

MA-004-88-1

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'enseignement supérieur et de la recherche scientifique

UNIVERSITE SAAD DAHLED BLIDA

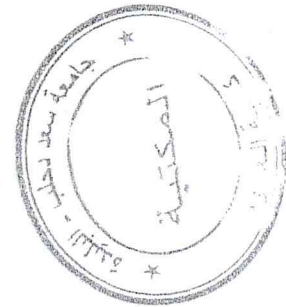
Faculté des sciences

Département d'informatique.



Mémoire Présenté par :

- Mokhbat Cherifa
- Sidi Moussa Zineb



Pour l'obtention du diplôme Master
Domaine : Mathématique et Informatique.

Filière : Informatique.

Spécialité : Informatique.

Option : Ingénierie du logiciel.

Sujet :

RESTAURATION DES DOCUMENTS

ANCIENS :

TEXTURE INPAINTING

Soutenu le : 1/07/2012 devant le jury composé de :

Mlle. Boustia	Président
M. Hamouda	Examineur
Mlle. Guendouz	Examineur
M. Hadj Yahia.O	Promoteur
Dr. Meziane Abdelkrim	Encadreur

Année universitaire 2011/2012

MA-004-88-1

ملخص

إن المسح الضوئي لكميات كبيرة من صور الوثائق القديمة خلق بيئة عمل جديدة بسبب الأضرار التي لحقت بهذه الوثائق. من بين هذه الأضرار سببها بالنقص الموجود في هذه المخطوطات نظرا لنوعية المواد الكيميائية للحبر و العوامل البيولوجية مثل القوارض و الفطريات اللاتي تخفض من سهولة قراءة المخطوطة .

مهمتنا هي الوقوف ضد هذه الأضرار من أجل إيجاد إقتراح جديد لترميم صور المخطوطات الملونة.

هذا النهج يتمثل في الجمع بين تقنية من أجل إيجاد النمذج المماثل الموجود في الرقعة التي على حواف البيكسل المتضرر و تقنية متابعة المحيطات من أجل الحصول على صورة مترجمة.

الكلمات الرئيسية:

المسح الضوئي ، مخطوطة، تدهور، ترميم، inpainting، أشكال مماثلة، محيط

RÉSUMÉ

La numérisation de grandes masses d'images de documents anciens a créé un nouveau contexte de travail à cause des dégradations apportées aux documents. Parmi ces dégradations, nous allons nous intéresser aux parties manquantes dans ces manuscrits dues à la qualité chimique de l'encre et aux agents biologiques, comme les rongeurs et les champignons qui diminuent la lisibilité des documents.

Notre travail consiste à faire face aux méfaits de détériorations, afin de trouver une nouvelle solution pour restaurer les images des manuscrits en couleurs.

L'approche consiste à combiner une technique qui permet de trouver la forme la plus similaire qui se trouve dans le patch choisis autour des bords du pixel dégradé, et la continuation des contours afin d'obtenir une image restaurée du document.

MOTS-CLÉS :

Numérisation, manuscrit, dégradation, restauration, inpainting, formes similaires, contour.



ABSTRACT:

The large mass of digitization of document's image had created a new context of work because of documents' damages. Among these degradations, we will be interested in the missing area in these manuscripts caused by the chemical quality of the ink and the biological agents such as rodents and fungi that decrease readability of documents.

Our job is to harm to these damages, to find a new solution for restoring colored images of manuscripts.

The approach is to combine technique of finding the similar form; which is chosen from the patch around the degraded pixel, and the continuation of the image's contour to obtain a restored image of the document.

KEYWORDS:

Digitization, manuscript, degradation, restoration, inpainting, similar forms, contour.

Remerciements

Tout d'abord le plus grand merci est réservé à notre dieu tout puissant qui nous a fait éclairer le chemin durant tout notre cursus.

Nous exprimons toute notre gratitude à monsieur Sadj Vahia, pour l'effort fourni, ses conseils, sa patience et sa persévérance dans le suivi.

Un grand merci pour notre encadreur monsieur Meziane au sein duquel nous avons trouvé tout ce qu'il nous fallait de mieux ainsi qu'à ma demoiselle Insaf qui nous a beaucoup aidé et enrichi nos informations.

Nous tenons aussi à remercier l'ensemble du département de l'informatique de nous avoir accepté, supporter et nous avoir donné tout ce qu'on avait besoin en matière d'approvisionnement en informations et études durant les 5 années de notre vie universitaire.

Un grand merci aussi est adressé au membre du jury qui nous a honorés de leur présence pour nous en juger et faire tous les points sur notre travail

A toute personne qui a participé de pré ou de loin pour l'achèvement de ce modeste travail.



Dédicace

Je dédie ce modeste travail à :

Mes parents, pour ses conseils.

Mon marie pour son aide et sa patience.

Mes frères et mes sœurs surtout ma sœur Sara la lumière de mon sentier.

Mes beaux et belles sœurs.

Mes amis surtout Amel.

Ma petite coquine de la famille Maria.

Mon binôme Cherifa.

Tous ceux qui m'ont prêté leurs cœurs, je leurs offre le mien.

Tous ceux qui me sont aidé, j'éterniserai de leur dire,

Que je les aime beaucoup.

Sans oublier la section 2ème année master Informatique

Promotion 2011/2012.

Sidi Moussa Zineb



Dédicace

Je dédie ce modeste travail à :

Mon cher père, ma fierté et ma dignité.

Ma chère mère, je n'aurais jamais pu réussir sans tes prières.

Mon cher grand-père, mon rayon de lumière qui éclaire ma vie.

Ma chère grand-mère, ma joie de vivre et mon rayon de soleil.

Mon cher frère, mon air que je respire.

Mes chères sœurs, mes bouquets de fleurs qui m'inspirent.

Ma chère Nina, la coquine de la famille.

Ma chère Lilia la fleur de toujours.

Mon cher frère Riadh pour ta tendresse infinie.

Ma chère Hanane pour ta gentillesse infinie.

Mon binôme Zineb.

Ma chère Zahra, mon amie de l'enfance.

Ma chère Amina ma chérie de toujours.

Toute la section de 2eme année Master du département d'Informatique

Mokhbat Cherifa



LISTE DES FIGURES

Figure 1.1 : Représentation binaire d'image	4
Figure 1.2 : Résolution d'image	5
Figure 1.3 : Dimensions d'une image numérique.....	5
Figure 1.4: L'histogramme d'une image	6
Figure1.5: Exemple d'image en niveau de gris	7
Figure1.6 : Représentation du codage RGB.....	9
Figure 1.7: Représentation du codage HSL.....	9
Figure 1.8: Représentation du codage CIE.....	10
Figure1.9 : Types de traitements d'image	13
Figure 1.10: Filtrage Passe-bas	15
Figure 1.11: Filtrage Passe haut.....	15
Figure 1.12: Filtrage Passe-bande.....	16
Figure1.13: Exemples de filtres utilisés pour la convolution.....	17
Figure 1.14 : Le principe de la convolution.....	17
Figure 2.1 : Evolution du support de manuscrits.....	19
Figure 2.2: Exemple de colophon en forme triangulaire.....	20
Figure 2.3 : Illustration de décors de manuscrits	21
Figure 2.4 : Exemple de manuscrits algériens	24
Figure 3.1 : Processus de numérisation.....	26
Figure 3.2 : Scanner à plat	27
Figure 3.3 : Scanner à défilement.....	27
Figure 3.4 : Scanner à main.....	28

Figure 3.5 : Scanner à diapositive	28
Figure 3.6 : Scanner à tambour.....	29
Figure 3.7 : Exemple des anciens documents dégradés.....	34
Figure 3.8 : Etat des manuscrits en Algérie	36
Figure 4.1 : Exemple de dégradations chimiques.....	41
Figure 4.2 : Exemple de dégradations biologiques.....	41
Figure 4.3 : Document dégradé par l'acidité de l'encre	42
Figure 4.4 : Exemple de dégradations d'origine des sinistres naturels ou accidentels.....	43
Figure 4.5 : Exemple de dégradation d'origine humaine	45
Figure 4.6 : Exemple de dégradations matérielles : Perte d'information couleur sur les images numérisées des microfilms.....	45
Figure 4.7 : Finesse de la grille pour une faible et une forte résolution	46
Figure 4.8 : Exemple de problème de la numérisation en binaire : Binarisation de documents dont le support est dégradé.....	47
Figure 4.9 : Exemple de dégradations du fond.....	47
Figure 4.10 : Exemple de dégradations des formes	48
Figure 4.11 : Une lampe à fluorescence équipée de filtres anti-UV (ultra-violetée)	50
Figure 4.12 : Méfaits de la température sur les documents	52
Figure 5.1 : Suppression des taches en transparence par recalage	59
Figure 5.2 : Exemple2 de suppression des taches en transparence par recalage	59
Figure 5.3 : Suppression des taches en transparence par analyse de l'orientation.....	60
Figure 5.4 : (a) Image d'origine, (b) Image finale de recto séparée du verso.....	60
Figure 5.5 : Séparation recto/verso par segmentation et agrégation de régions.....	61
Figure 5.6 : Application de l'approche de [33] : (a) Image originale (b) Résultat pour K=2, le nombre de classes.....	62
Figure 5.7 : Principe de la restauration d'une lettre cassée par les contours actifs	63

Figure 5.8 : Cas de dégradations de la lettre «e»	63
Figure 5.9 : Restauration par prototypage des formes dégradées de lettres	64
Figure 5.10 : Limite d'une restauration par prototypage.....	64
Figure 5.11 : Défauts géométriques de la photographie d'un livre ouvert à gauche et sa version restaurée à droite.....	65
Figure 5.12 : Suppression des taches en transparence par analyse de la couleur	67
Figure 5.13 : Principe de l'analyse d'image par la morphologie mathématique.....	67
Figure 5.14 : Restauration de la dégradation colorimétrique par ajustement du contraste..	69
Figure 6.1 : Principe de reconstruction par jonction des isophotes.....	71
Figure 6.2 : Direction de la propagation des informations situées le long du contour.....	72
Figure 6.3 : Limite d'une reconstruction par la totale variation.....	73
Figure 6.4 : Les avantages des modèles de diffusion dirigée par la courbure (CDD) et de l'élastique d'Euler par rapport à la méthode de diffusion basée sur la totale variation (TV)..	74
Figure 6.5 : Image originale (IRHT) et superposition du masque en blanc.....	74
Figure 6.6 : Reconstruction par inpainting avec le modèle de Beltramio	75
Figure 6.7 : Reconstruction par inpainting avec le modèle de Chan et l'élastique d'Euler	75
Figure 6.8 : Limite de l'inpainting par manque de continuité dans les isophotes.....	76
Figure 6.9 : Limites de l'inpainting pour un masque plus large que l'épaisseur des traits	76
Figure 6.10 : De gauche à droite respectivement : Forme régulière que l'inpainting pourra reconstruire et Formes irrégulières que les modèles actuels ne pourront pas reconstruire	76
Figure 6.11 : Différence entre la restauration par diffusion et par Exemplar-based.....	78
Figure 6.12 : Résultat de l'algorithme de synthèse de texture	80
Figure 6.13 : Exemples de dégradations obtenues à partir d'un caractère 'R' écrit en Romain.....	80
Figure 7.1 : Diagramme d'inpainting	82
Figure 7.2 : Comparaison du modèle d'inpainting par notre approche sur une police Arial..	87

Figure 7.3 : Comparaison du modèle d’inpainting par notre approche sur une police Andalouse.....	87
Figure 7.4 : Comparaison du modèle d’inpainting par notre approche sur une police Microsoft Sans Serif	87
Figure 7.5 : Suppression avec le modele « exemplar-based »	88
Figure7.6 : Image originale1 et superposition du masque en gris	88
Figure7.7 : Reconstruction avec notre approche	88
Figure7.8 : Image originale2 et superposition du masque en gris	89
Figure7.9 : Reconstruction avec notre approche de l’exemple 2.....	89
Figure 7.10 : Reconstruction avec notre approche avec les filtres Sobel et Kernel de l’exemple de la figure7.2.....	89
Figure 7.11 : Reconstruction avec notre approche avec les filtres Sobel et Kernel de l’exemple de la figure7.8.....	90
Figure7.12 : Exemple de reconstruction de manuscrit écrit avec le latin	90
Figure8.1 : Aperçu de l’application	93
Figure8.2 : Le 1 ^{er} menu de l’application	94
Figure8.3 : Le 2eme menu de l’application	94
Figure8.4 : Le 3 ^{eme} menu de l’application.....	95
Figure8.5 : Le 3 ^{eme} menu de l’application.....	95

LISTE DES TABLEAUX

Tableau 4.1 : Conditions relatives à la température et l’humidité et l’éclairage.....	49
Tableau 4.2 : Conditions relatives à la pollution et les agents chimiques et biologiques et choc	49

TABLE DES MATIÈRES

Introduction générale	1
-----------------------------	---

Partie1: GENERALITES

Chapitre1:Généralités sur le traitement d'image.....	3
1. Introduction	3
2. Image et image numérique	3
3. Carracteristiques de l'image numérique.....	4
3.1.Pixel.....	4
3.2.Résolution.....	4
3.3.Dimension.....	5
3.4.Histogramme	5
3.5.Luminance	6
3.6.Contraste.....	6
3.7.Niveau de gris	7
3.8.Région.....	7
3.9.Contour	7
3.10. Bruit.....	7
4. Les representations des couleurs	8
4.1.Le codage RGB	8
4.2.Le codage HSL	9
4.3.Le codage CMY.....	10
4.4.Le codage CIE	10
4.5.Le codage YUV	11
4.6.Le codage YIQ.....	11

5. Les differents types d'images numeriques	11
5.1.Les formats vectoriels.....	12
5.2.Les formats bitmap.....	12
5.3.Les formats meta fichiers.....	12
6. Le traitement d'image.....	12
6.1.Introduction au traitement d'image	12
6.2.Types de traitements d'image.....	13
6.2.1. Traitement de bas niveau.....	13
6.2.2. Traitement de haut niveau	13
6.3.Etapes de traitement d'image.....	13
6.3.1. Aquisition	13
6.3.2. Prétraitement	14
6.3.2.1.Traitement numerique des images.....	14
6.3.2.2.Filtrage.....	14
6.3.3. Analyse d'image.....	16
7. Quelques traitements d'image	16
7.1.Le gradient.....	16
7.2.La convolution.....	16
8. Conclusion.....	17
Chapitre2:Généralités sur les manuscrits	18
1. Introduction	18
2. Definition du manuscrit.....	18
3. Defintion du codex	19
4. Les elements descriptifs des manuscrits.....	20
4.1.La mention de responsabilité.....	20
4.2.Le nom du proprietaire	20

4.3.Le tittre du manuscrit.....	20
4.4.Le colophon	20
4.5.Le décor	20
4.6.La reliure.....	21
4.7.Organisation du volume.....	21
4.8.Type de support	22
5. Etat des manuscrits en Algerie	22
5.1.La bibiotheque nationale d’Algerie (BNA)	22
5.2.Le centre nationnal des manuscrits D’Adrar (CNMA)	23
6. Conclusion.....	24
Chapitre3:La numérisation	25
1. Introduction	25
2. La numerisation des documents du patrimoine	25
2.1.La numerisation	26
2.2.Types de scanners utilisés pour la numerisation.....	27
2.2.1. Scanner à plat	27
2.2.2. Scanner à defilement.....	27
2.2.3. Scanner à main.....	37
2.2.4. Scanner à diapositive.....	28
2.2.5. Scanner à tombours	28
2.3. L’importance de la numerisation	29
2.3.2. Pourquoi la numérisation est devenue si populaire ?.....	29
2.3.1. Pourquoi numeriser.....	30
2.3.2. Incovenients, limites et risques de la numerisation	32
2.3.3. Quelques grands programmes de numerisation	33
3. La numerisation du patrimoine en Algerie	34

3.1. Bibliothèque nationale d'Algérie.....	34
3.2. Centre national de manuscrit d'Adrar.....	35
3.3. Centre algérien de cinématographie.....	36
4. Conclusion.....	36

Partie 2 : DEGRADATIONS ET RESTAURATIONS DES IMAGES DES DOCUMENTS

Chapitre 4: Degradations des images des documents.....37

1. Introduction.....	37
2. Qu'est ce qu'une dégradation?.....	37
3. Les agents de dégradations.....	37
3.1. Les agents de dégradations biologiques.....	38
3.1.1. Les champignons.....	38
3.1.2. Les bactéries.....	38
3.1.3. Les insectes.....	39
3.1.4. Les rongeurs.....	39
3.2. Les agents de dégradations physico-chimiques.....	39
3.2.1. La lumière.....	39
3.2.2. L'humidité et la température.....	40
3.2.3. La pollution atmosphérique.....	41
3.2.4. Dégradations dues à la capture.....	42
3.3. La mauvaise qualité des matériaux constitutifs des documents.....	42
3.4. Les sinistres naturels ou accidentels.....	43
3.4.1. Les inondations et cyclones.....	43
3.4.2. Les vents de sable.....	43
3.4.3. Le feu.....	43
3.4.4. L'eau.....	44
3.5. Les méfaits de l'homme.....	44
3.5.1. Le vandalisme et le manque de soin des lecteurs ou visiteurs.....	44
3.5.2. Les manipulations maladroitement.....	44
3.6. Les agents de dégradations dues à la capture.....	45

