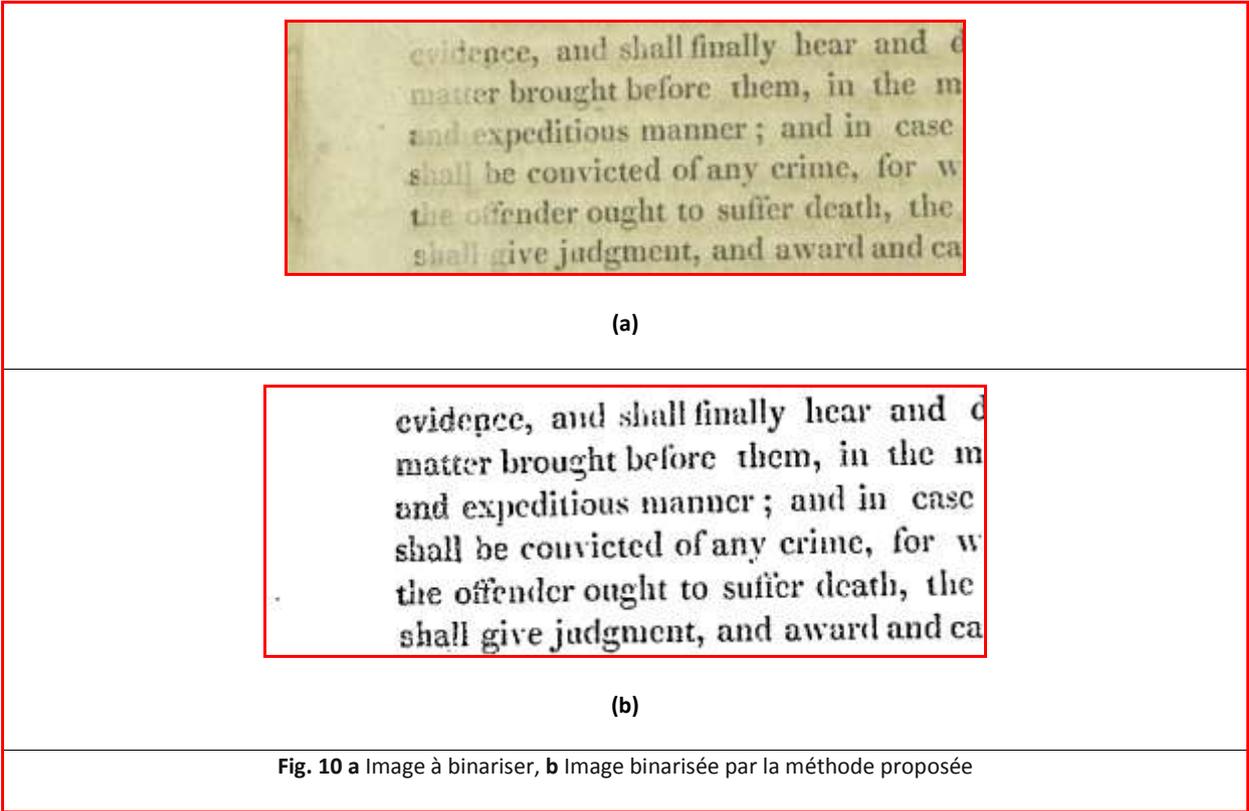
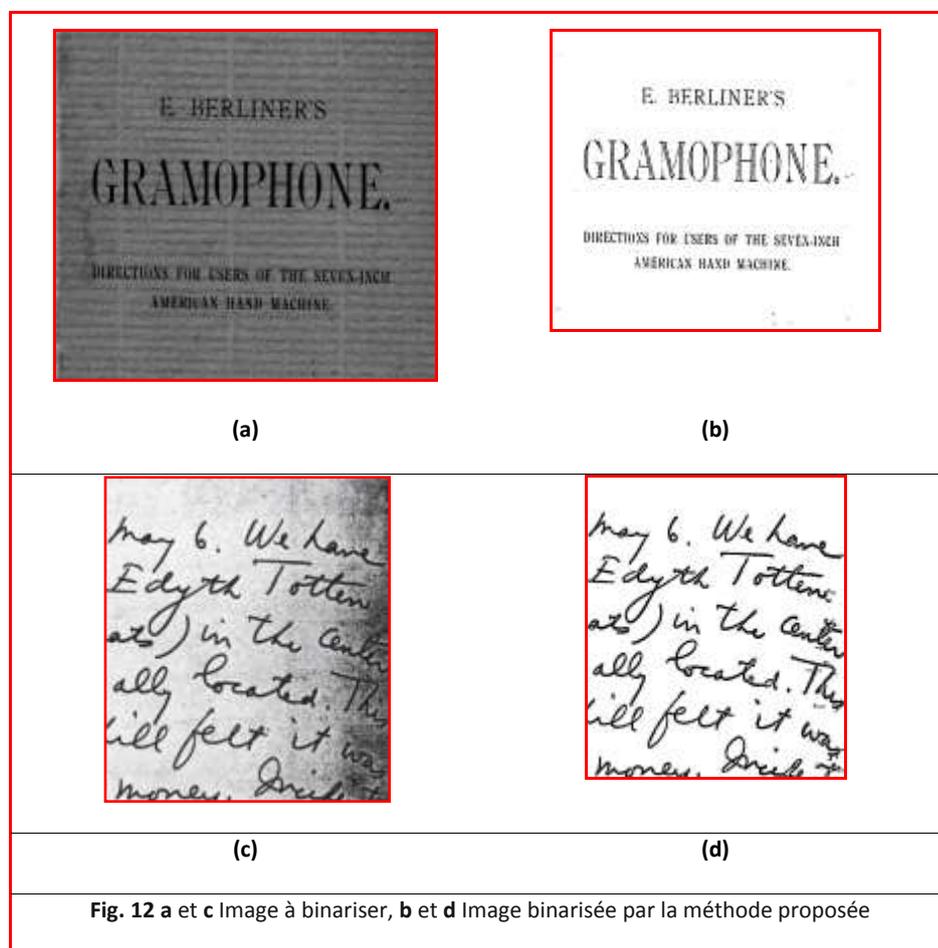
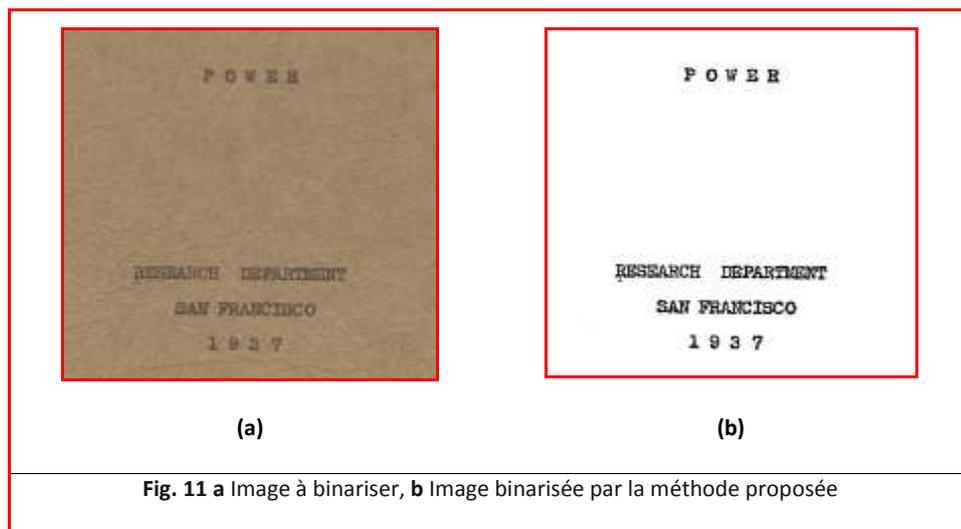


Fig. 10 a Image à binariser, b Image binarisée par la méthode proposée

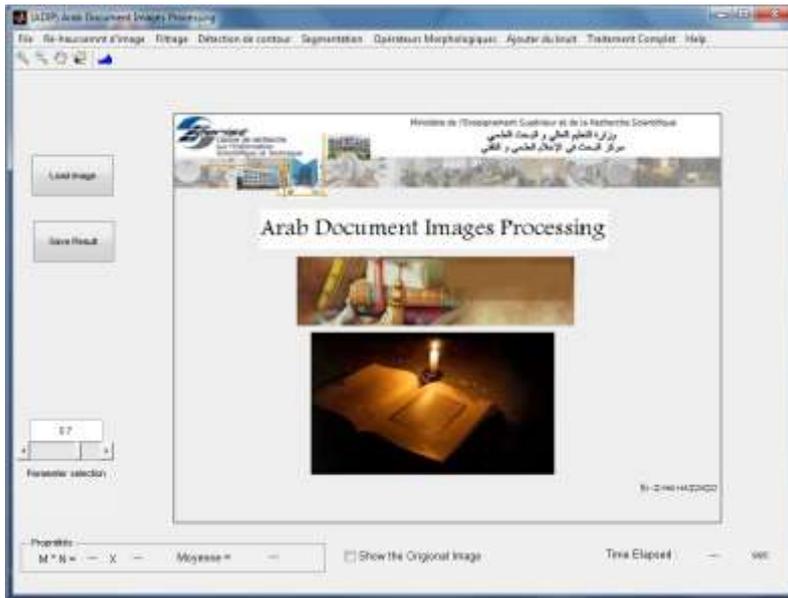
Dans la **Figure (11)** et **Figure (12)**, d'autres exemples de binarisation par cette méthode proposée,



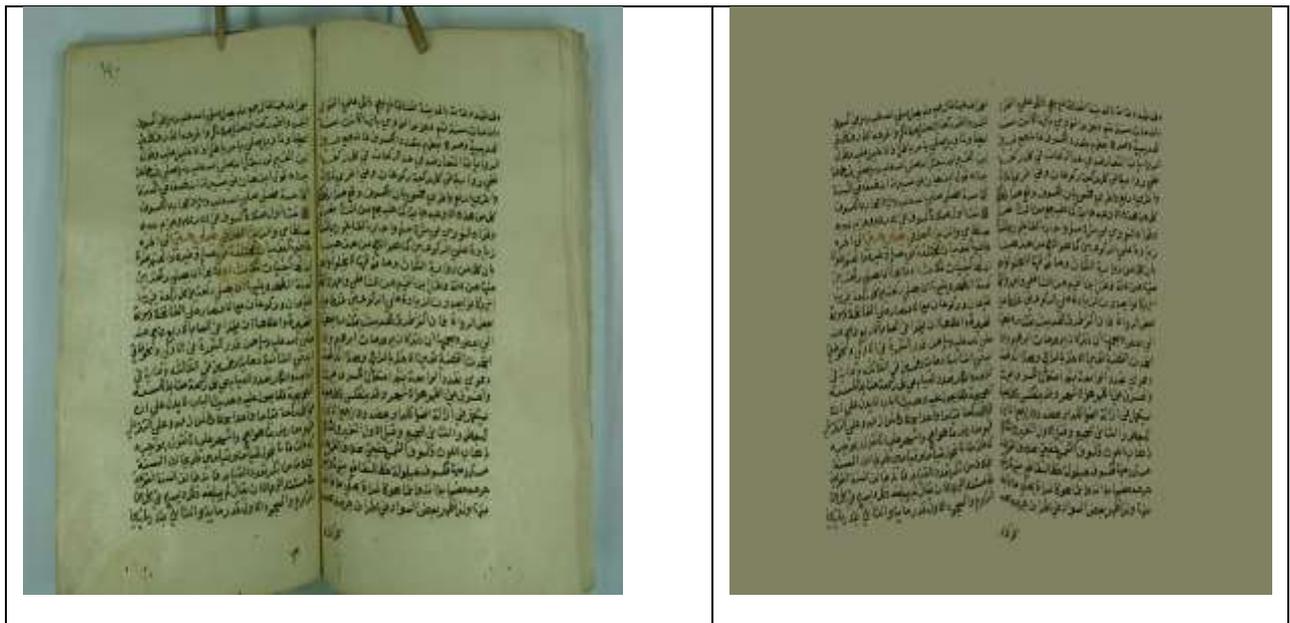
Donc, les deux étapes de notre méthode proposée sont :

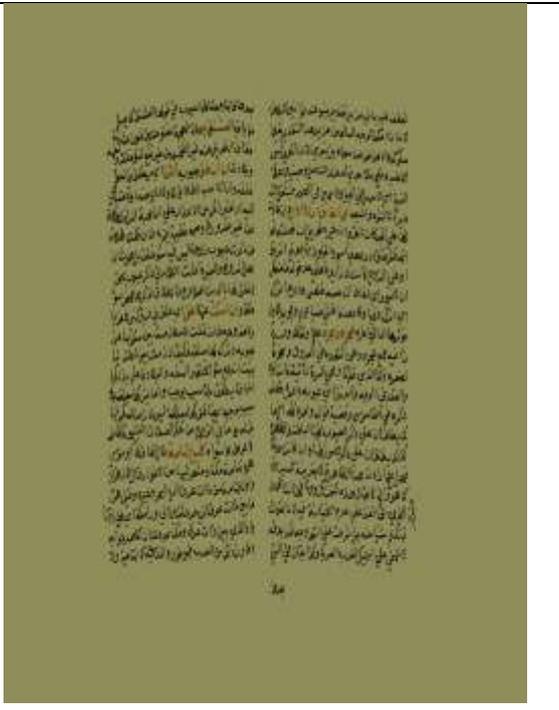
1. Appliquer à l'image originale un masque obtenu par un average thresholding ;
2. Segmenter l'image résultat de l'étape (1) par MD (en utilisant le contour de Canny comme contour initial) ;