3.6.1.	Degradations matérielles.....	45
3.6.2.	Degradations numériques.....	45
3.6.2.1.	L'inadéquation de la résolution à la taille des textes.....	46
3.6.2.2.	Le problème de la résolution en binaire	46
4.	Typologie des dégradations.....	47
4.1.	Dégradation du fond (image du papier)	47
4.2.	Dégradation des formes (images des traits, caractères, dessins).....	47
5.	Mesures de prévention	48
5.1.	Recommandations en matière de conservation des documents	48
5.2.	Prévention des dégradations dues à l'homme.....	49
5.3.	Prévention des désastres	49
5.4.	Prévention de la pollution	50
5.5.	Prévention des altérations dues à la lumière	50
5.6.	Prévention des altérations dues à la température et l'humidité relative	51
5.6.1.	Influence de la température	51
5.6.2.	Influence de l'humidité relative	52
5.6.2.1.	humidité relative trop élevée	53
5.6.2.2.	humidité relative trop basse	53
5.6.2.3.	humidité relative fluctuante	53
5.6.3.	La conservation dans des régions climatiques difficiles : le cas des bibliothèques des zones tropicales	54
6.	Conclusion	56
	 Chapitre5: Restaurations des images des documents	57
1.	Introduction	57
2.	Type de restauration	57
2.1.	La restauration physique	57
2.2.	La restauration virtuelle	57
2.3.	La restauration numérique	57
3.	Méthodes de restauration supervisées :.....	58
3.1.	Séparation recto/verso par recalage	58
3.2.	Séparation recto/verso par analyse de l'orientation	59

3.3.Séparation recto/verso par analyse des formes.....	60
3.4.Séparation recto/verso par analyse de la couleur.....	60
3.5.Restauration des formes par modèle.....	62
3.6.Restauration des formes par prototypage	63
3.7.Restauration géométrique de la courbure	64
3.7.1. Restauration par un modèle 2D	65
3.7.2. Restauration par un modèle 3D	65
4. Méthodes de restauration non-supervisées.....	66
4.1.Restauration du recto/verso par séparation de sources.....	66
4.2.Restauration des formes par morphologie mathématique	67
4.3.Restauration Colorimétrique	68
5. Conclusion	69

Chapitre6: Texture Inpainting

1. Introduction	70
2. Problématique	70
3. L'état de l'art	70
3.1.Qu'est ce que l'inpainting ?	70
3.2.Les méthodes de l'inpainting.....	71
3.2.1. Reconstruction méthodes variationnelles et équations différentielles partielles(EDP)	71
3.2.1.1.Etapes de l'inpainting	71
3.2.1.2.Expérimentation.....	74
3.2.1.3.Discussion sur la méthode de reconstruction par EDP	75
3.2.2. Méthodes basées sur la recherche de similarités (« Exemplar-based »)	77
3.2.2.1.Etapes de l'inpainting	77
3.2.2.2.Discussion sur la méthode de reconstruction par Exemplar-based.....	78
3.2.3. Reconstruction par apprentissage	78
3.2.3.1.Les champs de Markov (Markov Random Field)	79
3.2.3.2.L'approche par les contours actifs guidée par un modèle	80
3.2.3.3.Discussion de la méthode d'inpainting par apprentissage	80
4. Conclusion.....	81

Partie3 : METHODOLOGIE ET IMPLEMENTATION

Chapitre7:Methodologie	82
1. Introduction	82
2. Méthodologie	82
2.1.Principe	82
2.1.1. Phase 1	83
2.1.2. Phase 2	83
2.1.3. Phase 3	84
3. Expérimentation et évaluation.....	86
3.1.Expérimentation sur des images de synthèse.....	86
3.2.Expérimentation sur des images réelles de documents.....	98
4. Discussion sur notre approche.....	90
5. Conclusion.....	91
Chapitre8:Implementation	92
1. Introduction	92
2. Outils de travail	92
2.1.Choix du langage	92
2.1.1. Aperçu	92
2.1.2. Avantages de java	92
2.2.Choix de l'environnement de développement	93
2.2.1. Aperçu	93
2.2.2. Objectif d'Eclipse	93
3. L'application.....	93
4. Conclusion.....	95
Conclusion Generale	96

INTRODUCTION GENERALE:

Le progrès de nos jours est exceptionnellement hors de toutes les pensées. Ce progrès croissant n'aurait pu atteindre ce niveau sans l'aide des anciennes réflexions et de la succession des sciences à travers des générations toutes entières. Ces travaux antérieurs étaient plutôt écrits à la main sur des documents et des manuscrits, ce qui a amené à l'apparition d'une multitude de problèmes lors de leurs découvertes que ce soit sur leur vue d'ensemble à l'œil nu qu'à l'intérêt à leur étude, leur analyse et leur compréhension. Des dégradations partielles et parfois quasi totale sont souvent un frein à l'exploitation de ces documents. Manque d'une lettre, d'une phrase ou même d'un paragraphe tout entier sont des exemples de dégradations qui atteignent les documents anciens. Pour palier à ces problèmes, plusieurs travaux ont été entamés ; parmi ceux là, les travaux inspirés des restaurations de vieilles peintures qui visaient alors à rendre la vie qui s'était enfui au cœur de ces richesses culturelles.

Notre travail consiste à bien étudier le comment, le pourquoi et le par quoi peut on faire renaître les manuscrits arabes anciens en Algérie, à distinguer les différents outils qui permettent de faire ceci et à choisir la meilleure méthode et approche pour palier à l'un des multiples problèmes et difficultés que connaît le manuscrit arabe algérien. Effectivement, le manuscrit court plusieurs types de dégradations, la plupart des dégradations sont physiques ce qui implique une interaction humaine et des procédés parfois chimiques pour le restaurer. En revanche, lorsque la restauration physique est impossible, les outils informatiques interviennent. Numériser le manuscrit et le restaurer numériquement permet de conserver le document d'origine, d'éviter de le manipuler et de restaurer son contenu sans pour autant le dégrader d'avantage. En plus d'éviter des dégradations supplémentaires, la restauration des manuscrits permet de les ouvrir à un plus grand public et à présenter une richesse sans égale sans l'abimer.

Pour arriver à notre résultat, une méthode de travail a été suivie. Les parties et chapitres de ce mémoire résument notre travail.

Partie1: GENERALITES. Cette partie contient les chapitres suivants :

- Le chapitre 1: Dénommé généralités sur l'analyse et le traitement d'image, contient les notions de bases sur l'image et le traitement d'images.
- Le chapitre 2: Dénommé généralités sur les manuscrits définit ce qu'est un manuscrit et les différents caractéristiques de ce dernier. Un état de l'art sur les manuscrits algériens sera à la fin présenté.
- Le chapitre 3 : Dénommé la numérisation définit la numérisation et son grand enjeu envers la conservation des informations issues des manuscrits. A la fin de ce chapitre, une petite étude sur l'état du patrimoine en Algérie est présentée.

Partie 2 : DEGRADATIONS ET RESTAURATIONS DES IMAGES DES DOCUMENTS. Cette partie est divisée en 3 chapitres :

- Le chapitre 4 : Dénommé dégradations des images des documents contient les dégradations présentes sur les différents manuscrits, leurs causes principales, une typologie et les mesures aidant à la préservation de l'état des documents.
- Le chapitre 5 : Restauration des documents contient les différents types de restaurations existantes aidant à faire renaître le document afin de bien extraire l'information.
- Le chapitre 6 : Texture Inpainting contient un état de l'art sur les différentes méthodes qui existent pour le remplissage de la partie manquante de l'information sur les images des manuscrits et une comparaison entre les méthodes existantes est réalisée.

Partie3 : METHODOLOGIE ET IMPLMENTATION .Cette partie contient :

- Le chapitre 7 : Méthodologie contient la présentation de notre proposition pour traiter les parties manquantes des documents anciens et les expériences de l'approche choisie pour remplir ces parties manquantes.
- Le chapitre8 : Implémentation contient la présentation des outils de travail comme le langage de programmation et l'environnement de développement ainsi que notre architecture d'application.

Des perspectives de notre projet seront enfin présentées.

LA PREMIÈRE PARTIE :
GÉNÉRALITÉS

CHAPITRE 1 :
GÉNÉRALITÉS SUR
L'ANALYSE ET LE
TRAITEMENT
D'IMAGE

Sommaire :

Dans ce chapitre nous présentons les différentes notions de base sur l'image, leurs traitements numériques ainsi que les domaines d'application.

Notre chapitre se divise en :

- 1- Introduction
- 2- Image et image numérique
- 3- Caractéristiques de l'image numérique
- 4- La représentation de la couleur
- 5- Les différents types de l'image numérique
- 6- Traitement d'image
- 7- Quelques traitements d'image
- 8- Conclusion

1. Introduction :

L'image est considérée aujourd'hui comme le moyen le plus utilisé dans la communication humaine, partout dans le monde l'être humain a laissé des traces de ses facultés imaginaires sous forme de dessin sur les rochers, qui vont depuis le temps ancien jusqu'au temps moderne. Ces dessins étaient et ils resteront utilisés pour communiquer des messages et dégager de l'information.

Pour cela le traitement d'image est le moyen le plus efficace pour améliorer l'aspect visuel de l'image afin d'extraire une information précise et pertinente.

2. Image et image numérique [1,2]: Nous pouvons trouver de nombreuses définitions du mot « IMAGE », parmi ces définitions nous pouvons citer :

- Une image est une représentation visuelle de quelque chose (objet, être vivant), elle peut être naturelle (ombre, reflet) ou artificielle (dessin, peinture, photographie, sculpture, film) qui a une signification à l'œil. On peut la décrire par une fonction L de brillance analogique définie dans un domaine borné tel que x, y sont les coordonnées spatiales d'un point de l'image et I est une fonction d'intensité lumineuse et de couleur.

- Une image est une reproduction exacte ou une représentation analogique d'un objet.
- Mathématiquement, une image plane, généralement un rectangle, constitue un ensemble de points distincts appelés pixels.
- Les images numériques sont des clichés électroniques d'une scène ou numérisés à partir de documents tels que photographies, manuscrits, textes imprimés, et œuvres d'art. L'image numérique est échantillonnée et mappée comme une grille de points ou éléments d'images.

3. Caractéristique de l'image numérique: Une image se caractérise par plusieurs paramètres dont :

3.1. Pixel (Picture élément ou pixel) [1]:

Il est le plus petit élément qui constitue une image numérique, il est représenté comme une grille de points ou éléments d'images. A chaque pixel correspond une valeur tonale (noir, blanc, niveaux de gris ou couleur), exprimée en code binaire (zéros et uns). Les chiffres binaires ("bits") de chaque pixel sont stockés dans une séquence par l'ordinateur : un octet pour la représentation monochrome et dans une image de couleurs RVB le pixel est représenté sur 3 octets, un octet pour chacune des couleurs : rouge, bleu, vert.

Valeurs des Pixels:

Comme indiqué dans l'image bitonale, chaque pixel se voit assigné une valeur tonale, dans ce cas 0 pour le noir et 1 pour le blanc.

1			1	1	1			1	
1	1		1	1	1	1		1	1
1	1		1	1	1	1		1	1
1	1		1	1	1	1		1	1
1	1							1	1
1	1		1	1	1	1		1	1
1	1		1	1	1	1		1	1
1	1		1	1	1	1		1	1
1			1	1	1			1	

Figure 1.1 : Représentation binaire d'image

3.2. Résolution [2]: Elle est la capacité à distinguer les détails fins dans l'espace. La fréquence spatiale à laquelle une image numérique est échantillonnée, elle est généralement un bon indicateur de la résolution. C'est pourquoi les termes points par

pouce ou pixels par pouces (ppp ou dpi et ppi en anglais) sont les expressions courantes et synonymes indiquant la résolution des images numériques. En général, l'augmentation de la fréquence d'échantillonnage augmente la résolution, mais seulement jusqu'à un certain point.

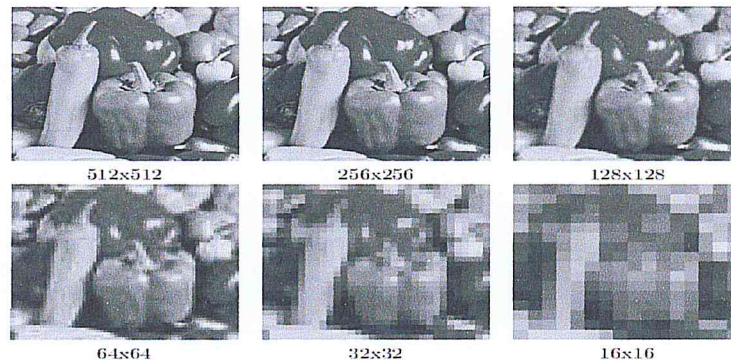
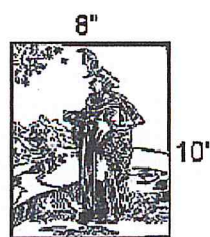


Figure 1.2 : Résolution d'image

3.3. Dimension [4] : C'est la taille de l'image et les mesures horizontales et verticales d'une image exprimées en pixels. Les dimensions en pixels peuvent être déterminées en multipliant la largeur et la longueur de l'image par le dpi. Un appareil photo numérique possède également des dimensions en pixels, exprimés par le nombre de pixels horizontaux et verticaux définissant sa résolution (p.ex. 2048 par 3072). Le calcul de la résolution se fait en divisant une des dimensions en pixels du document par la dimension en pouces correspondante.



Un document de 8" x 10" scanné à 300 dpi a une dimension en pixels de 2400 pixels (8" x 300 dpi) par 3000 pixels (10" x 300 dpi).

Figure 1.3 : Dimensions d'une image numérique

3.4. Histogramme [5] : L'histogramme de couleur est une observation statique qui permet de donner la fréquence de chaque niveau de couleur dans une image, il est utilisé pour diminuer l'erreur de quantification et pour comparer deux images obtenues dans des éclairages différents. Il permet de donner un grand nombre

d'informations sur les distributions des niveaux de couleurs et de voir entre quelle borne sont réparties les couleurs.

La forme de l'histogramme est significative car elle peut être utilisée pour améliorer la qualité d'une image.

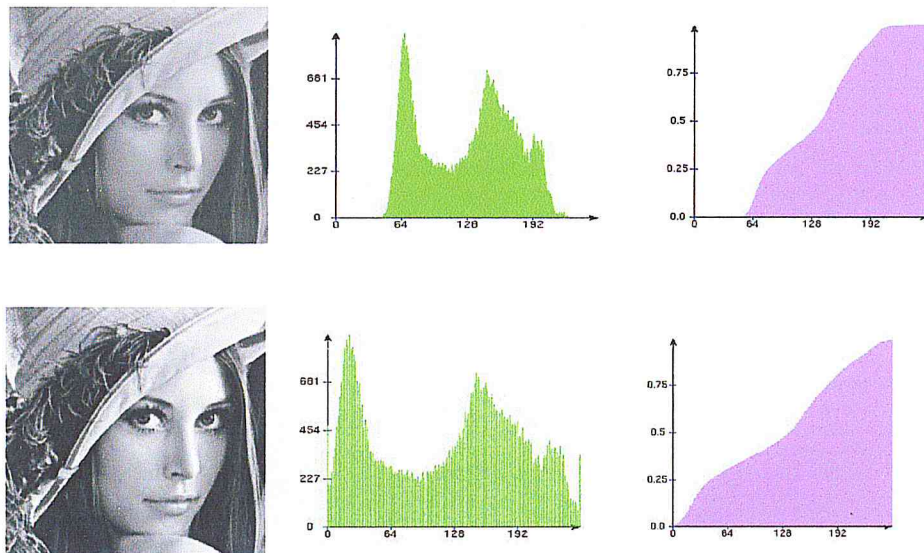


Figure 1.4: L'histogramme d'une image

3.5. Luminance [7]: C'est le degré de luminosité des points de l'image. Elle est définie aussi comme étant le quotient de l'intensité lumineuse d'une surface par l'aire apparente de cette surface pour un observateur lointain. Une bonne luminance se caractérise par :

- Des images lumineuses.
- Un bon contraste : il faut éviter les images où le contraste (défini un peu plus loin) tend vers le blanc et le noir, ces images entraînent des pertes de détails dans les zones sombres ou lumineuses.
- L'absence de parasites.

3.6. Contraste [6]: C'est l'opposition marquée entre les régions sombres et les régions claires de l'image. Le contraste est défini en fonction de la luminance de deux zones d'images.

Si L_1 et L_2 sont les degrés de luminosité respectivement de 2 zones voisines A_1 et A_2 d'une image, le contraste C est définie par le rapport : $C=L_1-L_2/L_1+L_2$.

3.7. Niveau de gris [3]: Le niveau de gris est la valeur de l'intensité lumineuse en un point. La couleur de pixel peut prendre des valeurs allant du noir au blanc en passant par un nombre fini de niveaux intermédiaire.

Pour ce type d'images le travail se fait avec 256 valeurs finies(figure1.5),pour 3 couleurs différentes le rouge, le vert et le bleu qui ont des valeurs égales et cela pour les 256 valeurs définies, ce qui donne une tonalité entre le noir et le blanc, quand R, V et B ont pour valeur 0, la valeur ainsi définie est un noir profond, quand les trois part de couleurs ont pour valeur 256, c'est un blanc pure qui est défini, une valeur intermédiaire, l'exemple 192,192,192,défini une couleur grise d'une certaine luminosité. La palette de l'image en niveau de gris est constituée de 256 couleurs comprises entre 0, 0,0 et 255, 255,255 qui sont sauvegardées, donc 256 nuances de gris. Voici un exemple d'image en niveau de gris.



Figure1.5: Exemple d'image en niveau de gris

3.8. Région [12]: C'est un ensemble de pixels connexes et homogènes .Si un pixel appartient à une région donnée alors il doit vérifier les caractéristiques de cette dernière (intensité, centre de gravité..).Une région est limitée par un contour.

3.9. Contour [12]: C'est un ensemble de pixels formant une frontière entre deux ou plusieurs régions voisines.

3.10. Bruit (parasite) [12] : Il est considéré comme un phénomène de brusque variation de l'intensité d'un pixel par rapport à ses voisins, il provient de l'éclairage des dispositifs optiques et électroniques mais aussi de l'objet lui-même. Prenons comme exemple une surface de gris rigoureusement homogène, l'image numérique

de cette surface devrait être une matrice dont tous les éléments ont la même valeur, or on peut trouver deux pixels voisins qui n'ont pas la même luminance.