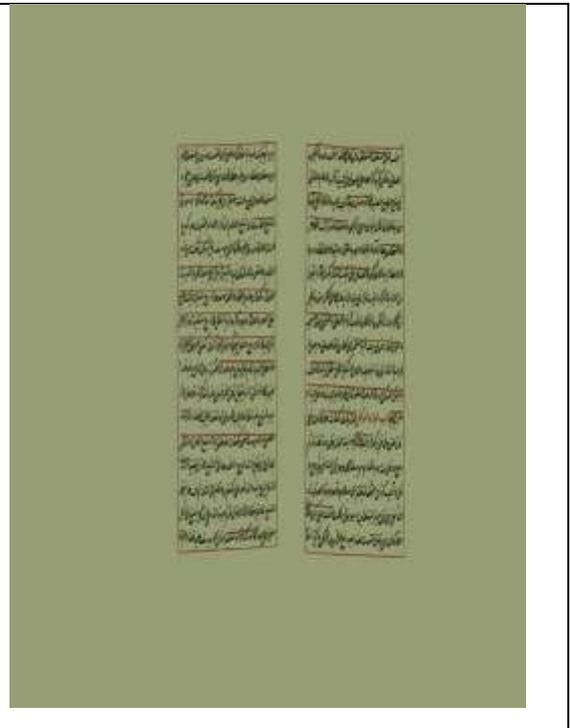
L'interface



Quelques résultats sur des images de manuscrit :







6.4 Application4 : Suppression du verso dans un document manuscrit ancien.

La numérisation de grandes quantités d'images de documents anciens a créé un nouveau contexte de travail, notamment au niveau des dégradations. Parmi ces dernières, le problème du passage en transparence du verso sur le recto, conséquence de la porosité du papier, de la qualité chimique de l'encre ou des conditions de numérisation. Cette dégradation se traduit par des marques diminuant la lisibilité de l'image du document. L'application des méthodes de restauration sur ces images dégradées peut être une solution pour extraire le texte original du recto du fond bruité.



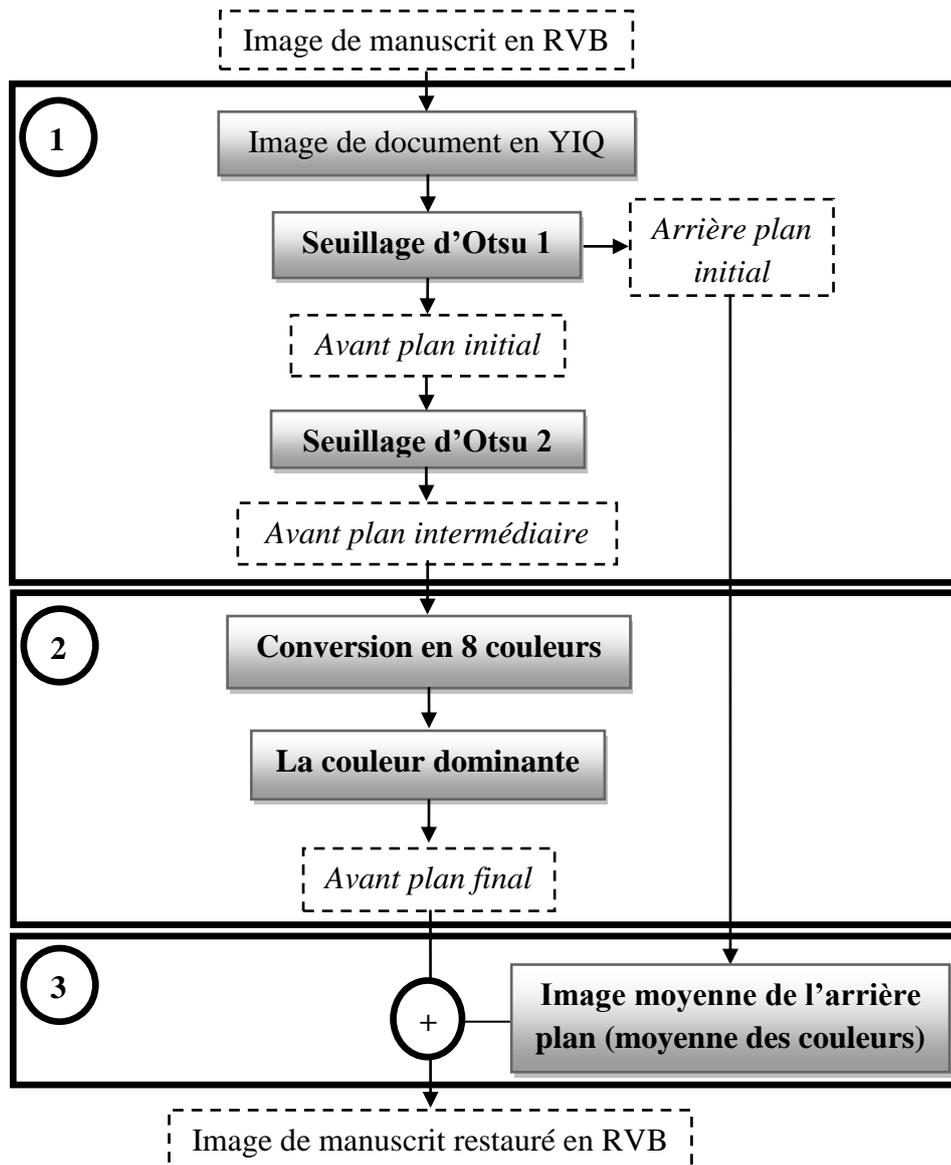
Extrait des images de documents dégradés.

Méthodologie :

Notre approche consiste à combiner une technique de seuillage pour la séparation avant/arrière plan avec une technique de classification qui consiste à trouver la couleur dominante de l'avant plan résultant (la couleur du texte).

Elle procède en trois étapes : séparation de l'avant/arrière plan en utilisant la technique de seuillage global, extraction du texte en gardant la couleur dominante, reconstitution de l'image finale en remplaçant l'arrière plan bruité par une couleur uniforme (moyenne des couleurs).

Le schéma présenté ci-dessous décrit l'architecture générale de notre système.

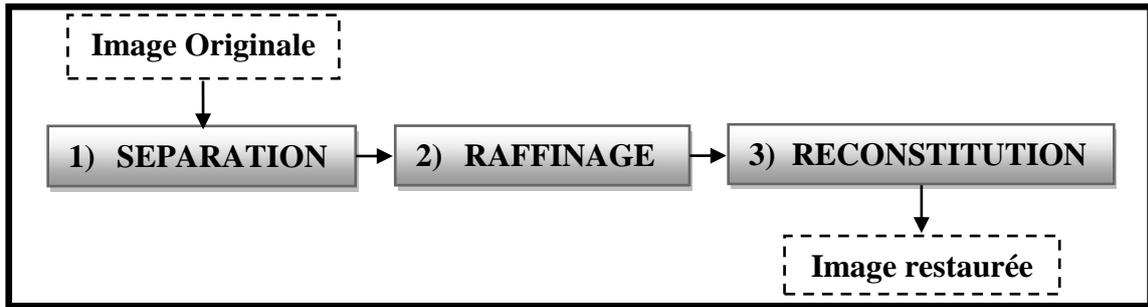


Architecture du système proposé.

Les étapes ci-dessous seront développées dans les sections suivantes :

- Séparation Arrière/Avant-plan en utilisant une technique de seuillage global d'Otsu 1 (avant plan initial et arrière plan initial) et rectification des déformations visuelles de l'avant plan initial obtenue en utilisant la même technique de seuillage d'Otsu 2 (Avant plan intermédiaire).
- Extraction du texte en gardant la couleur dominante après la conversion de l'avant plan intermédiaire en 8 couleurs (Palette 8 couleurs) pour obtenir la couleur dominante de l'avant plan final (couleur du texte).
- Reconstitution de l'image finale en remplaçant l'arrière plan bruité (arrière plan initial) par une couleur uniforme (moyenne des valeurs RVB) avec l'avant plan final on obtient l'image du manuscrit restaurée en RVB.

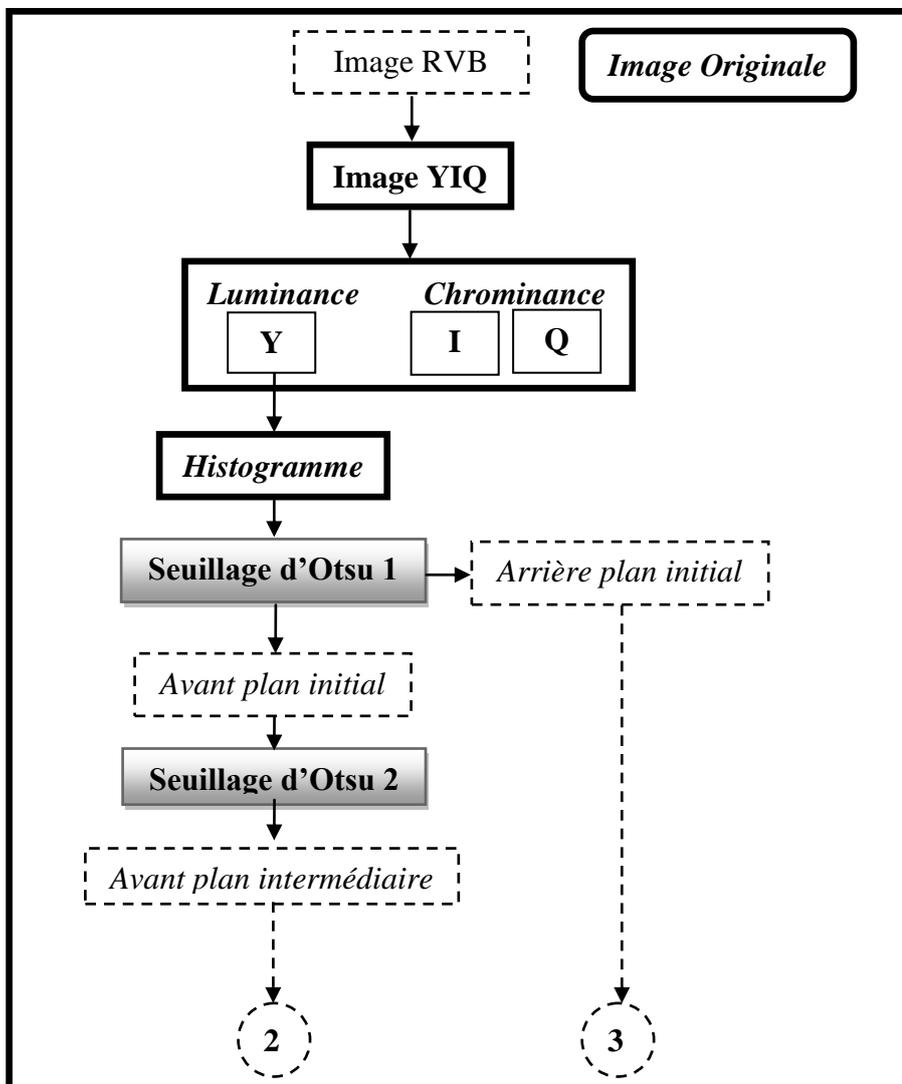
Description détaillée de notre système proposé :



Il est décomposé de trois parties :

La première partie : *Séparation de l'avant/arrière plan.*

- ✓ Tous d'abord on convertit l'image originale en RVB à l'image en YIQ -image en niveau de gris- en construisant l'histogramme de la luminance de la composante Y, pour obtenir la valeur du seuil grâce à la méthode d'Otsu. Après la première utilisation de cette technique (Otsu 1), on obtient l'avant plan initial (texte et transparence avec un fond blanc) et arrière plan avec la couleur du fond originale. Et après la deuxième utilisation (Otsu 2) on obtient l'avant plan intermédiaire (texte et des traces de transparence avec un fond blanc).
- ✓ **Les techniques adoptées:** Espace couleur YIQ et Technique de seuillage d'Otsu.
- ✓ La description de l'étape de *SEPARATION* est présentée dans la figure 6.3.1.a.



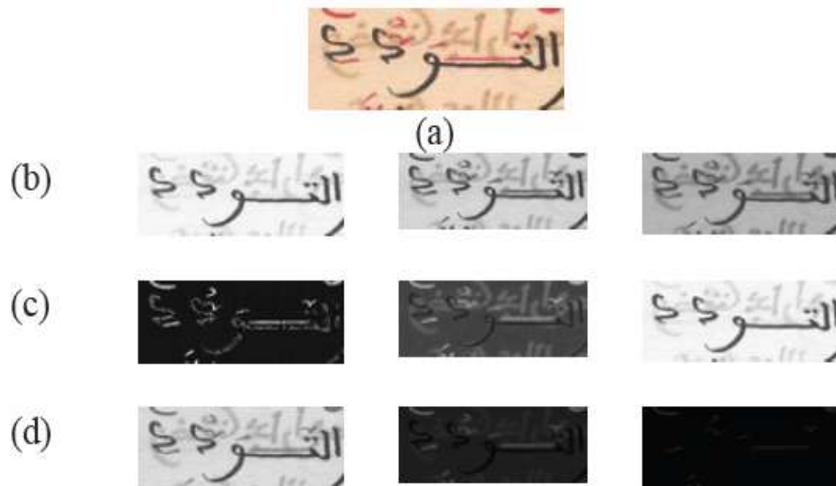
A. Choix de l'espace de couleur :

Nous avons choisi l'espace couleur YIQ, qui a été adopté en premier par le NTSC (National Television System Committee) pour assurer une diffusion efficace pour les téléviseurs couleur en garantissant en même temps la compatibilité avec les téléviseurs noir et blanc.