4. Les représentations de la couleur [5] : Afin de pouvoir bien manipuler les images en couleurs pour bien extraire l'information, il est nécessaire de choisir la ou les couleurs indispensables à la représentation de l'image. Dans ce cas, une palette de couleurs est présentée, dans laquelle la couleur convenant le mieux aux besoins est choisie.

En informatique, de la même façon, il est essentiel de disposer d'un moyen pour choisir une couleur parmi toutes celles utilisables. Or la gamme de couleurs possibles est très vaste et la chaîne de traitement de l'image passe par différents périphériques : par exemple un numériseur (scanner), puis un logiciel de retouche d'images et enfin une imprimante. Il est donc nécessaire de pouvoir représenter de façon fiable la couleur afin d'assurer la cohérence entre ces différents périphériques.

On appelle ainsi «espace de couleurs» la représentation mathématique d'un ensemble de couleurs. Il en existe plusieurs, les plus connus sont :

4.1. Le codage RGB : Il consiste à représenter l'espace des couleurs à partir de trois rayonnements monochromatiques de couleurs :

- rouge (de longueur d'onde égale à 700,0 nm).
- vert (de longueur d'onde égale à 546,1 nm).
- bleu (de longueur d'onde égale à 435,8 nm).

Ainsi, le modèle RGB propose de coder sur un octet chaque composante de couleur, ce qui correspond à 256 intensités de rouge (2^8), 256 intensités de vert et 256 intensités de bleu, soient 16777216 possibilités théoriques de couleurs différentes, c'est-à-dire plus que l'œil humain n'en peut discerner (environ 2 millions). Toutefois, cette valeur n'est que théorique car elle dépend fortement du matériel d'affichage utilisé.

Le codage RGB est dit aussi codage à synthèse additive car on obtient les autres couleurs par addition de ces couleurs.

Etant donné que le codage RGB repose sur trois composantes proposant la même gamme de valeur, on le représente généralement graphiquement par un cube dont chacun des axes correspond à une couleur primaire [5].

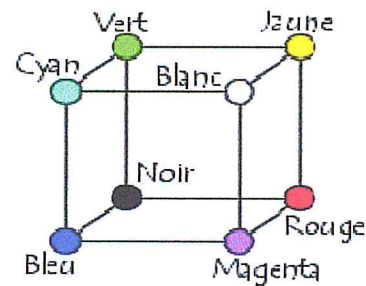


Figure1.6 : Représentation du codage RGB

4.2. Le codage HSL :(Hue, Saturation, Luminance, ou en français TSL)

C'est un modèle de représentation dit "naturel", c'est-à-dire proche de la perception physiologique de la couleur par l'œil humain. Il consiste à décomposer la couleur selon des critères physiologiques :

- **La teinte** :(en anglais Hue) Correspondant à la perception de la couleur (T-shirt mauve ou orange).
- **La saturation** : Décrivant la pureté de la couleur, c'est-à-dire son caractère vif ou terne (T-shirt neuf ou délavé).
- **La luminance** : Indiquant la quantité de lumière de la couleur, c'est-à-dire son aspect clair ou sombre (T-shirt au soleil ou à l'ombre).

Voici une représentation graphique du modèle HSL, dans lequel la teinte est représentée par un cercle chromatique, la luminance et la saturation par deux axes :

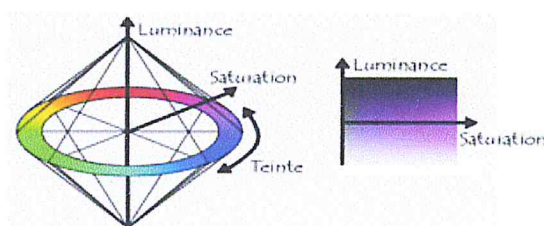


Figure 1.7: Représentation du codage HSL

Le modèle HSL a été mis au point dans le but de permettre un choix interactif rapide d'une couleur, pour autant il n'est pas adapté à une description quantitative d'une couleur.

Il existe d'autres modèles naturels de représentation proches du modèle HSL :

- **HSB** : Hue, Saturation, Brightness soit Teinte, Saturation, Brillance en français. La brillance décrit la perception de la lumière émise par une surface.
- **HSV** : Hue, Saturation, Value soit Teinte, Saturation, Valeur en français
- **HSI** : Hue, Saturation, Intensity soit Teinte, Saturation, Intensité en français
- **HCI** : Hue, Chrominance, Intensity soit Teinte, Chrominance, Intensité.

4.3. Le codage CMY : Le codage CMY (Cyan, Magenta, Yellow, ou Cyan, Magenta, Jaune en français, soit CMJ) est à la synthèse soustractive, par exemple si on soustrait la lumière Magenta et la lumière Blanche on obtient la lumière verte. Ce modèle consiste à décomposer une couleur en valeurs de Cyan, de Magenta et de Jaune.

L'absence de ces trois composantes donne du blanc tandis que leur ajout donne du noir. Toutefois, le noir obtenu par l'ajout des trois couleurs : Cyan, Magenta et Jaune n'étant que partiellement noir en pratique (et coûtant cher), les imprimeurs rajoutent une composante d'encre noire que l'on appelle noir pure. On parle alors de quadrichromie, ou modèle CMYK (Cyan, Magenta, Jaune, Noir pure, ou en français CMJN).

4.4. Le codage CIE : La Commission Internationale de l'Eclairage CIE a élaboré le système colorimétrique xyY représentant les couleurs selon leur chromaticité (axes x et y) et leur luminance (axes- Y). Le diagramme de chromaticité (ou diagramme chromatique), issu d'une transformation mathématique représente sur la périphérie les couleurs pures, c'est-à-dire les rayonnements monochromatiques correspondant aux couleurs du spectre (couleurs de l'arc en ciel), repérées par leur longueur d'onde. La ligne fermant le diagramme (donc fermant les deux extrémités du spectre visible) se nomme la droite des pourpres, car elle correspond à la couleur pourpre, composée des deux rayonnements monochromatiques bleu (420 nm) et rouge (680 nm) :

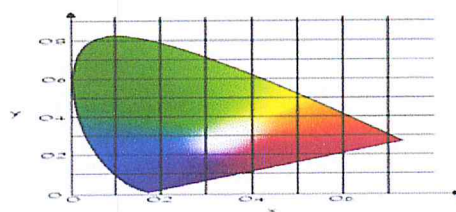


Figure 1.8: Représentation du codage CIE

Les modèles de la CIE ne sont pas intuitifs, toutefois le fait de les utiliser garantit qu'une couleur créée selon ces modèles sera vue de la même façon par tous.

Il s'agit d'un mode très utilisé dans l'industrie, mais peu retenu dans la plupart des logiciels étant donné qu'il est difficile à manipuler.

4.5. Le codage YUV : Le modèle YUV (appelé aussi CCIR 601) est un modèle de représentation de la couleur dédié à la vidéo analogique. Il s'agit du format utilisé dans les standards PAL (Phase Alternation Line) et SECAM (Séquentiel Couleur avec Mémoire). Le paramètre Y représente la luminance (c'est-à-dire l'information en noir et blanc), tandis que U et V permettent de représenter la chrominance, c'est-à-dire l'information sur la couleur. Ce modèle a été mis au point afin de permettre la transmission des informations colorées aux téléviseurs couleurs, tout en s'assurant que les téléviseurs noir et blanc existant continuent d'afficher une image en tons de gris.

Voici les relations liant Y, U, V à R, G et B :

- $Y = 0.299R + 0.587G + 0.114B$
- $U = -0.147R - 0.289G + 0.436B = 0.492(B - Y)$
- $V = 0.615R - 0.515G - 0.100B = 0.877(R - Y)$

Ainsi U est parfois noté Cr et V noté Cb, d'où la notation YCrCb.

4.6. Le codage YIQ : Le modèle YIQ est très proche du modèle YUV. Il est notamment utilisé dans le standard vidéo NTSC (utilisé entre autres aux États-Unis et au Japon).

Le paramètre Y représente la luminance. I et Q représentent respectivement l'interpolation et la quadrature. Les relations entre ces paramètres et le modèle RGB sont les suivantes :

- $Y = 0.299R + 0.587G + 0.114B$
- $I = 0.596R - 0.275G - 0.321B$
- $Q = 0.212R - 0.523G + 0.311B$

5. Les types d'images numériques : Les images numériques sont classés selon le format or la multitude des formats vient de la volonté de chacun des fabricants de logiciel d'imposer son format car on trouve très peu de différences hormis que

quelques-uns sont compressés et d'autres ne le sont pas aussi par rapport aux en-têtes du fichier. Parmi ces formats nous pouvons citer:

5.1. Les formats vectoriels [9]: Ils sont en fait une suite d'objets géométriques (rond, carré, droite, image clipart) défini par leurs coordonnées polaires. Les formats les plus connus sont : DXF (Autocad); SYLK; Lotus PIC et Lotus DDIF; cependant, de part le fait qu'ils soient vectoriels ces formats supportent les images de type photo. Ainsi hormis dans le secteur du dessin industriel avec le logiciel Autocad, les images vectorielles sont peu utilisées, au profit des métas fichiers.

5.2. Les formats bitmaps : Ils sont appelées aussi image raster, il s'agit d'images pixellisées, c'est-à-dire un ensemble de points (pixels) contenus dans un tableau, chacun de ces points possédant une ou plusieurs valeurs décrivant sa couleur. Il existe plusieurs types tel que : BMP, GIF, PNG et JPEG /JPEG2000.

5.3. Les formats méta fichiers [8]: Ils sont très répandus. Ils offrent la possibilité d'intégrer à la fois des images bitmap et des objets vectoriels. Les formats les plus connus sont : EPS Mackintosh; PICT RIF et WMF. Les métas fichiers ont la facilité de porter les images d'une plateforme à une autre. Les données étant stockées, la plupart du temps, en ASCII, les problèmes d'ordre de bit n'apparaissent pas. L'ASCII offre aussi une plus grande aptitude à la compression des fichiers.

6. Le traitement d'image :

6.1. Introduction au traitement d'image : Le traitement d'image est un ensemble de techniques appliquées aux images permettant de :

- Faciliter la manipulation des images (compression pour le stockage, visualisation).
- Eliminer ou réduire les bruits et les parasites introduits lors de l'acquisition.
- Donner une meilleure interprétation de l'image, de son contenu et ceci en détectant la présence de certaines formes.

Le traitement d'image est utilisé dans divers domaines tels que :

- La physique
- La biologie.
- La météorologie.
- L'industrie (la robotique, la sécurité).

- La médecine (radiologie, scanner, échographie...).

6.2. Type de traitement d'image [12]: On distingue deux types de traitement d'image :

6.2.1. Traitement de bas niveau : Il appréhende l'extraction des objets constituant une scène donnée, ces objectifs sont quantifiés « numériques » en vue de leur identification.

6.2.2. Traitement de haut niveau: Il permet l'interprétation de l'image et sa compréhension dans le but de concevoir un système intelligent qui permet d'extraire une information utile.

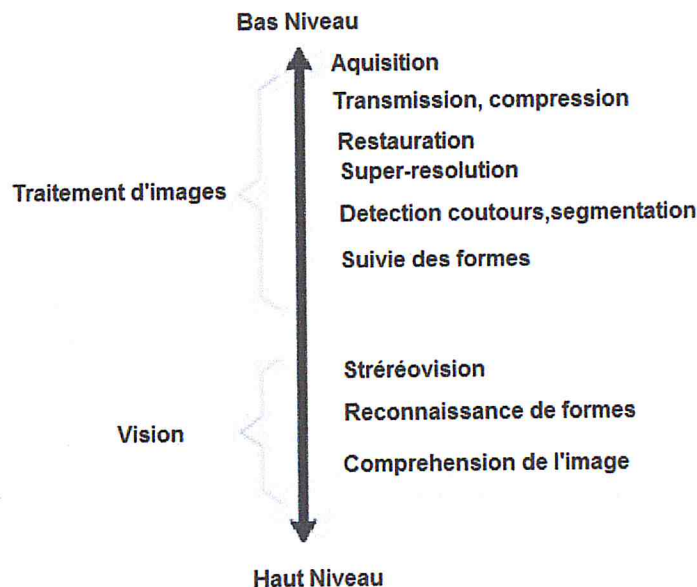


Figure 1.9 : Types de traitements d'image

6.3. Etapes du traitement d'images [12,13] : L'image suit un système de traitement d'image afin d'arriver à une information précise, ce système est composé comme suit :

6.3.1. Acquisition : Cette étape est cruciale dans le processus de traitement d'image car elle fait la transformation de l'image prise ou trouvée dans des documents « image d'origine » en une donnée binaire utilisée ou manipulée par la machine, c'est la conversion de l'image analogique en numérique. Celle-ci doit passer par une procédure de numérisation. Il existe plusieurs outils pour cette étape allant du scanner et photo /camera numérique simples à main de satellite de télédétection.

6.3.2. Prétraitement : Le traitement, souvent appelé prétraitement, regroupe toutes les techniques visant à améliorer la qualité d'une image, et cela dans le but de faciliter l'extraction des différents objets de la scène. De ce fait, la donnée de départ est l'image initiale et le résultat est également une image. L'idéal est d'obtenir un résultat sans bruit. C'est la raison pour laquelle les techniques ne sont pas les mêmes.

La notion de qualité est une notion très subjective, assujettie à la réalisation d'un objectif. La qualité d'une image n'est pas forcément la même pour un ordinateur ou pour un opérateur humain. Les capacités de vision de l'être humain étant limitées, il est indispensable d'adapter la dynamique de l'image à notre vision. Dans cette étape il existe plusieurs sortes de traitement :

6.3.2.1. Traitement numérique des images : On peut citer les traitements suivants :

- **Restauration :** La restauration a pour but d'inverser l'effet du phénomène dégradant. Il s'agit donc de produire une image la plus proche de la réalité physique de la scène observée. Le plus souvent, cette étape est la première dans la chaîne de traitements constituant un système de vision.
- **Amélioration :** L'amélioration a pour but de satisfaire l'œil de l'observateur humain. C'est pourquoi l'image produite peut être différente de la réalité. Cette amélioration peut servir dans un premier temps à faciliter la visualisation de l'image sur un écran d'ordinateur.
- **Compression :** On classe les techniques de compression par extension du fichier informatique. Il s'agit là de faciliter le traitement et surtout le stockage des images par une réduction adéquate de leur volume d'information.
- **Segmentation :** Il existe deux grandes catégories de segmentations : la segmentation de région et la segmentation de contour. Les pixels présentant une même caractéristique sont décrits par un niveau de gris compris dans un certain intervalle.

6.3.2.2. Filtrage: Le principe du filtrage est de modifier la valeur des pixels d'une image, généralement dans le but d'améliorer son aspect. En pratique, il s'agit de créer une nouvelle image en se servant des valeurs des pixels de l'image d'origine et plus précisément il consiste à modifier la distribution fréquentielle des composantes d'un

signal selon des spécifications données. Il existe plusieurs types de système pour ce type de traitements on peut citer par exemple :

- **Filtrage de lissage (passe-bas):** Les filtres de lissage sont des opérateurs qui éliminent des éléments perturbateurs / non significatifs dans les images numériques, soit pour améliorer leur visualisation, soit pour les simplifier en but d'un traitement postérieur. Ce type de filtre n'affecte pas les composantes de basse fréquence dans les données d'une image, mais il s'intéresse aux composantes de haute fréquence c'est-à-dire qu'il supprime les composante de haute fréquence. Il peut être répété plusieurs fois.

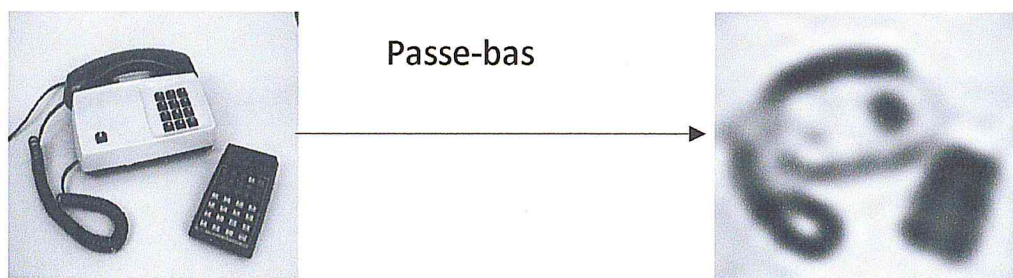


Figure 1.10: Filtrage Passe-bas

- **Filtres passe haut :** Le filtrage passe-haut correspond à la multiplication dans le domaine fréquentiel d'une fonction porte-complémentaire. Il a les caractéristiques inverses du filtre passe bas car il s'intéresse aux composantes de basse fréquence.



Figure 1.11: Filtrage Passe haut

- **Filtres passe-bande :** Le filtrage passe-bande correspond à la multiplication dans le domaine fréquentiel par une fonction bande symétrique. Elle consiste à éliminer la redondance d'informations entre l'image originale et l'image obtenue par le filtre passe bas.

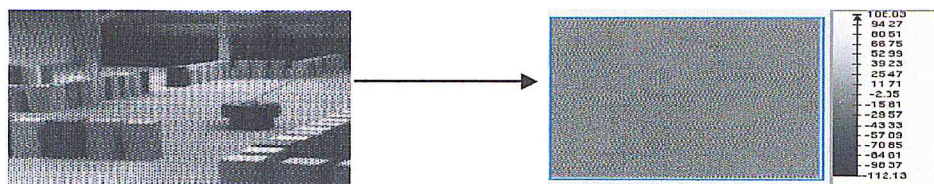


Figure 1.12: Filtrage Passe-bande

6.3.3. Analyse de l'image: En analyse d'image on distingue 2 types de traitements : le traitement de bas niveau et de haut niveau. L'ordinateur traite l'image comme un ensemble de points ayant chacun un niveau de gris voir un niveau de couleur.