Ce système de couleurs est défini par la transformation linéaire de l'espace RVB :

- $Y = 0.2992 R + 0.5868 V + 0.1140 B$
- $I = 0.5960 R - 0.2742 V - 0.3219 B$
- $Q = 0.2109 R - 0.5229 V + 0.3120 B$

La composante Y capture l'intensité lumineuse ; I et Q correspondent à la chrominance. Dans cet espace de représentation la composante Y est favorisée par rapport à I et Q. Le choix de cet espace de couleurs se justifie par le fait que le système visuel humain est plus sensible au changement de l'intensité lumineuse. De plus, la variation de l'intensité lumineuse provoquée par l'arrière plan irrégulier dans les images de documents historiques est capturée, dans sa plus grande partie, sur la composante Y. En comparaison avec l'espace de couleur TSL, la teinte est peu sensible à la variation de la luminance et permet la séparation des objets de différentes couleurs. Un exemple de décomposition de l'image RVB et sa transformation aux espaces de couleurs YIQ, HSV est montré dans la figure 6.4.



Espace de représentation couleur : (a) Image originale, (b) Espace RVB, (c) Espace TSL, (d) Espace YIQ.

B. Technique de seuillage global :

Le seuillage global consiste à partitionner l'image en deux classes grâce à un seuil optimal qui est calculé à partir d'une mesure globale sur toute l'image. L'histogramme est une mesure utilisée le plus souvent dans les méthodes de seuillage. Dans ce cas, le seuil attendu est celui qui correspond à la vallée de l'histogramme, celui qui distingue le plus possible les deux classes : fond et objet.

✓ Parmi les différentes méthodes de seuillage global qui existent dans la littérature et d'après les résultats obtenus on a choisi la méthode d'Otsu car elle est simple à implanter et donne de bons résultats en général sur les documents manuscrits couleurs.

✓ **Résultats obtenus :**

Nous avons appliqué la méthode de seuillage d'Otsu présentée précédemment, sur notre collection de test composée d'images d'anciens documents comportant la superposition de verso sur le recto. Nous présentons dans la figure ci dessous les résultats de la première partie de notre méthode.

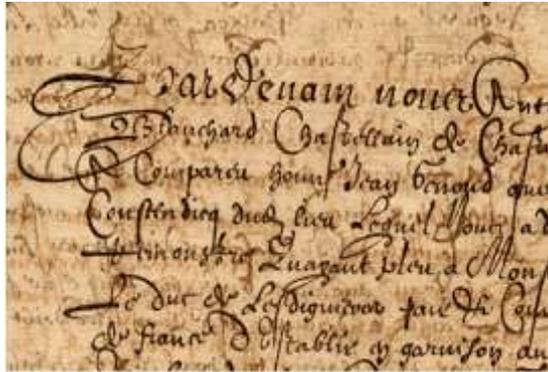


Image Originale en RVB.

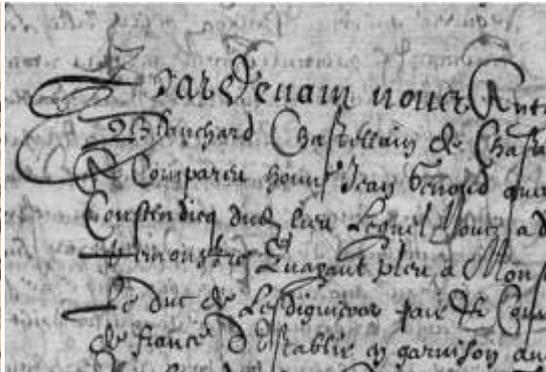
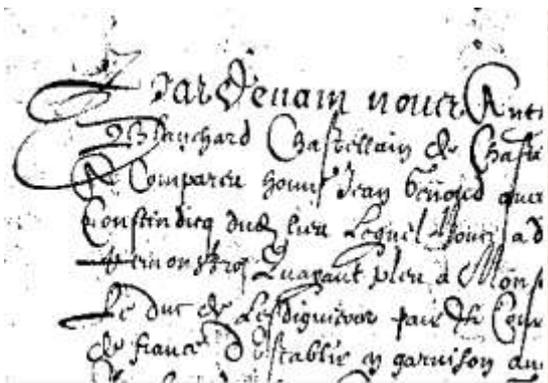
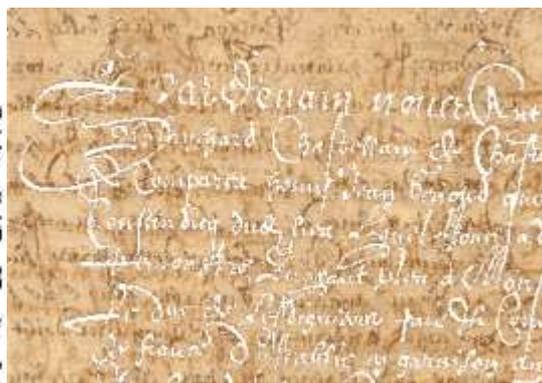


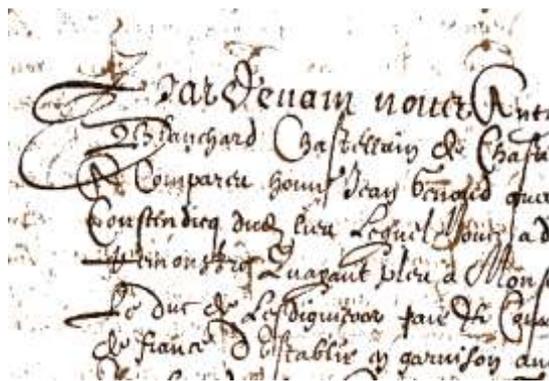
Image au niveau de gris.



Avant plan initial après Seuillage d'Otsu1.



Arrière plan initial après Seuillage d'Otsu1.