7. Quelques traitements d'image [37,41] : Dans les modèles différentiels, l'image est une fonction continue F tel que : $F:I \times I \rightarrow [0, 255]$

Nous pouvons étudier le comportement local à l'aide de ses dérivées. Une telle étude n'a de sens que si la fonction f est assez régulière. Ce n'est pas toujours le cas ! Une image noir et blanc sera discontinue (en fait continue par morceaux) les zones de discontinuité sont les contours.

7.1. Le gradient : Le gradient d'une image est un changement d'intensité ou de couleurs dans une image, il peut être utilisé pour extraire des informations à partir de l'image (généralement par une convolution avec un filtre) dans les directions x, y .

Une des utilisations les plus courantes c'est la détection des contours .

Le gradient est calculé par l'équation suivante :

$$\nabla f = \frac{\partial f}{\partial x} \hat{x} + \frac{\partial f}{\partial y} \hat{y} \quad \text{Equation 1.1}$$

Avec

$\frac{\partial f}{\partial x}$: Le gradient dans la direction x de la fonction de l'image.

$\frac{\partial f}{\partial y}$: Le gradient dans la direction y de la fonction de l'image.

7.2. La convolution : La convolution d'une image est une technique de traitement d'image qui permet de changer l'intensité des pixels en reflétant l'intensité des pixels voisins, en utilisant la convolution nous pouvons obtenir des effets d'images

populaires tel que le flou, la netteté ou la détection de contours comme c'est mentionné avant.

On peut utiliser plusieurs filtres comme le filtre Prewitt, Roberts, Kirsh.... Mais les plus utilisés pour la détection de contour c'est les filtres Laplacien, Kernel, Sobel que voici leurs interprétations :

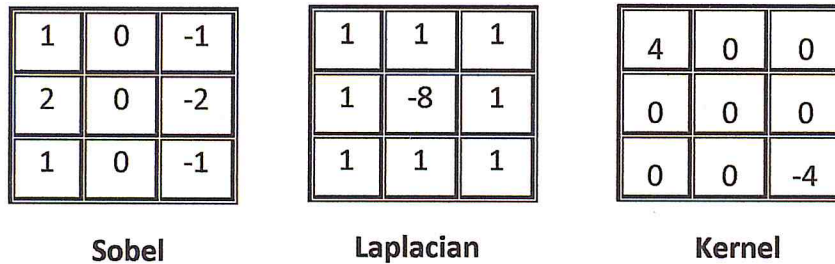


Figure 1.13: Exemples de filtres utilisés pour la convolution

La convolution peut se calculer par un produit de la matrice pixels qui représente les valeurs des pixels et l'un filtre (figure 1.14).

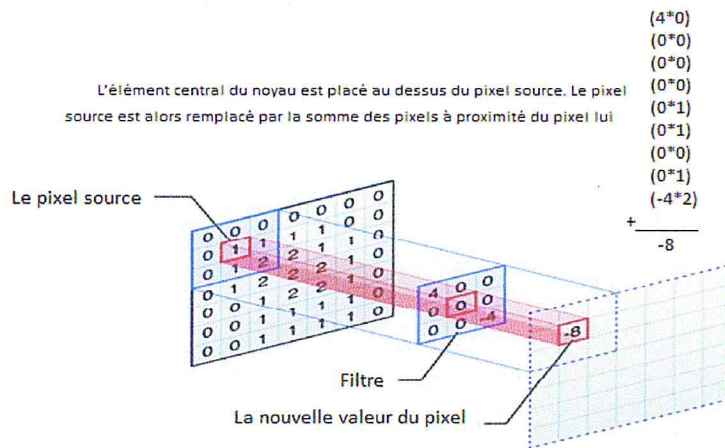


Figure 1.14 : Le principe de la convolution

8. Conclusion :

Nous venons de présenter quelques notions et concepts de base en imagerie et son traitement, ainsi que les différentes étapes de cette discipline.

Désormais nous pouvons aborder le chapitre suivant qui concerne les manuscrits en général et les manuscrits algériens précisément.

CHAPITRE 2 :

GÉNÉRALITÉS SUR LES

MANUSCRITS

Sommaire :

Dans ce chapitre, nous allons faire une petite étude sur les documents anciens et plus précisément les manuscrits. Notre chapitre se divise en :

- 1- Introduction
- 2- Définition du manuscrit
- 3- Définition du codex
- 4- Les éléments descriptifs des manuscrits
- 5- Etat des manuscrits en Algérie
- 6- Conclusion

1. Introduction :

Les manuscrits anciens représentent une richesse infinie pour les pays car ils sont les porteurs de l'information à travers les générations.

2. Définition du manuscrit [23]: Un manuscrit en latin « manu scriptus » est tout texte écrit à la main sur un support souple, que ce soit par son auteur ou non, avant ou après l'invention de l'imprimerie. L'expression s'applique à différents supports (papyrus, rouleaux, livres de parchemin ou de papier) mais on ne l'emploie pas pour d'autres (les tablettes de cire, par exemple).

Aujourd'hui, on qualifie parfois ces « manuscrits » de « tapuscrits » quand il s'agit de documents dactylographiés produits à l'aide d'un logiciel de traitement de texte.

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, le style de l'écriture, la couleur de l'encre, la présence d'annotations, etc. Il n'existe donc jamais deux manuscrits identiques, même s'il peut arriver que le texte de deux manuscrits soit identique.

L'étude des textes manuscrits est l'objet de la paléographie et de la philologie, tandis que l'étude des livres manuscrits en tant qu'objets relève de la codicologie.

Les manuscrits ne respectent aucune forme ou structure bien définie. Ils sont, généralement de dimensions inégales et formés d'un nombre de pages très varié allant parfois jusqu'à 500 pages.

3. Définition du codex [24]: Un codex (bloc de bois ou livre en latin) est un livre manuscrit du même format que celui utilisé pour les livres modernes, avec des pages reliées ensemble et une couverture. Par la possibilité qu'elle offre d'accéder directement à n'importe quelle partie du texte, cette invention romaine a remplacé le rouleau de parchemin et est la première forme de livre de toutes les cultures d'Eurasie.

Le codex est un livre de forme parallélépipédique, résultat de l'assemblage de manuscrits, d'abord en parchemin à partir des I^{er} et II^e siècles dans l'empire romain puis en papier depuis le XIII^e siècle, historiquement issu de la forme ancienne du rouleau et du volumen horizontal. Sa caractéristique principale est la reliure, par leur marge, des feuillets qui le constituent. Cette présentation des textes a constitué une véritable révolution au début de l'ère chrétienne car à l'inverse du rouleau (volumen), qui impose une lecture continue, le codex permet d'accéder aux chapitres (structure du texte) de manière directe. L'habitude de numéroter les pages (par des lettres) accompagna cette innovation, la principale évolution introduite par le codex étant la notion de « page ». Son adoption dans la chrétienté est d'autant plus marquée que, support de la Bible, le codex permet de se différencier des rouleaux sur lesquels les juifs écrivent la Torah (Sefer Torah). Par extension, le terme codex a été employé pour des recueils de lois (comme le Codex Theodosianus) d'où le nom de Code employé aujourd'hui. L'imprimerie n'a pas fait disparaître les codex : ceux de Léonard de Vinci en constituent une preuve frappante. Le codex comme objet est étudié par une science spécifique : la codicologie.

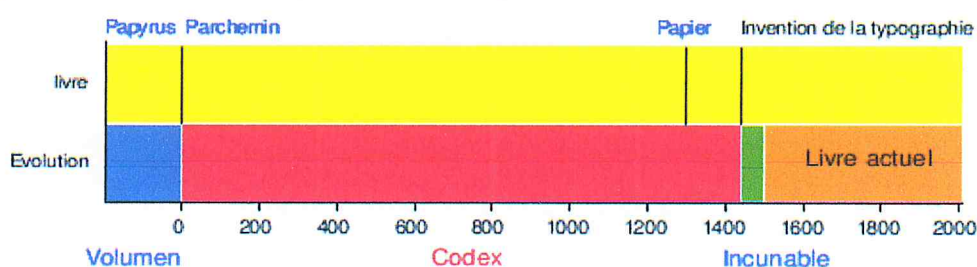


Figure 2.1 : Evolution du support de manuscrits

4. Les éléments descriptifs des manuscrits [25]: Le caractère d'unicité des manuscrits leur donne la particularité d'être différents les uns des autres : Il s'agit d'un ensemble d'éléments définissant d'une manière unique les manuscrits. Parmi les éléments d'identification nous retrouvons principalement :

4.1. La mention de responsabilité : Il s'agit d'identifier le responsable de l'écriture du manuscrit. En effet, bien souvent des manuscrits sont écrits par d'autres personnes autres que l'auteur : ce sont des copistes.

4.2. Le nom du propriétaire : Il peut être le nom d'une personne, d'une bibliothèque ou d'une institution. Cette information figure soit sur des médaillons, ou au premier recto du manuscrit. Elle est très utile pour le chercheur afin d'identifier le développement historique du manuscrit.

4.3. Le titre du manuscrit : C'est l'élément principal pour l'identification du manuscrit. Il peut être retrouvé dans plusieurs endroits du manuscrit. Le plus souvent, nous le retrouvons au recto de la première page accompagné du nom de l'auteur.

4.4. Le colophon : Il s'agit de la formule finale dans laquelle le copiste mentionne son nom, la date selon les calendriers hégire et grégorien et le lieu de la copie. Le colophon prend souvent la forme d'un triangle. Parfois, il ne se distingue pas du corps du texte. Il contient le nom du copiste, le nom du possesseur, la date d'achèvement et le nom de la bibliothèque pour laquelle le manuscrit a été copié.



Figure 2.2: Exemple de colophon en forme triangulaire

4.5. Le décor : Le décor d'un manuscrit joue plusieurs rôles : didactique, symbolique, historique et parfois, purement esthétique. Par conséquent, il permet à la fois d'articuler le texte et au lecteur de se repérer. Le décor peut se présenter sous une

forme peinte ou dessinée. Il apporte des indices de datation et de localisation du manuscrit. L'iconographie peut fournir la clé d'identification d'un possesseur ou d'un destinataire (portraits, emblèmes, armoiries, allusions biographiques). L'étude du décor est soumise à une hiérarchie, la peinture prime sur le dessin, les métaux sur les couleurs, l'image sur l'ornement, le grand sur le petit. La décoration du texte est faite en utilisant des enlumineurs, considérées comme un moyen d'expression de l'art artistique de l'enlumineur. Ainsi, nous retrouvons le décor en plein texte, le décor de section, les signes de division de texte, les encadrements, les décors de reliures,... etc.

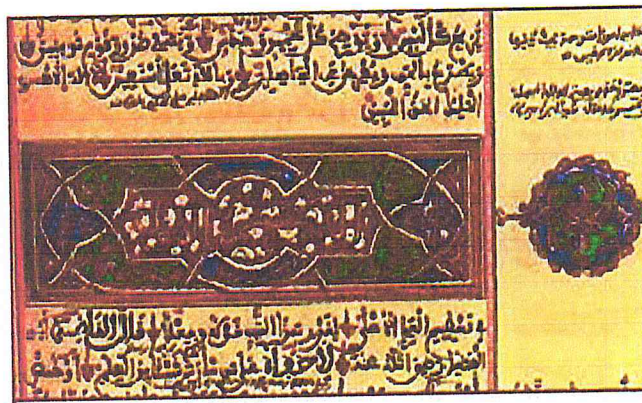


Figure 2.3 : Illustration de décors de manuscrits

4.6. La reliure : C'est l'assemblage des cahiers formant un codex. De ce fait, elle doit être décrite au même titre que les autres éléments du codex. La reliure intéresse d'une part l'histoire du livre et des techniques artisanales, d'autre part l'histoire intellectuelle. La description de la reliure suit les étapes de fabrication, c'est-à-dire qu'elle se fait selon la séquence "bloc, aspect du dos, les plats, les pièces de renfort, la couverture, le décor, les titres et les étiquettes et enfin la date de la reliure".

4.7. Organisation du volume : Un manuscrit peut être composé de plusieurs cahiers de parchemin ou de papier, on parle de volume. Il est donc important d'observer l'organisation du volume. Cette opération permet de déterminer si le manuscrit est homogène ou composite, complet ou mutilé, si les cahiers sont reliés en ordre ou dans le désordre. Pour ce faire, il faut considérer les signes de numérotation et de renvoi. Ils permettent de se repérer dans le manuscrit. Parmi les signes de numérotation nous distinguons la foliotation, la signature :

- **La foliotation** : C'est la numérotation de chaque feuillet formé par l'ensemble recto verso.
- **La signature** : C'est la numérotation des cahiers. Elle se trouve soit sur le premier feuillet du cahier, au recto, soit sur le dernier au verso ou soit, sur les deux.

4.8. Type de support : Plusieurs types de supports ont été utilisés pour rédiger les manuscrits arabes : le papyrus, le parchemin et le papier. Par ailleurs, il semble que la production des corans sur papier n'a commencé qu'au 11^{ème} siècle, jusqu'à lors le parchemin était considéré comme prestigieux.

L'étude du support doit être accompagnée par celle des ponctuations, qui sont utilisées pour divers objectifs (séparation entre les versets dans les textes coraniques, marquage de pause, décor) et de la description de la page (dimension, mise en page, réglure, structuration des lignes dans la page).

5. Etat des manuscrits en Algérie : L'Algérie possède un riche trésor de manuscrits et il est difficile d'accès. Il est estimé à 12 000 volumes. Elle possède 2 centres ou établissements chargés de la numérisation des manuscrits :

5.1. La bibliothèque nationale d'Algérie BNA : Elle est la première institution culturelle en Algérie. Elle est incontestablement une référence dans tous les domaines de la connaissance culturelle et artistique, du savoir et de l'information. Elle est la préservatrice de la mémoire collective et dispose de moyens adéquats pour mener à bien ses missions. Elle joue un rôle primordial dans la satisfaction des désirs de savoir et de culture de la communauté. Elle contient plusieurs services nous pouvons citer :

- **Service de l'audiovisuel** : Le service audiovisuel de la Bibliothèque Nationale d'Algérie occupe une place à part dans le paysage de la lecture publique en Algérie. Il a la taille d'une bibliothèque médiatique parce qu'il offre comme service en matière d'audiovisuel.
- **Service des manuscrits et ouvrages rares** : Ce service contient 3.743 manuscrits en langue arabe et étrangère. Ils sont conservés et mis à la disposition des lecteurs pour consultation sur place. Parmi ses activités on trouve :

❖ La présentation et la conservation des manuscrits et des ouvrages rares et précieux.

❖ Le traitement, le classement des manuscrits et la reproduction des ouvrages de références microfilms, CD-ROM et sur papier.

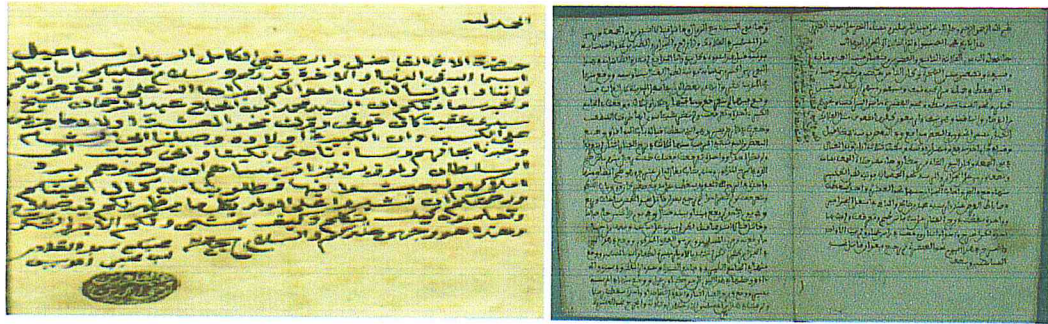
❖ L'organisation de plusieurs expositions pour le compte des chercheurs qui fréquentent le service en particulier, des historiens ou des théologiens.

❖ La reproduction des documents.

❖ Les demandes de reproduction de manuscrits sur papier, sur microfilm ou CD-ROM doivent obtenir, au préalable, un accord de la part du responsable du service des manuscrits.

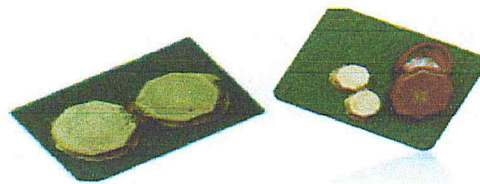
Le service s'attache à acquérir ce qui peut compléter ses collections ou en faciliter la compréhension. Il s'étoffe actuellement grâce aux différentes sources comme les dons ou legs émanant de personnes physiques ou morales généreuses et grâce aussi aux achats pour lesquels un budget est prévu. Le fond ainsi recueilli exige souvent un important travail de classement et de recherche avant de pouvoir relier, reproduire, coter et cataloguer tous les documents.

5.2. Le centre national des manuscrits d'Adrar CNMA [26]: Il est l'outil qui permet de sauvegarder et de restaurer le patrimoine culturel du Touat ainsi que d'autres régions d'Algérie. Quelque 30.000 manuscrits avaient été recensés l'année 1908, période de la colonisation. Ce nombre s'est réduit à 17.000 en 1962, puis à 12.000 actuellement; parmi lesquels 25% dans un état moyen, 30% dans un mauvais état et 45% dans un état critique. Il est à noter que 44% de ces manuscrits traitent des sciences islamiques, 34% de littérature, 12% du droit, 3% de mathématiques et logique, 1% de médecine et le reste dans le divers. Les propriétaires de ce trésor de manuscrits disposent de 28 «khizanat» (bibliothèques) dans les régions d'Akbli et Tit, à 300 km au sud d'Adrar, à Ouled Saïd, à 270 km au nord d'Adrar. Parmi les plus importantes bibliothèques, on trouve celles d'Akbli, Metarfa et Zaouïa El-Bakria, qui dispose de manuscrits datant du huitième siècle.