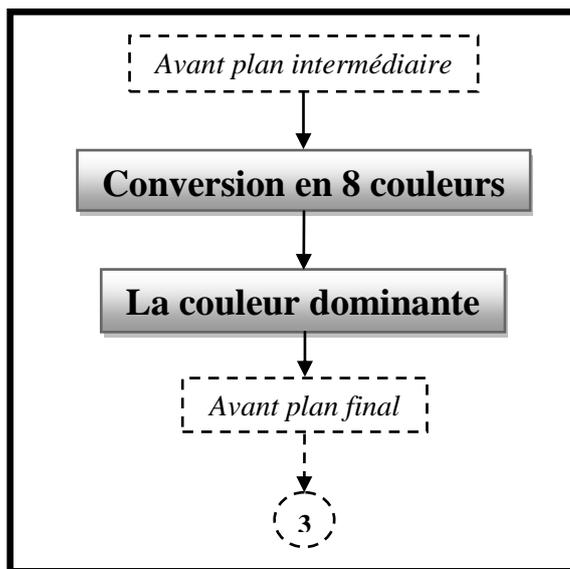


Avant plan Intermédiaire après Seuillage d'Otsu2.

Les résultats de la technique de Séparation proposée.

La deuxième partie : Le Raffinage (extraction de l'avant plan final).

- ✓ Après la séparation de l'objet (texte) et le fond, on prend l'image intermédiaire de l'avant plan, où il contient toujours des traces de la transparence qui ont l'intensité lumineuse moins faible par rapport au texte (niveau de gris différent) donc pour obtenir la couleur dominante de l'avant plan final, on commence par la conversion de l'image intermédiaire de l'avant plan en palette de 8 couleurs (Palette fixée) où on classe chaque pixel de cette image dans un tableau, par la suite on distingue la couleur dominante et on la garde comme couleur du texte final de l'image restaurée.
- ✓ La description de l'étape de **RAFFINAGE** (extraction de l'avant plan final) est présentée dans la figure :

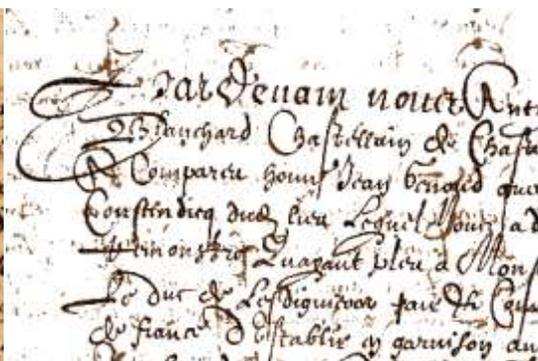


Technique d'extraction de l'avant plan final par notre méthode proposée.

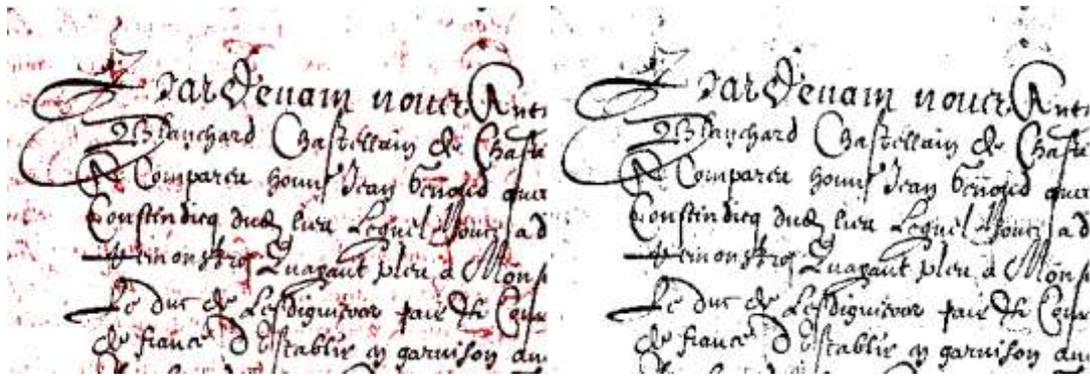
- ✓ **Résultats obtenus :** Nous avons classifié l'image de l'avant plan intermédiaire en palette fixée de 8 couleurs. Nous présentons dans la figure 6.3.2.b les résultats de la deuxième partie de notre méthode.



Image Originale en RVB



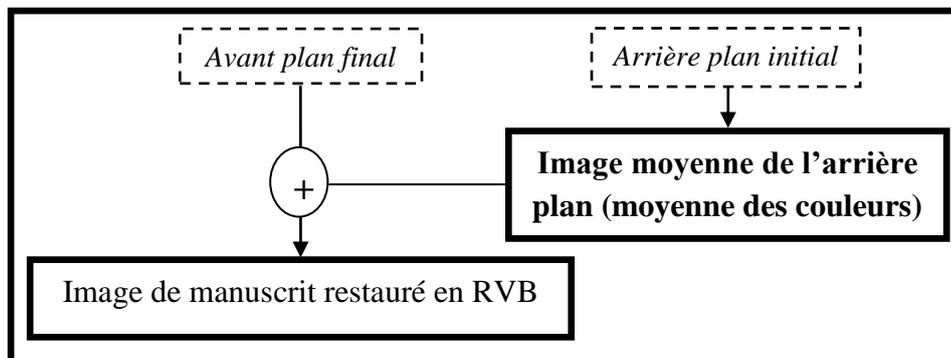
Avant plan Intermédiaire



Les résultats de la technique de Raffinage proposée.

La troisième partie : Reconstitution de l'image en remplaçant l'arrière plan bruité en une couleur uniforme (moyenne des couleurs).

- Tout d'abord on prend l'image de l'arrière plan initial où on calcule la moyenne des couleurs de cette dernière (image moyenne de l'arrière plan), combiné par superposition avec l'avant plan final pour obtenir l'image restaurée du document.
- La description de l'étape de **RECONSTITUTION** est présentée dans la figure suivante.



Technique de reconstitution de l'image restaurée de notre approche proposée.

✓ **Résultats obtenus :**

Pour obtenir l'image de l'avant plan final, on prend l'arrière plan initial et en calculant la moyenne des couleurs de ce dernier pour établir la couleur du fond de l'image restauré, on combine les deux images pour reconstituer l'image de manuscrit restaurée. Nous présentons dans la figure 6.3.3.b les résultats de la troisième partie de notre méthode.

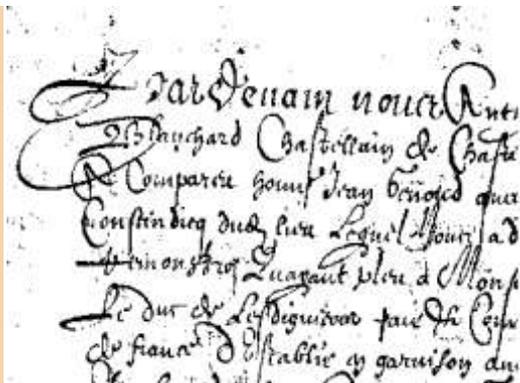


Image Originale en RVB

Arrière plan initial



Image moyenne de l'arrière plan final



Avant plan final

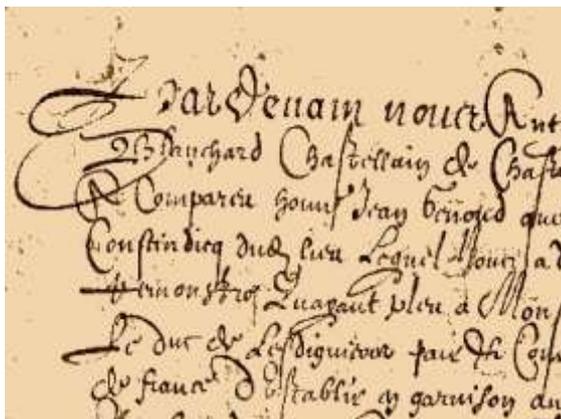


Image de manuscrit restauré en RVB.

Les résultats de la technique de Reconstitution proposée.

Expérimentation et résultats finaux:



Exemples d'images de documents de notre base de test.

Résultats obtenus :

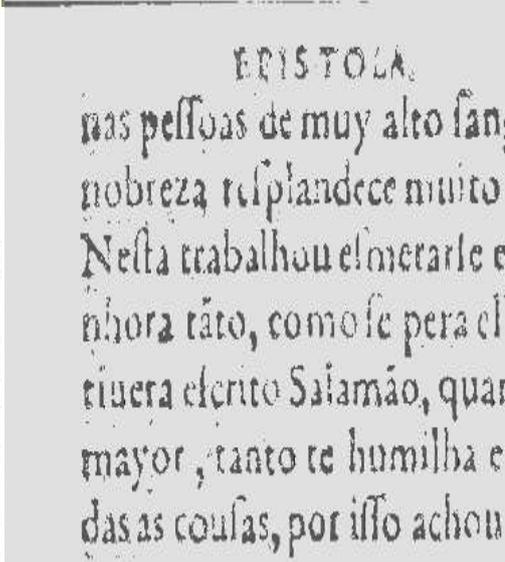
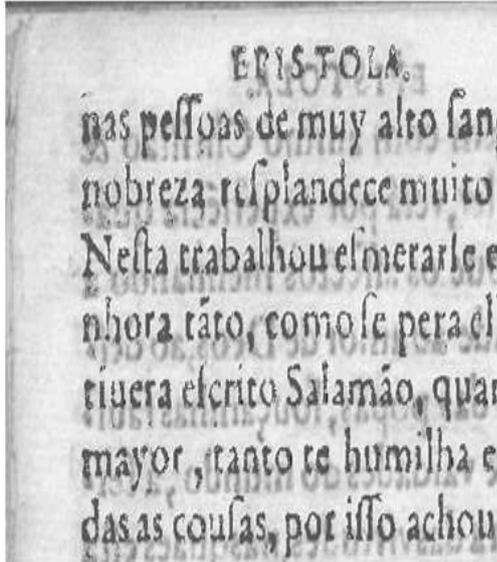
Nous avons appliqué notre approche présentée précédemment, sur notre collection de tests composée d'images de documents. Nous présentons les résultats des différents tests.

80 Years of
Bibliotheca
Islamica

From Manuscripts to Books:
Arab Heritage edited in Beirut

80 Years of
Bibliotheca
Islamica

From Manuscripts to Books:
Arab Heritage edited in Beirut



Images originaux

Images restaurées

les résultats des différents tests de notre approche.

7. BIBLIOGRAPHIE

Kricha A., Ghardallou Lasmar A., Essoukri Ben Amara N., *Exploration des Ondelettes en Prétraitement des Documents Anciens*. Dans Actes du 9ème Colloque International Francophone sur l'Écrit et le Document - (2006)

Boussellaa W., Zahour A., Taconet B., Benabdelhafid A., Alimi A., *Méthode hybride de séparation Avant/arrière-plan pour la restauration des manuscrits arabes anciens couleur*. In Actes du 9ème Colloque International Francophone sur l'Écrit et le Document - (2006)

Boulehmi H., Seddik B., Kricha A., Essoukri Ben Amara N., *Prétraitement de documents anciens*. Dans 10ème Colloque International Francophone sur l'Écrit et le Document, France (2008)

L.Likforman-Sulem, *Apport du traitement des images à la numérisation des documents manuscrits anciens*, Document Numérique, Vol. 7/3-4, 2003, p.13-26.

Li J., Gray M., "Context based Multiscale classification of document wavelets coefficients distributions", *IEEE Transactions on Image Processing*, Vol. NO. 9, September 2000

Menoti D., Borges D., Facon L., J. , Britto Jr A. S., "Segmentation of Postal Envelopes for Address Block Location: an approach based on feature selection in wavelet space", *IEEE proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)*, 0-7695-1960-1/03, 2003.

Smigiel E., Belaid A., and Hamza H., "Selforganizing Maps and Ancient Documents", S. Marinai and A. Dengel (Eds.): DAS 2004, LNCS 3163, pp. 125–134, Springer-Verlag Berlin Heidelberg, 2004.

DEBORA, Digital Access to Books of Renaissance, projet européen n°. LB 5608 A. Coordinateur R. Bouché, juin 2000.

Fadoua Drira, *Contribution à la restauration des images de documents anciens*. Thèse de doctorat, institut national des sciences appliquées de Lyon, Ecole doctorale informatique et information pour la société (EDIS) , 2007-ISAL-0111 (2007).

Terebes Romulus Mircea, *Diffusion directionnelle. Applications à la restauration et à l'amélioration d'images de documents anciens*. Thèse de Doctorat, Université Bordeaux I, Ecole doctorale des sciences physiques et de l'ingénieur. 2004.

Thibault Lelore encadré par José Martinez, *Segmentation d'image. Application aux documents anciens*. Mémoire de master de recherche « Architectures logicielles distribuées », Laboratoire des sciences de l'information et des systèmes, mai 2007

HAMMOUM Zoubida, Les services en ligne liés aux activités culturelles, <http://www.m-culture.gov.dz,17/05/2010>

(2)Algérie 360. Nouveaux manuscrits acquis par la Bibliothèque Nationale d'Algérie. Consultation (06 Mars 2010). Disponible sur Internet : <http://www.algerie360.com/algerie/164-nouveaux-manuscrits-acquis-par-la-bibliothequenationale-en-2009/>.