Lettre de l'Emir Abdelkader contenant son tampon

Manuscrit décrivant la rentrée des espagnoles en Algérie



Coran de forme microscopique de forme octogonale doré sur tranche. Exécuté dans un très joli neskhi persan entièrement vocalisé écrit en 1016 de l'hégire.

Figure 2.4 : Exemple de manuscrits algériens

6. Conclusion :

Nous avons vu dans ce chapitre un petit résumé sur les manuscrits en général et les manuscrits algériens plus précisément. Nous n'oublions pas de mentionner que beaucoup de personnes possèdent un nombre important de manuscrits hérité à travers les temps et leur exploitation est très difficile.

CHAPITRE 3 :

LA NUMÉRISATION

Sommaire:

La masse très importante des documents anciens stockés dans des différentes places, rend leurs utilisations très difficiles, d'où le besoin de les numériser. Nous introduisons dans ce chapitre le principe de la numérisation des documents, ses principaux avantages et limites en générale et la numérisation du patrimoine algérien plus précisément. Notre chapitre se divise en :

- 1- Introduction
- 2- La numérisation des documents du patrimoine
- 3- La numérisation du patrimoine en Algérie
- 4- Conclusion

1. Introduction :

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 qui constitue le cadre de notre travail.

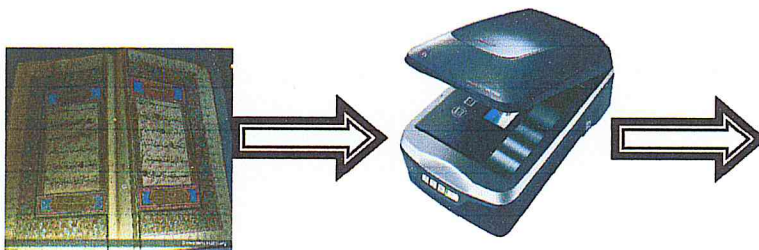
2. La numérisation des documents du patrimoine : Le document joue un rôle important dans le développement des civilisations et de leurs cultures, il est le teneur et le transmetteur de l'information à travers les générations. Au fil de ces générations le document ou plus précisément le support physique du document a beaucoup évolué allant des supports rigides (la pierre, le marbre...) aux supports souples (le parchemin, les manuscrits, le papyrus, le papier...). Le papier est considéré comme le principal support mais il est fragile donc la conservation est plus difficile.

Par ailleurs, l'apparition des technologies numériques de l'information et de la communication a révolutionné la transmission des connaissances, de ce fait le support papier n'est plus le support unique pour diffuser de l'information. Le document électronique est aujourd'hui le vecteur de diffusion le plus utilisé dans le monde.

Les technologies de la numérisation sont multiples, elles aident dans la diffusion de l'information car le document électronique est plus facile à manipuler (modification, enregistrement, transmission et impression) n'importe où et n'importe quand sans aucune perte de l'information ou de la qualité. La numérisation du patrimoine offre de nouveaux services comme la consultation à distance, l'accès à des documents rares, de nouvelles formes de lecture et de navigation, une recherche plus performante de l'information et enfin la possibilité de partager l'information. Pour cela nous avons besoin d'élargir le champ d'utilisation de ces techniques de numérisation afin de convertir le plus grand nombre possible des documents sous forme numérique.

2.1. La numérisation [14]: La numérisation est la technologie qui permet de convertir des documents physiques en images électroniques. La numérisation au sens large du terme est l'ensemble des opérations nécessaires pour capturer les images et les transformer en données exploitables informatiquement. La numérisation fournit les images et les métadonnées associées ainsi que les informations nécessaires au fonctionnement des services offerts aux usagers.

Cette définition de la numérisation est encore plus large que celle de la rétro-conversion qui consiste à convertir les documents papier dans un format logique pivot les rendant de nouveau réédictables et donc réutilisables. La rétro-conversion concerne la transformation complète d'un document papier en un document électronique proche de l'original alors que la numérisation, au sens large, s'intéresse aussi à l'accessibilité et aux services que l'on peut offrir aux usagers. Au sens strict du terme, la numérisation consiste seulement à effectuer la capture des images des documents, on parlera alors de capture d'images.



L'image est découpée en lignes, chaque ligne en petits carrés (pixels), chaque pixel est transformé en une valeur numérique selon sa couleur, puis codée sur un ou plusieurs octets.

Figure 3.1 : Processus de numérisation

2.2. Type de scanners utilisés pour la numérisation [21]:

2.2.1. Scanner à plat : Le support sur lequel on pose les originaux est une vitre en verre incassable. Plus la vitre est grande, plus les originaux qu'on numérise peuvent être grands, évidemment. Une grande vitre permet aussi d'accueillir plusieurs petits documents à numériser. Le document est posé contre la vitre. Avec certains scanners, la vitre est mobile et entre dans le scanner. Dans ce cas on peut attacher le document en se servant de rubans adhésifs.



Figure 3.2 : Scanner à plat

2.2.2. Scanner à défilement : Le document est entraîné, comme pour un fax, c'est lui qui se déplace et non les capteurs ou un miroir. Ce procédé est principalement utilisé dans le monde bureautique. Ils permettent aujourd'hui de numériser des documents de grammages variables de format allant de la carte de visite à l'A3. Certains d'entre eux, peuvent scanner jusqu'à plus de 160 pages par minute en recto/verso (ex: Kodak i840). D'autres types de scanners à défilement, plus spécifiques, sont prévus pour numériser des documents grands formats (scanners de plan supérieur à l'A0). On note également l'existence d'autres scanners à défilement spécifiques, pour par exemple des photos, des chèques, des cartes de visite, etc.



Figure 3.3 : Scanner à défilement

2.2.3. Scanner à main : L'opérateur fait défiler l'outil contre le document à scanner. Au début des années 1990, les scanners à main ont été les premiers scanners « grand public », du fait notamment de leur faible prix.

Ils sont aujourd'hui beaucoup moins fréquemment utilisés, eu égard à l'habileté que réclame leur manipulation et aux erreurs dont ils proviennent lors de leurs utilisations. Ils restent intéressants dans des domaines tels que l'analyse de l'écriture, notamment pour le chinois et le japonais.

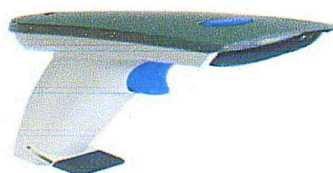


Figure 3.4 : Scanner à main

2.2.4. Scanner à diapositive : Dans un scanner à diapositive, la lumière est directe, elle traverse le document transparent, ce qui permet d'obtenir une qualité chromatique et une résolution très importante.

Certains scanners à plat sont aussi équipés d'un boîtier lumineux, dit « dos diapo » qui leur permet donc de scanner en lumière réfléchie les documents opaques et en lumière directe les documents transparents. Bien que ces scanners soient rarement d'excellents scanners diapo, ils peuvent constituer un compromis intéressant.



Figure 3.5 : Scanner à diapositive

2.2.5. Scanner à tambour : Le support sur lequel on pose les originaux est un cylindre en acrylique. L'acrylique, est une matière qu'on appelle communément du « plexiglas ». Plus le cylindre est grand, plus les originaux qu'on numérise peuvent être grands, évidemment. Un grand cylindre permet aussi d'accueillir plusieurs petits documents à numériser. Le document est enroulé autour de ce cylindre. Il y est attaché solidement en se servant de rubans adhésifs car quand on numérise, ce cylindre tourne très vite.

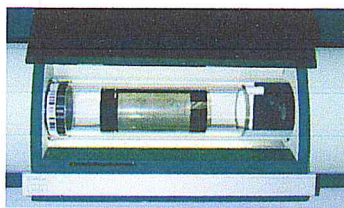


Figure 3.6 : Scanner à tambour

2.3. L'importance de la numérisation [14, 22,23]: Les documents numériques issus d'une numérisation servent à :

- L'impression
- La documentation
- La recherche
- La publication en ligne
- Et le plus souvent pour la gestion de collections, la préparation de catalogues et la promotion d'exposition ...

2.3.1. Pourquoi la numérisation est devenue si populaire ?

- Elle permet de rendre les collections beaucoup plus accessibles (collections autrefois trop lointaines, objets trop fragiles pour être manipulés ou exposés).
- Elle permet de mettre en valeur des aspects particuliers de l'histoire locale ou d'atteindre un public national ou international.
- Elle permet de rassembler, à des fins de comparaison et de recherche, des objets ou des collections dispersés en plusieurs lieux
- Elle peut devenir un puissant outil d'enseignement.
- La numérisation peut également contribuer à la gestion des collections en procurant à tout le personnel une meilleure information sur le contenu des collections.
- Le simple fait de sélectionner des objets à numériser fournit au personnel une occasion supplémentaire d'évaluer et de consigner l'état des objets. La numérisation incite en outre à améliorer la documentation, en particulier lorsque des enregistrements d'accompagnement sont destinés à être rendus public en même temps que les images numérisées : le besoin de normes et d'une meilleure information devient rapidement évident.

- Les objets numérisés sont utilisés dans une vaste gamme de produits de diffusion comme les sites Internet, le matériel de promotion, de nouveaux articles pour la boutique souvenir d'un musée, etc.
- Comme la technologie numérique permet de faire une recherche dans un grand nombre d'enregistrements, de modifier et de traiter des images et des textes et de rassembler des éléments disparus selon des modalités nouvelles, elle peut se révéler un outil souple et précieux dans tous les secteurs d'un musée.
- La facilité d'exécution des tâches mentionnées ci-dessus favorise également une collaboration plus étroite avec d'autres établissements.

2.3.2. Pourquoi numériser ?

- **Facilité de reproduction** : Les données numériques peuvent être dupliquées c'est-à-dire sauvegardées rapidement sur plusieurs supports différents (bases de données, cd, disque dur...). De ce fait elles peuvent être exploitées par la machine à tout moment : échangées, stockées, manipulées... à des coûts réduits.
- **Facilité de consultation** : Aujourd'hui la consultation des documents numériques est plus utilisable que celle des supports physiques (livres, microfilms, manuscrits). Le public peut consulter et accéder à des documents anciens ou rares.
- **Facilité d'échange d'information** : L'échange des supports physiques (livre...) peut être lent et très coûteux à être exécuté, ils peuvent être envoyés par courrier ou moyen de transport. Par contre les documents numériques peuvent être envoyés à tout moment par des courriers électroniques, et cela peut prendre quelques minutes.
- **Facilité de recherche** : La recherche d'information sur des supports numériques est plus performante et efficace en utilisant des moteurs de recherche permettant de comparer rapidement les informations recherchées sur plusieurs bibliothèques numériques et d'en extraire les informations pertinentes à partir du contenu du support numérique.
- **Indépendance du support physique** : Les données numériques peuvent être stockées sur des supports numériques (bases de données, cd, disque dure...). Ces données peuvent être éditées sur plusieurs supports (écran, papier...) et manipulées simultanément à tout moments et à tout endroit, par n'importe quel moyen informatique (PC, téléphone portable...). De nombreux chercheurs peuvent travailler

simultanément à distance sur les mêmes documents ou bien consulter des documents rares autrefois et peu accessibles.

- **Préservation des documents originaux** : Les supports physiques sont conservés dans des bibliothèques ou des endroits qui peuvent être détruits, volés ou altérés (l'acidité par exemple) à tout moment. Pour cela il est nécessaire de garder des copies sur des supports différents afin de préserver notre patrimoine.

- **Plus de détérioration des supports en consultation** : Une fois numérisés, les documents originaux peuvent être conservés dans des conditions optimales, sans le danger que représente leur manipulation. On augmente donc leur sécurité.

- **Accessibilité** :

- ❖ « **A distance** » Elle tient à la capacité d'envoyer un signal numérique sur les réseaux de télécommunications sans perte de qualité.

- ❖ « **temporelle** » La collection numérisée est accessible 24 heures sur 24 avec un minimum de personnel.

Finalement, l'accessibilité est grandement améliorée pour le public.

- **Qualité** : Elle tient à la possibilité de numériser un document avec une très forte résolution, selon les besoins. Il est également très facile de réaliser des copies de qualité inférieure à partir de la copie de haute qualité chaque fois que nécessaire.

- **Espace nécessaire** : La haute densité du stockage de l'information sur les supports numériques peut se traduire par une réduction majeure de l'espace de rayonnage nécessaire. D'où, également, une réduction de l'espace nécessitant une régulation climatique conforme aux normes archivistiques et, en conséquence, un abaissement des frais de fonctionnement. Par ailleurs, un signal numérique peut être fortement compressé avec ou sans perte d'information. Une compression avec perte d'information peut s'effectuer sans perte d'information « sensible », c'est-à-dire détectable. Une information numérique est facilement stockable sur des supports dont la capacité augmente et dont le coût et la taille diminuent rapidement au fil des années.

- **Amélioration de la lisibilité des documents** : Le traitement d'images peut améliorer la lisibilité de documents défraîchis, tachés voire carrément illisibles.

- La transmission numérique est plus résistante que l'analogique aux défauts de transmission puisqu'il suffit de reconnaître, à la réception du message, sa présence et

non sa forme pour le reconstituer ; la puissance nécessaire pour la transmission est donc plus faible et les équipements de réception souvent plus petits.

2.3.3. Inconvénients, limites et risques de la numérisation :

- **Le coût** : C'est un des reproches souvent fait aux projets de numérisation. Les points soulevés ci-après sont parfois judicieux, parfois exagérés. Ils méritent en tout cas qu'on sache y répondre le cas échéant.

- ❖ **Investissement initial** : Le matériel nécessaire pour effectuer la numérisation peut être coûteux et demande souvent à être utilisé par des opérateurs qualifiés pour donner les meilleurs résultats.

- ❖ **Contraintes de rangement** : Les supports utilisés pour stocker l'information numérisée devront être conservés dans un environnement très propre et très stable sur le plan climatique, ce qui accroîtra la consommation d'énergie imputable à la collection. L'aménagement de cet espace de rangement à l'atmosphère stabilisée exigerait en outre un investissement initial.

- ❖ **Coûts de préparation** : Un document doit être préparé (parfois de façon destructrice, comme le massicotage des ouvrages avant leur passage par le scanner) pour la saisie numérique. Outre l'éventuelle préparation physique requise, le contenu initial doit en être classé et indexé et les références textuelles introduites dans la base de données par du personnel spécialisé, opérations qui peuvent être coûteuses. On pense que la copie de l'information des supports existants sur de nouveaux supports implique d'importantes contraintes de main-d'œuvre.

- **Le risque important de mauvais choix technologique** :

- ❖ **Lors de la numérisation** : Des erreurs notamment lors du choix de l'échantillonnage ou du format de fichier (formats propriétaires) peuvent empêcher une exploitation future des données.

- ❖ **Lors de la gestion des documents numériques** : Des erreurs lors du stockage (supports, conditions de stockage) et lors de la préservation des documents peuvent conduire à des pertes sévères de données (cas des données sur les vols lunaires de la NASA). Ne jamais oublier que les débuts de l'ère informatique ont été marqués par la plus grande perte de données de l'histoire de l'humanité.

- **La perte de repère physique :** Une partie de l'information d'un document est contenue dans le support physique. La numérisation ne retient donc pas tout : l'odeur, le poids, la texture ne sont pas reproduits comme le démontre ce compte-rendu d'expérience concernant le « petit Cartulaire » : « Le traitement de l'image connaît aussi ses limites, qui sont celles de la source. Comme plusieurs microfilms ont beaucoup vieilli, il serait souhaitable que la numérisation à haute résolution et en couleur soit faite directement sur les manuscrits à l'aide d'un appareil-photo numérique. De plus, le déchiffrement des écritures cursives, bien que facilité par la possibilité d'agrandir la taille de l'image à souhait, n'est guère simplifié. Les changements dans les teintes d'encre demeurent également imperceptibles. Le repérage de mots dans la source est aussi impossible si ceux-ci n'ont pas fait l'objet préalable d'une transcription. Aucune machine à ce jour ne peut donc remplacer l'œil du spécialiste. »

2.3.4. Quelques grands programmes de numérisation :

- **Projet NUMDAM de la cellule MATHDOC à Grenoble:** (Numérisation de Documents Anciens Mathématiques) Dans le but de soutenir les revues de mathématiques, le programme NUMDAM met en place un libre accès aux données bibliographiques et au texte des articles qui y sont apparus. Pour chaque revue concernée, la totalité des volumes publiés jusqu'en l'an 2000 a été convertie au format numérique, ce qui représente actuellement plus de 560 000 pages numérisées et 27 000 articles mis en ligne. Les articles eux-mêmes sont disponibles pour consultation en ligne à l'issue d'un délai pendant lequel ils sont réservés aux seuls abonnés. Il est possible de rechercher directement un article par nom d'auteur, mots du titre ou mots clés présents dans le texte. Il est également possible de feuilleter les sommaires de l'ensemble des volumes.

- **Internet d'archive :** Est une organisation à but non commercial fondée dans le but d'être la « bibliothèque d'internet ». Son intérêt est de s'adresser aux chercheurs, aux historiens, aux étudiants, aux personnes présentant un handicap et au grand public pour leur offrir des collections historiques qui existent en format numérique. Fondée en 1996, puis refinancée en 1999, cette organisation a connu une grande croissance et inclut maintenant des collections conséquentes. Elle propose des textes, de l'audio, des films, des logiciels ainsi que des pages web archivées. Elle travaille à

fournir des services plus spécialisés liés à l'enseignement, à la formation et à l'accès aux collections par des personnes présentant un handicap.