(1) La numérisation des documents anciens et son application au Centre Historique des Archives Nationales (CHAN): Archim

Contribution à la Restauration des Images de Documents Anciens, Thèse de doctorat, L'Institut National des Sciences Appliquées de Lyon, École Doctorale Informatique et Information pour la Société (EDIIS)

- J. André, M.A. Chabin. Numériser les documents anciens : et après ?, 1999, Les documents anciens, numéro spécial de Document Numérique, vol. 3, n°1-2, pp. 7-11.
- BoukerzazaKamel et BouderbaneAzzedine-Maîtres de Conférences- Les Manuscrits de la Bibliothèque Nationale d'Algérie : La Numérisation et la Recherche Scientifique, Département de Bibliothéconomie, Université de Constantine.
- B. Bringier. Élaboration de modèles d'évaluation de la qualité du rendu couleur, Thèse de doctorat, Université de Poitier, 2005.
- H. Emptoz, F. Le Bourgeois, V. Eglin, Y Leydier. La reconnaissance dans les images numérisées, OCR et transcription, reconnaissance des structures fonctionnelles et des méta-données, dans l'ouvrage La numérisation des textes et des images : Techniques et réalisations, éditions Presse de Lille 3, ISBN 2-84467-050-4, pp. 105-129, 2003.
- L. Bottou, P. Haffner, P.G. Howard, P. Simard, Y. Bengio, Y. LeCun, High quality document image compression with DjVu, journal of electronics imaging, 7(3) :410-428, 1998.
- F. Le Bourgeois, H. Emptoz. DEBORA: Digital AccEss to BOoks of the RenAissance, International Journal on Document Analysis and Recognition (IJ DAR), Special issue on the analysis of historical documents, Springer, Vo. 9, Nr 2-4, pp. 193-221, 2007.
- F. LEBOURGEOIS - H. EMPTOZ - F. DRIRA ; Séparation recto/verso des images de documents anciens par une approche « aveugle » ; Laboratoire d.InFoRmatique en image et Système d.Information (LIRIS) INSA de LYON- Bât Jules Verne 20 avenue Albert Einstein 69621 Villeurbanne Cedex France.
- G. Leedham, S. Varma, A. Patankar, V. Govindaraju, Separating text and background in degraded document images, *IWFHR*, 2002, pp.244. 249.
- G. Sharma. Cancellation of show-through in duplex scanning, Appears in Proceedings IEEE International Conference on Image processing, vol.2, pp.609-612, 2000.
- V. Rabeux, N. Journet, J.P. Domenger ; Evaluation de la transparence des documents anciens ; LaBRI September 21, 2010
- E. Dubois, A. Pathak, Reduction of bleedthrough in scanned manuscripts documents, *IS&T conference on image processing, image quality, image capture systems*, 2001.
- Q. Wang, T. Xia, C. L. Tan, L. Li. Directional Wavelet Approach to Remove Document Image Interference, ICDAR 2003, pp. 736-740, Edinburgh, 2003.
- C. Wolf. Document Ink bleed-through removal with two hidden Markov random fields and a single observation field, Technical Report RRLIRIS- 2006-019, 2006
- E. Smigiel, A. belaid, H. Hamza. Self-organizing Maps and Ancient Documents, Document Analysis Systems VI, 6th international workshop, pp.125-134, Florence, ITALY, 2004.
- U. Garain, T. Paquet, L. Heutte. On foreground-background separation in low quality document images, International Journal on Document Analysis and Recognition, vol. 8, no.1, pp. 47-63, 2006.

- Y Leydier., F. LeBourgeois, H. Emptoz. Serialized k-means for adaptative color image segmentation: Application to document images and others, Document Analysis Systems VI, 6th international workshop, pp. 252-263, Florence, ITALY, 2004.
- Kapur P. J.N. and Wong A.K.C., "A new method for gray-level picture thresholding using the entropy of the histogram", *Computer Vision, Graphics, and Image Processing*, 1985, vol. 29, pp.273-285.
- Kittler J. and Illingworth J., "Minimum error thresholding", *Pattern Recognition*, 1986, vol. 19, no.1, pp. 41-47.
- Leedham G., Varma S., Patankar A., and Govindaraju V., "Separating text and background in degraded document images - a comparison of global thresholding techniques for multi-stage thresholding", in *Proceedings Eighth International Workshop on Frontiers of Handwriting Recognition*, September 2002.
- Leedham G. Yan, C., Takru K., Tan Joie H. Nata and Mian L., "Comparison of Some Thresholding Algorithms for Text/Background Segmentation in Difficult Document Images", in *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, Janvier 2003.
- Leydier Y., Bourgeois F.L., and Emptoz H., "Serialized K-Means for Adaptive Color Image Segmentation-Application to Document Images and Others", in *6th International Workshop on Document Analysis systems (DAS)*, Itay, , 2004, LNCS vol.3163, 252-263.
- Mello C.A.B. and Lins R.D., "Image segmentation of historical documents", in *Visual2000, Mex-ico City, Mexico*, September 2000.
- Mello C.A.B. and Lins R.D., "Generation of images of historical documents by composition", in *ACM Symposium on Document Engineering*, McLean, VA, USA, 2002.
- Otsu N., "A threshold selection method from gray level histogram", *IEEE Transactions in Systems, Man, and Cybernetics*, 1979, vol. 9, pp. 62-66.
- Shi Z. and Govindaraju V., "Historical Document Image Enhancement Using Background Light Intensity Normalization", *17th International Conference on Pattern Recognition, Cambridge, United Kingdom*, 23-26 August 2004.
- Shi Z., S Setlur. and Govindaraju V., "Digital Enhancement of Palm Leaf Manuscript", *Based Computer Systems*, Hyderabad, India, December 19-22, 2004.
- Shi Z. and Govindaraju V., "Historical Document Image Segmentation Using Background Light Intensity Normalization", *SPIE Document Recognition and Retrieval XII*, San Jose, California, USA, 16-20 January 2005.
- Tan C. L., Cao R., Shen P., "Restoration of archival documents using a wavelet technique", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.24, No. 10, October 2002, pp. 1399-1404.

8- Bilan financier

Bilan financier : (exprimés en DA)

| <i>Intitulés des postes de dépenses par année</i> | Réalisé 1^{ère} année | Réalisé 2^{ème} année |
|---|--------------------------------------|--------------------------------------|
| Frais de séjour scientifique et de déplacement à l'étranger | | 88.627 |
| Frais de séjour scientifique et de déplacement en Algérie | | |
| Frais d'organisation de rencontres scientifiques | | |
| Honoraires des enquêteurs | | |
| Honoraires des guides | | |
| Frais de travaux et de prestations | | |
| Matériels et instruments scientifiques | | |
| Matériel informatique | 338.000 | 276.980 |
| Matériels d'expérience (animaux, végétaux, etc..) | | |
| Mobilier de bureau et de laboratoire | | |
| Entretien et réparation | | |
| Produits chimiques | | |
| Produits consommables | | |
| Composants électroniques, mécaniques et audio- visuels | | |
| Accessoires et consommables informatiques | | |
| Papeterie et fournitures de bureau | | |
| Périodiques | | |
| Ouvrages et documentation scientifiques et techniques | | |
| Logiciels | | |
| Impression et Edition | | |
| Affranchissements Postaux | | |
| Communications téléphoniques, Fax, Internet | | |
| Droits de douanes, Assurances | | |
| Carburant | | |
| Total | 338.000 | 365.607 |