- **Le programme de l'UNESCO (organisation des nations unies pour l'éducation, la science et la culture)** Baptisé « mémoire du monde » a pour objectif de promouvoir la conservation des collections d'archives et de bibliothèques partout dans le monde et d'en assurer la plus large diffusion.

3. La numérisation du patrimoine en Algérie : Actuellement l'Algérie dispose de nombreux documents anciens issus du moyen âge jusqu'à nos jours. Ces documents peuvent être des manuscrits ou imprimés de différents alphabets (arabe, latin,..). Ils peuvent avoir des formes différentes de mise en page. Les différentes bases d'images nous ont permis de voir les problèmes liés à la qualité de ces images c'est pour cela que nous avons besoin d'une technologie aidant à la restauration des ces images.



Figure 3.7 : Exemple des anciens documents dégradés

L'Algérie dispose d'établissements chargés de la numérisation des documents anciens afin de sauvegarder les documents originaux, des manuscrits, des films... pour éviter leurs détériorations avec le temps. Parmi ceux-là nous citons :

3.1. Bibliothèque nationale d'Algérie (BNA) [17] : Elle a été créée le 13 Octobre 1835 par la décision de l'occupation algérienne ministre Adrian Verbruggen. En 1986 a commencé le projet de construction d'une bibliothèque nationale d'Algérie à nouveau en toute indépendance et elle a été inaugurée pour la 1ere fois le 1 novembre 1994. La BNA contient environ 600.000 ouvrages (livres, manuscrits, ouvrages rares...): 4291 manuscrits et 3901 titres de livres rares dans différentes langues (arabe -turque -perse...).Le plus ancien manuscrit date du 3^{ème} ou 4^{ème}

siècle (ère de l'hégire), c'est un fragment du Saint Coran écrit sur du parchemin. La collection est de nature encyclopédique et couvre plusieurs domaines comme le Saint Coran – hadith - théologie - droit - politique- administration - philosophie – logique – arithmétique – géométrie - astronomie - tables solaires et calendrier - cosmographie - géographie et histoire de l'Algérie, du grands Maghreb et de l'Afrique- médecine - poésie – littérature.

La BNA a à charge plusieurs fonctions, parmi ces fonctions nous pouvons citer :

- Collecter le patrimoine, conserver la culture du rapport national et assurer l'ouverture au patrimoine mondial.
- Acquérir, traiter, conserver et distribuer des livres et des documents écrits et audio-visuels au territoire national et à l'étranger.
- Recueillir des groupes de manuscrits, monnaies, médailles et documents rares et précieux d'importance nationale et ajuster l'indice de sortie.
- Être maintenue sur l'ensemble des divers documents, publications et manuscrits relatifs à l'Algérie, algériens ou autre, et publiées à l'étranger, ainsi que des cartes, des dessins et compositions musicales, audio-visuelles du son.
- Participer à l'élaboration de la recherche.

3.2. Centre national de manuscrit d'Adrar [27]: Créé par décret n°06 - 10 du 15 janvier 2006, il a pour missions principales :

- La conservation des manuscrits selon les méthodes scientifiques modernes
- La réalisation de l'inventaire général et la classification des manuscrits
- Le catalogage scientifique des manuscrits

A cet effet, le centre dispose d'un Service de reprographie qui a pour mission de :

- Numériser et photographier les manuscrits en utilisant les différents moyens modernes de photographie (photographie numérique, scanner).
- Archivage des documents numériques et création d'une banque de données.
- Création d'une bibliothèque virtuelle accessible.

LA DEUXIÈME PARTIE :
DEGRADATIONS ET
RESTAURATIONS DES
IMAGES DES DOCUMENTS



CHAPITRE 4 :
DÉGRADATIONS DES
IMAGES DES
DOCUMENTS

Sommaire :

La numérisation des manuscrits nous a permis de voir leur état. Dans ce chapitre nous allons voir les différents types de dégradations et leurs causes principales. Nous avons divisé ce chapitre en :

- 1- Introduction
- 2- Qu'est ce qu'une dégradation ?
- 3- Les agents de dégradation
- 4- Typologie de dégradations
- 5- Contrôle des conditions de préservation
- 6- Mesures de prévention
- 7- Conclusion

1. Introduction :

Après la numérisation des documents nous distinguons souvent plusieurs défauts. Ces défauts sont causés parfois par des modifications faites par l'homme (soulignement, annotations...) ou bien par l'état de conservation du document (vieillesse du support, l'humidité du lieu de stockage, qualité du papier..). D'autres sont causées par le processus de numérisation (résolution insuffisante...).

2. Qu'est ce qu'une dégradation ? La dégradation est tout changement négatif ou indésirable dans une image, elles ont plusieurs causes ou origines et l'accumulation de ces problèmes et défauts rend l'extraction de l'information très difficile.

3. Les agents de dégradations [14, 16,18]: Tous les matériaux organiques entrant dans la composition des documents graphiques sont extrêmement fragiles et facilement dégradés par des agents physiques, chimiques et biologiques. A cela viennent s'ajouter les dangers d'inondation et d'incendie. L'homme peut également, d'une façon délibérée ou inconsciente, être à l'origine de très grands désastres : le vol, la guerre ou les émeutes, les manipulations trop brutales, l'application des techniques de conservation et de restauration inadaptées sont des facteurs de destruction.

Nous devons donc étudier minutieusement tous les problèmes relatifs à la conservation de ces documents. A cette fin, il est nécessaire de connaître le mieux possible les différents agents de dégradation, ainsi que leurs effets sur les œuvres à conserver, de manière à mieux les combattre.

Les documents, s'ils ne sont pas conservés dans un gaz inerte (ce qui ne peut être le cas que pour quelques documents d'une très grande valeur historique ou artistique), sont soumis à un air ambiant qui peut, dans certains cas, leur être néfaste et provoquer des phénomènes de dégradations physiques, chimiques et biologiques.

Nous étudierons ici l'ensemble de ces facteurs.

3.1. Les agents de dégradations biologiques: Parmi les nombreux facteurs de dégradations des documents graphiques, ce sont certainement les micro-organismes et les insectes qui causent les dégâts les plus fréquents et les plus étendus.

3.1.1. Les champignons (papyriques, moisissure, mэрule): Les champignons sont des facteurs dangereux pour les dégradations des documents. Parmi eux nous pouvons citer :

- Les champignons papyriques qui s'attaquent tout particulièrement aux vieux livres, aux parchemins, aux reliures.
- Les champignons appelés moisissures, qui sécrètent des pigments qui se diffusent dans le papier en laissant des taches de différentes couleurs plus ou moins intenses.
- La mэрule qui est un champignon résistant et très nuisible aux matériaux sur lesquels elle se développe; elle peut se présenter sous des aspects très différents selon les conditions d'aération, de lumière et de localisation. Ses principales formes subissent une évolution : coussinets ouatés, fins filaments noirâtres (forme pauvre) ou toile; palmettes divergentes; cordonnets blancs ou rhizomorphes longs et assez épais; plaques orangées qui sont des réceptacles fertiles. Elle est donc assez difficile à reconnaître pour le non-spécialiste.

3.1.2. Les bactéries : Elles sont également des éléments aidant aux dégradations des documents, comme les taches brunes sur les manuscrits ou les gelées brunes sur le papier.

3.1.3. Les insectes : Les insectes qui ravagent les fonds de bibliothèques et d'archives sont nombreux et appartiennent à des espèces variées. Nous pouvons citer :

- **Les Thysanoures :** Seules les espèces appartenant à la famille des lépismes sont nuisibles pour les documents d'archives, notamment le *Lepisma saccharina*, connu sous le nom de "poisson d'argent". Leurs nourritures favorites sont la colle, l'amidon; aussi s'attaquent-ils surtout aux reliures dont ils "grignotent" la surface.
- **Dictyoptères. :** Il s'agit des blattes et en particulier de la blatte germanique (*Phyllodronia germanica*) et de la blatte orientale (*Blatta orientalis*). Ces insectes, fuyant la lumière, pullulent dans les endroits sombres, chauds et humides; ils endommagent surtout les reliures des registres.

3.1.4. Les rongeurs : Tous les rongeurs peuvent causer des dégâts très importants dans les fonds de bibliothèques et d'archives. Dans certains cas, les dégradations peuvent atteindre jusqu'à 20 % des documents. Ils dévorent avec la même avidité les papiers, les cuirs, les parchemins et les colles. A l'inverse de certains insectes qui provoquent des altérations souvent lentes et limitées, les rongeurs sont capables de détériorer complètement un document en très peu de temps.

3.2. Les agents de dégradations physico-chimiques : Les altérations provoquées par ces agents se manifestent souvent par un changement de couleur et une fragilité plus ou moins prononcée des documents altérés. Les dégradations physiques sont produites par la lumière, la chaleur et l'humidité, alors que les dégradations chimiques sont presque exclusivement dues à la pollution atmosphérique.

3.2.1. La lumière: La lumière est formée d'ondes électromagnétiques comparables en tous points aux ondes de radio ou de télévision, mais de longueur plus courte.

- **Le mécanisme de la détérioration photochimique :** Le cuir et le parchemin sont généralement moins sensibles à l'effet de la lumière. En revanche, les encres métallo galliques peuvent pâlir dans certains cas. D'autres facteurs interviennent également dans le processus de détérioration, en particulier le temps d'exposition et le niveau d'éclairement. En effet, si ces deux facteurs ne peuvent empêcher des réactions photochimiques, ils les modifient notablement.

- **Les sources lumineuses :** Les sources lumineuses, qu'elles soient naturelles (soleil...) ou artificielles (lampes à incandescence ou tubes fluorescents), émettent un rayonnement qui contient généralement, en plus des radiations visibles, une certaine proportion d'ultraviolets et d'infrarouges nuisibles pour les matériaux organiques.

3.2.2. L'humidité et la température: Les matériaux constituant les documents graphiques sont extrêmement sensibles aux variations climatiques. Le papier, essentiellement composé de fibres cellulosiques, est très hygroscopique. Ses propriétés physiques et chimiques dépendent donc de la teneur en eau de l'atmosphère. Il en est de même pour les manuscrits. Cette humidité exerce une action très marquée sur les propriétés chimiques et physiques des matériaux organiques. Elle favorise également des dégradations biologiques.

- **Dégradation physique :** Les matériaux hygroscopiques, en particulier le papier et le manuscrit, gonflent lorsqu'ils absorbent de l'humidité et se rétractent lorsqu'ils la libèrent. Cela entraîne d'importants changements dimensionnels : perte d'élasticité, de souplesse et de solidité. Ce phénomène est particulièrement frappant dans le cas des manuscrits enluminés; l'inégalité des tensions entre le support et la couche picturale aboutit forcément à une séparation des couches. C'est pourquoi de brusques variations hygrométriques représentent un danger très important. Si un excès d'humidité est à proscrire, une très grande sécheresse est également nuisible.

En effet, le papier a besoin d'un taux d'humidité assez important (environ 50 %) pour garder sa souplesse et son élasticité : si ce taux s'abaisse et devient inférieur à 40 %, le papier et surtout les colles deviennent cassants et finissent par tomber en poussière, tandis que les reliures se craquellent.

- **Dégradation chimique :** Du point de vue chimique, la chaleur humide entraîne une hydrolyse des molécules qui, de ce fait, se subdivisent en chaînes moléculaires plus petites. Ce phénomène a déjà été étudié par Chapman, qui, en 1915, compara l'état d'une collection de livres dont une partie avait été conservée aux Indes et l'autre au Royaume-Uni. Sur des lots de huit ouvrages, les exemplaires conservés en Inde, dans un climat humide et chaud, étaient tous attaqués, tandis qu'au British Muséum de Londres six étaient intacts, un était piqué et un autre décoloré. Les mêmes constatations ont été faites au sujet de papiers conservés, les uns dans des

régions montagneuses (climat sec et frais), les autres dans des régions maritimes (air salin très humide). Il en est de même pour le collagène, constituant essentiel du cuir et du parchemin.



Figure 4.1 : Exemple de dégradations chimiques

- **Dégradation biologique :** L'humidité et la chaleur sont des facteurs essentiels de germination des spores des champignons et de fructification des bactéries, toujours présentes en suspension dans l'atmosphère. Il suffit que les conditions de température et d'hygrométrie soient favorables (humidité relative supérieure à 65 %) pour que les spores des champignons, au contact du papier, fructifient en puisant leur nourriture dans ce support dont elles affaiblissent très rapidement la résistance.



Figure 4.2 : Exemple de dégradations biologiques

3.2.3. La pollution atmosphérique : (dégradation du manuscrit lui-même) Le problème de la dégradation des œuvres d'art par la pollution atmosphérique n'est pas nouveau : en 1850 déjà, Eastlake et Faraday étudiaient les moyens de protéger des peintures de la National Gallery. Cependant, avec le développement industriel et l'accroissement de la circulation automobile, les phénomènes de dégradations dus à la pollution se sont très sérieusement accrus au cours des vingt dernières années. En effet, la majeure partie des agents polluants provient de la combustion du charbon et du fuel, du chauffage domestique, etc.

3.2.4. Dégradation dues à la capture : (dégradation de l'image) 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.

3.3. La mauvaise qualité des matériaux constitutifs des documents : Le papier, était autrefois constitué essentiellement de fibres cellulosiques de coton ou de lin, résistait aux atteintes du temps. Il n'en est pas de même pour certains papiers à base de pâte de bois, encollés à la colophane en milieu acide. En effet, l'acidité initiale déjà élevée de ces papiers peut considérablement augmenter en vieillissant. Ce qui entraîne, comme dans le cas de la pollution, l'hydrolyse de la cellulose, d'où une perte de sa résistance mécanique.

Enfin, la stabilité des encres dépend des produits qu'elles renferment. Les encres au carbone, généralement indélébiles, deviennent parfois pulvérulentes par décomposition du liant. Elles peuvent également s'écailler lorsqu'elles sont étalées sur un support non absorbant, tel que le parchemin, auquel elles adhèrent mal. Quant aux encres métallo galliques, à base de sulfate de fer, elles peuvent, au contact de l'humidité, libérer de l'acide sulfurique qui ronge le papier et le parchemin et les transforme en « dentelle ». Par ailleurs, il arrive que ces encres pâlissent et parfois même se décolorent complètement. Ce phénomène est dû à une décomposition partielle des substances tannantes qui ne peuvent plus s'assembler aux particules métalliques restant à la surface du document.



Figure 4.3 : Document dégradé par l'acidité de l'encre

3.4. Les sinistres naturels ou accidentels :

3.4.1. Les inondations et cyclones: L'eau entraîne souvent des dommages considérables pour les documents graphiques. L'origine de ce genre de sinistre peut être soit naturelle (crues de fleuves ou rivières, orages, tempêtes), soit accidentelle (rupture de canalisations, fuite de toiture, murs lézardés, eau utilisée lors d'incendies). En quelques minutes, des dizaines de milliers de livres, manuscrits et liasses d'archives peuvent être ainsi inondés. Le sauvetage des documents ainsi immergés est une des principales préoccupations de tous les responsables de collections. Cette tâche est difficile, car il faut agir très vite et sur une quantité très importante de documents.

3.4.2. Les vents de sable: Dans certaines régions au climat très sec, le vent transporte des particules minérales résultant de l'érosion des roches. Certaines de ces particules, composées de minéraux très durs (silice, matériaux volcaniques), exercent sur les documents d'archives et de bibliothèque une action abrasive qui peut aller jusqu'à l'effacement de l'écriture.

3.4.3. Le feu: De tous les ennemis des documents graphiques, le feu est évidemment, avec l'eau, le plus spectaculaire; il est aussi le plus redoutable, puisqu'il menace tous les matériaux d'une destruction complète, rapide et généralement irréversible. Malheureusement, depuis une vingtaine d'années, le nombre de foyers d'incendie n'a cessé de croître en fonction de l'emploi intensif de matériaux très inflammables et mal ou non traités. En effet, les matières plastiques ont trop souvent remplacé les matériaux traditionnels pour l'aménagement des locaux. Les sacs et enveloppes en polyéthylène ou même en polychlorure de vinyle sont utilisés à la place du papier kraft. Ces substances, si elles ne sont pas ignifugées, sont très inflammables et permettent au feu de se propager très rapidement. Par ailleurs, leurs produits de combustion sont très toxiques pour l'homme.

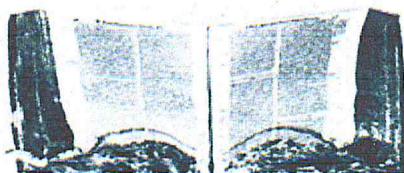


Figure 4.4 : Exemple de dégradations d'origine des sinistres naturels ou accidentels

3.4.4. L'eau : L'eau constitue une menace importante pour les collections. Les dommages causés peuvent avoir de multiples origines : fuites de canalisation, toitures non étanches, crues de rivière, ouragans, lutte contre des incendies. Lorsque les dégâts d'eau ne sont pas découverts à temps ou lorsque les mesures de sauvetage ne sont pas adaptées à l'ampleur des dégâts, ils entraînent en général des dégradations supplémentaires par des moisissures.

3.5. Les méfaits dus à l'homme : Trop souvent, la malveillance ou la négligence de l'homme entraîne des détériorations pouvant aller jusqu'à la destruction complète des documents.

3.5.1. Le vandalisme et le manque de soin des lecteurs ou des visiteurs: Si les actes de vandalisme tendent à diminuer grâce à la vigilance des conservateurs et des gardiens, on se doit cependant de signaler le cas des manuscrits à peintures qui ont été découpés, des notes marginales manuscrites gommées, des pages de livre et de manuscrit arrachées, des feuillets souillés par des tâches de graisse ou d'encre à stylo, de crayon feutre ou à bille.

3.5.2. Les manipulations maladroites: Celles-ci ont généralement lieu lors de l'inventaire, du classement ou de la communication des documents. Parmi les nombreux faits constatés au cours des inspections des magasins, signalons les plus marquants :

- L'estampillage des collections est trop souvent effectué avec des encres inadaptées et à des endroits mal choisis.
- Les étiquettes (en particulier les étiquettes auto-adhésives) portant la cote du document peuvent dans certains cas entraîner une détérioration ponctuelle des reliures, Les livres sont quelquefois mal rangés sur les rayonnages les grands formats écrasent les petits.
- Les documents trop serrés risquent de se déchirer lorsqu'on les retire des étagères.
- Les reliures non traitées se dessèchent, et les plats finissent à la longue par se détacher du dos.



Figure 4.5 : Exemple de dégradation d'origine humaine

3.6. les agents des dégradations dues à la capture : Nous pouvons citer :

3.6.1. **Dégradations matérielles:** Nous distinguons deux façons de numériser : la numérisation directe du document original et la numérisation indirecte faite à partir d'une copie du document original. La numérisation de copies est plus simple à mettre en œuvre et à automatiser mais la qualité des images produites est très inférieure à celle obtenue par la numérisation directe du document original.

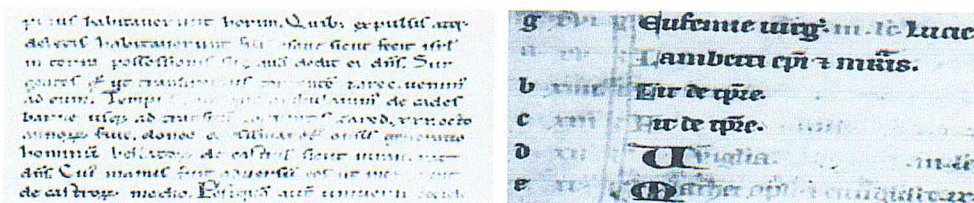


Figure 4.6 : Exemple de dégradations matérielles : Perte d'information couleur sur les images numérisées des microfilms

3.6.2. **Dégradations numériques :** Les dégradations numériques sont les conséquences de plusieurs causes 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'exemple, 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.

3.6.2.1. L'inadéquation de la résolution à la taille des textes : Une bonne résolution aide à mieux distinguer les détails fins, mais elle nécessite la réservation d'un grand espace de stockage. Pour réduire les coûts tout en générant des images numériques exploitables, il faut trouver un compromis entre la taille des images et leur qualité. En outre, lors de la numérisation d'une collection, la résolution est fixée à l'avance indépendamment des contenus des textes.

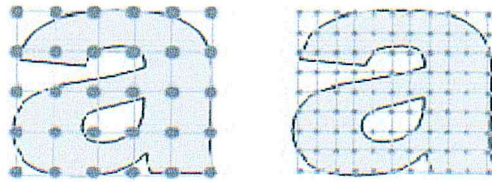


Figure 4.7 : Finesse de la grille pour une faible et une forte résolution

3.6.2.2. Le problème de la numérisation en binaire : La binarisation est un traitement irréversible qui permet de transformer une image en niveaux de gris ou en couleur en une image noir et blanc en fonction d'un seuil à définir. Toute valeur de pixel supérieure à un seuil fixe aura la valeur 255 (blanc), 0 (noir) sinon. Un document original dégradé aboutit généralement à une image binaire de mauvaise qualité. Les dégradations du support, comme l'apparition du verso par transparence sur le recto et la présence de tâches d'humidité, sont conservées dans l'image binaire.

En outre, le bruit du capteur étant plus élevé sur les zones de fortes variations d'intensité, la binarisation fait systématiquement apparaître des irrégularités aléatoires sur les contours des caractères. Les images en couleur ou en niveaux de gris sont beaucoup plus lisibles et plus facilement interprétables par une machine que les images binaires, les niveaux de gris ou la couleur apportent une information supplémentaire qui compense le manque de résolution.

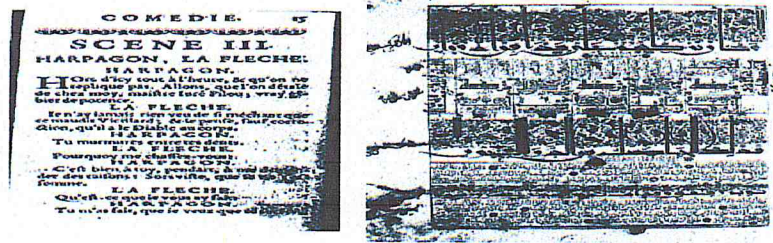


Figure 4.8 : Exemple de problème de la numérisation en binaire : Binarisation de documents dont le support est dégradé

4. Typologie des dégradations [20]:

4.1. **Dégradation du fond (image du papier):** Plusieurs effets indésirables sur le fond du document peut déclencher plusieurs types de dégradations (annotations, tâches dues à l'humidité, superposition du verso et ajouts indésirables...). Ces différentes dégradations se présentent sous forme de couches superposées à différents niveaux de gris. Restaurer le document revient alors à utiliser des méthodes de segmentation/classification.

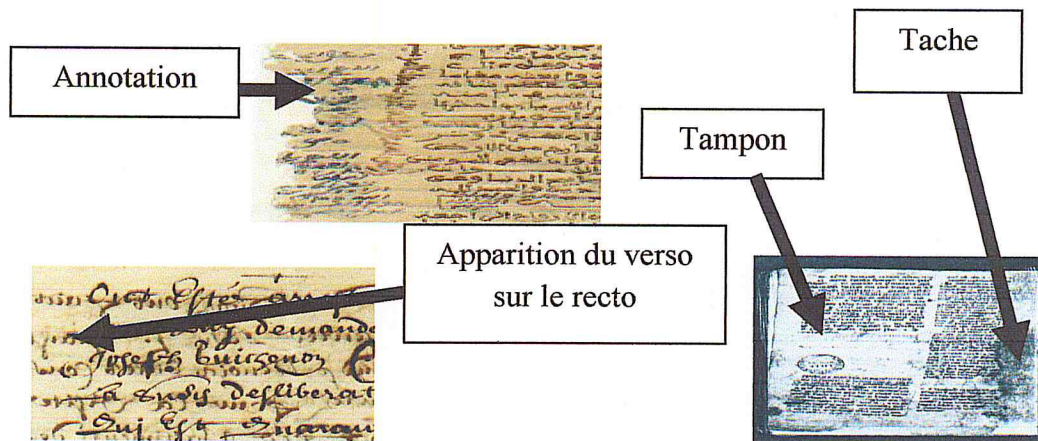


Figure 4.9 : Exemple de dégradations du fond

4.2. **Dégradation des formes (images des traits, caractères, dessins..):** Ces dégradations apparaissent sur les caractères (rupture des traits, caractères tronqués, vide dans les traits...). Les trous sur le support sont associés à ce type de dégradation s'il engendre des pertes de l'information. La compression JPEG est aussi une autre forme de cette dégradation puisqu'elle peut induire des dégradations de caractères. Cette classe est caractérisée par une perte aléatoire de l'information. Restaurer

revient ici à réparer les formes des traits ainsi qu'à extrapoler l'information perdue pour améliorer les performances de l'analyse de ses images.

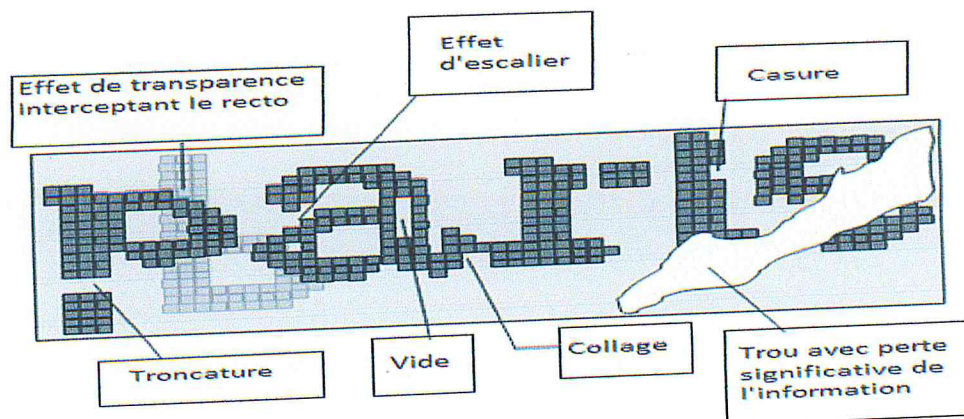


Figure 4.10: Exemple de dégradations des formes

5. Mesure de prévention : Il est possible de ralentir la dégradation des documents en agissant sur les facteurs tels que la température, l'humidité relative, la lumière, les agents de détérioration biologiques (moisissures, insectes et rongeurs), chimiques et mécaniques (polluants et poussières). Par ailleurs, il est possible d'éviter ou de limiter la dégradation causée par l'homme.

L'environnement idéal pour les collections de bibliothèques est un environnement où la température et l'humidité relative sont contrôlées, où des polluants sont absents, où une bonne ventilation de l'air est assurée, où la lumière est contrôlée, où les moisissures, insectes, rongeurs sont absents, où des bonnes pratiques de maintenance et de sécurité (feu, eau, vol) existent.

5.1. Recommandations en matière de conservation des documents : Voici des ci-dessus des tableaux récapitulant les différentes recommandations afin de préserver les documents.

Supports	Conditions environnementales				
	Température		Humidité relative		Eclairage
	Niveau	fluctuation	niveau	fluctuation	Intensité
	°C	°C	%	% / 24h	lux (1)
Parchemin et cuir	18	2	50-60	5	50-200 (2)
Papyrus	18	2	50-60	5	50 (2)
Papier	18	2	45-55	5	50-200 (2)

(1) l'effet de la lumière étant cumulatif, il faut multiplier le nombre de lux par le nombre d'heures d'exposition. Il est recommandé de ne pas dépasser 720 h/an (soit 8 heures par jour pendant 3 mois).

(2) l'intensité d'éclairage est fonction de la fragilité des supports.

Tableau 4.1: Conditions relatives à la température et l'humidité et l'éclairage

Supports	Conditions environnementales				
	Pollution atmosphérique			Agents chimiques et biologiques	Chocs
	dioxyde de soufre	oxyde d'azote	Ozone		
parchemin et cuir				Isolation des supports contaminants, entretien régulier	protection adéquate, bonne pratique de manipulation
papyrus				Idem	Idem
papiers	5 à 10	5 à 10	5 à 10	Idem	Idem

Tableau 4.2: Conditions relatives à la pollution et les agents chimiques et biologiques et choc

5.2. Prévention des dégradations dues à l'homme : Les dégradations causées par l'homme peuvent être minimisées. La négligence est souvent liée à l'absence de formation aux pratiques de conservation préventive et à la méconnaissance des conséquences de mauvaises conditions de conservation et de manipulation. La formation initiale et continue du personnel doit être une priorité absolue, surtout dans le domaine de la préservation des documents audiovisuels dont l'importance relative aux autres collections est en constante augmentation. Les dégradations par vandalisme et vol doivent être diminuées par la mise en place de systèmes de protection et de surveillance appropriés.

5.3. Prévention des désastres : Quant aux désastres (feu, inondations, tremblement de terre...), l'objectif de toute politique de conservation préventive doit être de limiter

autant que possible les risques potentiels et d'augmenter les chances de récupération des matériaux en cas de désastre. Il est fortement recommandé de préparer à l'avance un plan d'intervention en cas de désastre, afin de diminuer le temps d'intervention. Pour cela, il est essentiel de déterminer les procédures à appliquer, de les tester et de les inclure dans les exercices périodiques de sécurité impliquant tous les personnels de la bibliothèque. Il faut également tenir à jour une liste d'adresses pour savoir à qui on peut s'adresser en cas de désastre (pompiers, sociétés de transports frigorifiques, de congélation et de lyophilisation, par exemple).

5.4. Prévention de la pollution : Lorsque les magasins sont climatisés, le niveau de pollution peut être réduit grâce à l'utilisation d'une filtration d'air efficace. Les filtres électrostatiques sont cependant à éviter, car ils dégagent de l'ozone, un oxydant puissant. Dans tous les cas, une maintenance régulière des magasins et des collections par dépoussiérage minimisera les dégradations. Quant aux aspirateurs utilisés pour le dépoussiérage, ils doivent être obligatoirement munis de filtres absolus pour éviter la dispersion des spores de micro-organismes dans l'air.

5.5. Prévention des altérations dues à la lumière : La lumière naturelle, les lampes fluorescentes et les lampes tungstène-halogènes, contiennent toutes une partie non négligeable de rayons ultraviolets désastreux pour les matériaux organiques qui entrent dans la composition de la grande majorité de nos documents. Pour cette raison, ils doivent être impérativement éliminés pour éviter des dégradations irréversibles.

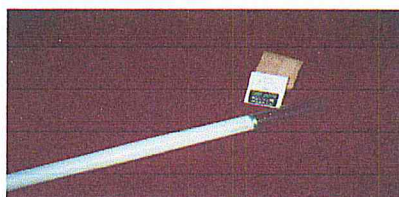


Figure 4.11: Une lampe à fluorescence équipée de filtres anti-UV (ultra-violette)

Naturellement, le rayonnement visible doit lui aussi être contrôlé. Il faut l'éliminer dans tous les cas en dehors d'une activité humaine. Dans le cadre d'une exposition, on respectera une exposition lumineuse, c'est-à-dire le produit d'un niveau d'éclairage par le nombre d'heures d'exposition, inférieur à 84 Klux.heures/an, voir pour certains documents en papier extrêmement sensibles (à

base de pâte de bois) on réduira cette valeur à 12,5 Klux.heures/an. Ces valeurs s'expliquent par le fait que l'action photochimique des rayonnements électromagnétiques est cumulative et, par exemple, que les dégradations causées à un document par un éclairage de 50 lux pendant 10 000 heures (3 ans à raison de 8 heures par jour) seront identiques à celles provoquées par un éclairage de 1 000 lux durant 500 heures (2 mois environ). Là aussi, vitrages, stores, films et autres systèmes permettent de maîtriser l'action de la lumière.

Les documents graphiques et photographiques sont à considérer comme faisant partie, dans leur grande majorité, des objets patrimoniaux les plus sensibles à la lumière. En fonction des matériaux les constituant ainsi que de leur état de conservation, il est convenu de les répartir en deux catégories suivant leur sensibilité à la lumière, soit :

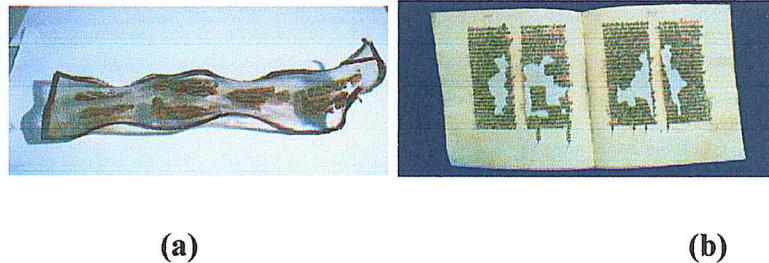
- Documents fortement sensibles (photos noir et blanc etc.).
- Documents très fortement sensibles (tirages noir et blanc sur papier plastifié, etc).

5.6. Prévention des altérations dues à la température et l'humidité relative : Si la plupart des facteurs de dégradation peuvent être minimisés ou même exclus, il est souvent difficile de maîtriser correctement les facteurs "température" et "humidité relative". En effet, il s'agit là de deux paramètres interdépendants dont les effets sur les œuvres sont plus variés et plus complexes que ceux des autres paramètres.

5.6.1. Influence de la température : Les matériaux qui sont en général sensibles aux fluctuations de température sont les objets composites dont les matériaux constitutifs possèdent des caractéristiques de dilatation différentes en fonction de la température.

Des températures trop basses, peuvent fragiliser des matériaux plastiques qui deviennent en général vitreux et de plus en plus friables. Des températures trop élevées accélèrent la dégradation des matériaux instables (papiers acides, films de nitrate et diacétate de cellulose, films en couleur). Pour l'archivage des films, on a souvent recours à un stockage à - 18° C. Il est devenu courant de faire la distinction entre un stockage de conservation et un stockage permettant un accès aux documents dans des conditions climatiques ambiantes pour la consultation. Théoriquement, chaque baisse de température de 10° C va doubler la durée de vie de ces matériaux. Il

n'est cependant pas très économique de maintenir constamment des matériaux bien en dessous de la température ambiante. Dans le cas d'une panne du système de refroidissement, les dégâts dus à la condensation de la vapeur d'eau à la surface froide des matériaux peuvent également être préjudiciables.



(a) : Conditionnement de fragments de manuscrits entre des plaques de diacétate de cellulose. Ce conditionnement est en pleine décomposition.

(b) : Manuscrit dégradé par les encres ferro-galliques, stabilisé avec du papier japon.

Figure 4.12 : Méfaits de la température sur les documents

5.6.2. Influence de l'humidité relative : Dans le cas des matériaux organiques, constituants majoritaires des documents de bibliothèques ou d'archives, les niveaux et fluctuations de l'humidité relative ont beaucoup plus d'incidence sur la conservation des collections que les niveaux et les fluctuations de la température (proportion 95 % contre 5 %). Quels sont alors les niveaux et variations acceptables de l'humidité relative ? Il n'existe pas de normes en matière d'humidité relative, mais seulement des recommandations.

La plupart des musées, archives et bibliothèques aux États-Unis et en Europe ont adopté le niveau de $50 \% \pm 5 \%$ d'humidité relative. Des musées dans des pays à climat froid en hiver (Scandinavie et Canada) préconisent des niveaux de $40 \% \pm 5 \%$, car des valeurs plus élevées risquent de provoquer des phénomènes de condensation sur les surfaces froides des bâtiments (vitres, murs). En fait, ces valeurs ont été adoptées en se basant sur la faisabilité technique plus que sur la connaissance de l'impact de ces niveaux d'humidité sur la conservation des collections.

Les conditions d'humidité contre-indiquées peuvent être classées en trois catégories : humidité relative trop élevée, trop basse, fluctuante.

5.6.2.1. humidité relative trop élevée : L'humidité excessive (supérieure à 65 %) entraîne une prolifération de moisissures et une corrosion rapide des métaux. Le risque croît rapidement avec chaque hausse au-delà de ce seuil. Par exemple, à température ambiante, le temps de développement de moisissures sera de quelques semaines à 75 % d'humidité relative, alors qu'à 90 % d'humidité relative il sera de quelques jours.

5.6.2.2. humidité relative trop basse : L'humidité joue un rôle important dans les processus d'altération chimique des matériaux. En théorie ces processus ne sont stoppés qu'à 0 % d'humidité relative. En revanche, une humidité relative trop basse entraîne la déshydratation des matériaux organiques et leur fragilisation.

Une humidité relative excessivement basse peut favoriser l'apparition d'électricité statique, ce qui dégrade la qualité de la restitution des documents sonores et audiovisuels.

5.6.2.3. humidité relative fluctuante : Les fluctuations d'humidité relative sont contre-indiquées pour la majorité des collections, car elles induisent des contraintes mécaniques plus ou moins fortes (dilatation, rétrécissement). Un cycle unique d'une variation brutale de l'humidité relative peut se traduire par des fissurations visibles sur des objets (ivoires par exemple). Certains objets qui ont subi une restauration récente sont particulièrement sensibles aux variations de l'humidité relative. Des cycles répétés de variations d'humidité relative se traduisent par une fatigue mécanique qui va fragiliser graduellement l'objet en question.

Actuellement, les recommandations sont de plus en plus différenciées en fonction des types de matériaux – organiques et minéraux – et de la composition des objets – homogène ou composite. Dans le cas d'une collection constituée, comme cela est le cas en général d'une multitude de matériaux, il faut soit trouver un compromis, soit isoler les documents les plus fragiles pour les entreposer dans des magasins à climatisation spéciale ou pour leur confectionner des "microclimats" (vitrines conditionnées, boîtes ou cadres contenant des substances "tampon" régulant l'humidité relative).

En général, on constate que la consigne du niveau de l'humidité relative $\pm 5 \%$ a tendance à être élargie à une variation de $\pm 10 \%$ autour d'une valeur moyenne.

Ainsi, pour beaucoup de collections, des niveaux d'humidité relative compris entre 40 et 60 % sont parfaitement admissibles (soit $50 \% \pm 10 \%$). Ces nouvelles recommandations en matière d'environnement offrent aux musées, archives et bibliothèques une certaine marge pour concilier les restrictions financières, les conséquences du caractère parfois historique du bâtiment et les exigences en matière d'humidité relative.

Dans les régions tempérées, les niveaux optimaux de température et d'humidité relative pour des collections traditionnelles se situent autour de 18° et 55 % d'humidité relative. Des fluctuations sont généralement admises dans la gamme comprise entre 16° et 21° et 40 et 60 % d'humidité relative. En revanche, pour les collections de photographies, de microfilms, de supports mécaniques, magnétiques et optiques, on préconise des valeurs plus basses de température et d'humidité relative : pas plus de 16° et 40 % d'humidité relative dans les magasins de stockage.

5.6.3. La conservation dans des régions climatiques difficiles : le cas des bibliothèques des zones tropicales : La conservation de collections patrimoniales dans les régions géographiques où les conditions climatiques sont éloignées des conditions générales d'une partie de l'hémisphère Nord est un problème difficile à résoudre. Pour respecter les données idéales de température et d'humidité relative, il faudrait imposer une climatisation performante qui serait elle-même à la source de deux problèmes aux conséquences non négligeables : les coûts de fonctionnement et de maintenance, et les risques de chocs thermiques.

Plus la température et l'humidité relative extérieures s'éloignent des conditions idéales, plus il est difficile de maintenir à l'intérieur une température et une humidité relatives recommandées. Pour y parvenir, il faut dépenser beaucoup d'énergie, dont il résulte un coût élevé de fonctionnement; ce qui suppose également une installation particulièrement efficace et à l'abri des pannes éventuelles, ce qui impose une maintenance régulière et coûteuse. Par ailleurs, et pour des raisons de confort, la différence de température entre l'extérieur et l'intérieur ne peut être trop grande (au-delà de 10° la différence est ressentie comme perturbante par l'organisme humain) : en adoptant des conditions de conservation "idéales", soit on contraint les documents à subir des chocs thermiques importants lors du passage des magasins aux

salles de lecture (avec un redoutable phénomène de condensation sur les surfaces froides comme les films et les bandes magnétiques), soit on contraint les usagers à subir eux-mêmes des chocs thermiques en les obligeant à supporter des différences importantes de température lors de leur entrée dans la bibliothèque ou dans les salles de consultation, ce qui n'est guère admissible.

Les solutions recommandées sont de deux ordres :

- Solution techniquement performante avec recours à la climatisation quasi généralisée dans des bâtiments qui doivent être impérativement adaptés pour éviter toute déperdition d'énergie, avec les conséquences financières.
- Solution de compromis par le recours à des méthodes qui utilisent le plus possible les qualités de certains matériaux (briques, terre, etc., en usage dans de nombreux pays de la zone tropicale par exemple), utilisation de techniques architecturales qui permettent une circulation permanente et un renouvellement constant de l'air dans les zones de stockage pour éviter la stagnation propice au développement de moisissures, utilisation de zones de mise à température et à humidité progressives. Les collections ne "bénéficient" certes pas des conditions idéales comme en métropole mais elles ne sont pas soumises aux variations brutales d'humidité et de température qui constituent un des risques majeurs de dégradation des matériaux.

En complément, on instaurera une politique rigoureuse de surveillance des locaux: surveillance des variations climatiques (avec des thermohygromètres enregistreurs), surveillance des développements de moisissures et prolifération d'insectes.

On recommandera enfin de considérer avec soin la fin recherchée et les moyens dont on peut disposer afin de trouver la solution la plus acceptable à défaut d'être la plus satisfaisante pour la conservation des collections; dans cette démarche, on procédera de manière prudente, en évitant de généraliser hâtivement et en tenant compte de tous les facteurs qui font de chaque cas un cas particulier.

6. Conclusion :

Nous avons vu dans ce chapitre les éléments déclencheurs des dégradations des manuscrits et des images des documents et comment ceux-là affectent la qualité visuelle des images et des manuscrits et engendrent la perte de l'information. Sans oublier la typologie de ces dégradations proposée par Fadoua Drira dans [14] et à la fin on fait une petite étude sur les techniques de conservations des documents.

Enfin, nous concluons dans ce chapitre qu'afin de faire face à ces dégradations, des techniques de restauration particulières s'imposent.

CHAPITRE 5 :
RESTAURATION DES
IMAGES DES
DOCUMENTS

Sommaire :

L'étude des dégradations et des différents agents de dégradation rend l'extraction de l'information très difficile pour cela nous allons étudier dans ce chapitre les techniques de restaurations. Nous avons divisé ce chapitre en :

- 1- Introduction
- 2- Types de restauration
- 3- Méthodes de restauration supervisées
- 4- Méthodes de restauration non supervisées
- 5- Conclusion

1. Introduction :

Après avoir vu les manuscrits en général les manuscrits algériens en particulier et tous les effets indésirables sur ces documents, nous avons besoin de faire une petite étude sur les technologies de restauration qui existent en général afin de bien extraire l'information essentielle.

2. Type de restauration : [14]

2.1. La restauration physique : Le document ancien du patrimoine n'échappe pas aux dégradations dues aux méfaits du temps « vieillissement, acidité de l'encre, humidité, présence des insectes et modification des couleurs ... etc ». Le principal objectif de ce type de restauration est de redonner au document sa forme originale ce sont des méthodes de restauration physico-chimiques qui permettent de ralentir le vieillissement des documents et de renforcer les supports.

2.2. La restauration virtuelle : Elle utilise les méthodes de traitements d'images pour retrouver la forme originale des documents. Elle est donc complémentaire à la restauration physique, elle cherche à réparer les dégradations physiques des documents: elle permet par exemple de retrouver la couleur d'origine, de remplir les trous, d'atténuer le verso quand il traverse le recto, d'enlever les taches ou de corriger la courbure après la restauration de la reliure.

2.3. La restauration numérique : Elle regroupe l'ensemble des opérations de traitements informatiques qui ont pour objectif de retrouver la qualité de l'image originale avant sa dégradation par le processus de numérisation. Le traitement de

l'image permet de corriger les défauts qui apparaissent lors du processus de capture d'images. Ces défauts sont notamment des défauts géométriques (inclinaison du document, courbure apparente le long de la reliure...), des défauts optiques (éclairage non-homogène, netteté de l'image...), des défauts électroniques (bruit et qualité du capteur), des défauts informatiques (traitements destructifs de l'image sans sauvegarde des originaux, compression avec perte d'information...) ou même des défauts extérieurs dus aux conditions de la prise d'image (présence d'éléments non-désirés, extérieurs au document). Dans ce cas précis, le traitement numérique doit être le dernier recours si on ne peut ni refaire une autre capture d'image ni améliorer les conditions de la numérisation; il peut rarement retrouver l'information déformée par le processus de numérisation.

Les restaurations numériques et virtuelles ont toutes les deux besoin de l'outil informatique ou plus précisément le traitement d'images pour retrouver l'image la plus fidèle possible. Elles sont classifiées en :

- Méthodes supervisées
- Méthodes non supervisées

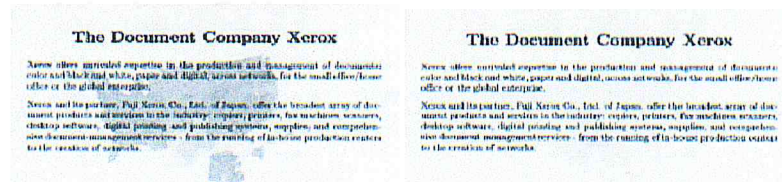
3. Méthodes de restauration supervisées : Les méthodes dites « non-aveugles » ou supervisées sont basées sur des informations externes fournies par l'utilisateur qui connaît à priori le contenu du document mais qui ne sont pas toujours disponibles. Ces méthodes sont très efficaces mais elles ne sont pas génériques pour traiter tous les types de documents. En outre, la restauration des formes de caractères par l'observation de toutes les occurrences du même caractère dans la page ne fonctionne que sur des documents imprimés où les caractères peuvent être segmentés et comparés. La séparation entre le recto et le verso sont les principaux travaux qui les utilisent.

3.1. Séparation recto/verso par recalage [28, 29]: Le recalage est le plus utilisé pour la suppression des marques du verso qui se présentent sur le recto.

Le principe est de mettre en correspondance les deux faces d'une même page pour extraire le texte original de chaque face.

- L'approche de Sharma applique les filtres linéaires sur les deux faces. Ce cas ne semble t'il pas été traité et nous donnons dans la figure 5.1 le seul résultat publié par

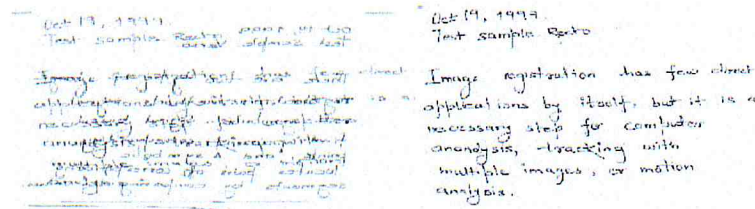
l'auteur, sans oublier de mentionner que les deux parties sont de natures différentes (texte et image).



(a) Recto du document dégradé (b) Recto du document restauré

Figure 5.1 : Suppression des taches en transparence par recalage

- La méthode proposée par E. Dubois et A. Pathak consiste à traiter les deux faces en même temps et à comparer les intensités des deux pixels. Si l'intensité du pixel du recto est moins élevée que celle du verso, sa valeur est conservée. La figure 5.2 montre un exemple d'application de cette méthode.



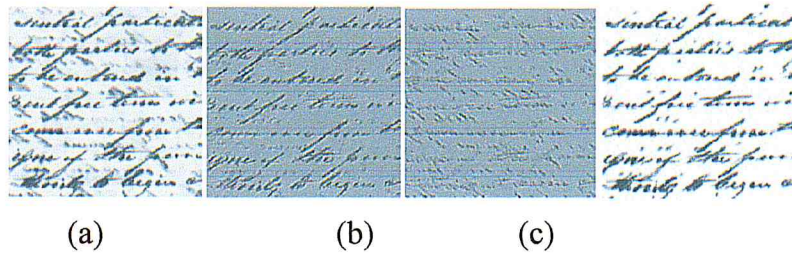
(a) Le recto d'un document de test avec une fuite significative du verso
(b) Le recto du document restauré après application de la méthode

Figure 5.2 : Exemple2 de suppression des taches en transparence par recalage

Si l'image du recto est géométriquement déformée par rapport à celle du verso, la méthode ne peut mettre une mise en correspondance parfaite des deux faces.

3.2. Séparation recto/verso par analyse de l'orientation [29]:Certaines écritures des documents sont orientées, l'orientation donc est efficace pour distinguer le recto du verso.

Q.Wang et al. proposent une analyse de l'orientation qui est capable de séparer les traits suivant les directions 45° et 135° . Ainsi, le recto va être obtenu dans l'image filtrée dans la direction 45° et le verso dans l'image filtrée dans la direction orthogonale à la première. Néanmoins, certains détails du recto peuvent se projeter sur l'image du verso cassant ainsi les caractères. Les auteurs proposent d'extraire du document original les parties manquantes sur un certain voisinage de taille 7×7 .



(a) image de test, (b) image associée à l'orientation 135o, (c) image associée à l'orientation 45o, (d) image restaurée de (a)

Figure 5.3 : Suppression des taches en transparence par analyse de l'orientation

Cette méthode est dans ces types de restauration car il faut connaître l'orientation du texte.

3.3. Séparation recto/verso par analyse des formes [30]:

- C. Wolf propose une méthode de séparation entre le recto et verso basée sur une régularisation de la segmentation par les champs de Markov. Cette séparation ne s'intéresse pas aux couleurs ou aux niveaux de gris de l'image. Les champs de Markov capturent la distribution spatiale des pixels et prédisent la configuration la plus probable à partir d'une configuration initiale donnée par la segmentation. Ils tendent à régulariser la segmentation pour préserver la continuité des traits du recto qui croisent ceux du verso.

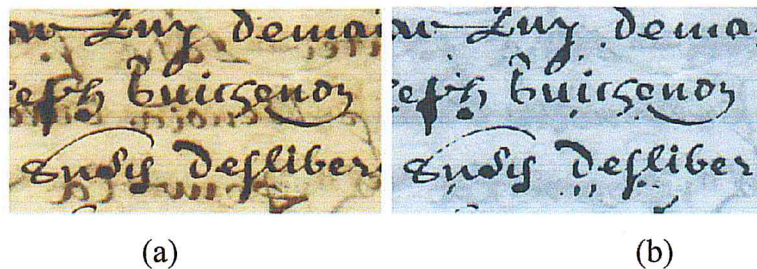


Figure 5.4 : (a) Image d'origine, (b) Image finale de recto séparée du verso

3.4. Séparation recto/verso par analyse de la couleur [31, 32, 33]:

Cette séparation est basée sur les couleurs de chaque face du document.

- E. Smigiel utilise une carte de Kohonen pour résoudre le problème de suppression des taches en transparence. Il détermine une classification des pixels en quatre classes (fond, texte, texte coloré et texte du verso). Cette approche nécessite une

fastidieuse phase d'apprentissage sur une portion représentative d'une page. La classification ne marche pas si la luminance varie sur une même page ou d'une page à l'autre. Les performances de la séparation par classification dépendent donc de l'apprentissage.

- U. Garain propose de séparer le recto du verso par une approche classique de segmentation en régions de couleur qui sont ensuite agrégées suivant des règles prédéfinies. Le résultat montre un effet de mosaïque autour des régions agrégées et présente de nombreuses fragmentations des traits. De plus, cette approche échoue sur cette image de référence (Figure5.5).

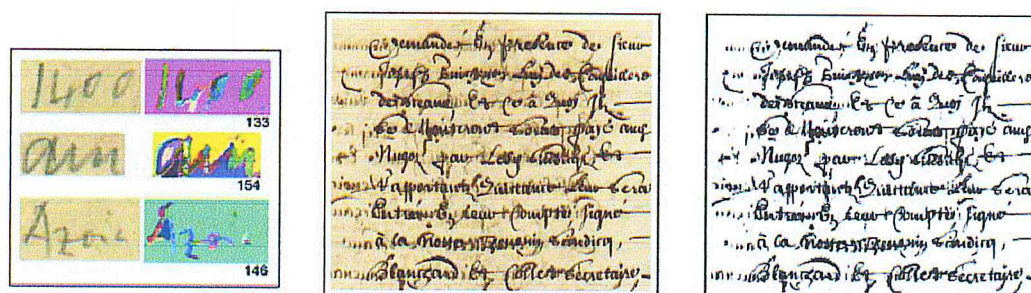


Figure 5.5 : Séparation recto/verso par segmentation et agrégation de régions couleurs : Résultats de la sur-segmentation en régions, image dégradée et restaurée

- La méthode de segmentation d'images couleurs d'Y. Leydier est basée sur la sérialisation de l'algorithme des nuées dynamiques ; elle donne de bons résultats pour la séparation recto/verso (Figure5.6). Il s'agit d'une classification automatique des couleurs sur une fenêtre glissante parcourant toute l'image. Durant le déplacement de la fenêtre, l'algorithme adapte automatiquement les centres des classes au contexte local. Cette approche permet donc de combiner les avantages d'une segmentation par classification et les méthodes de segmentation adaptatives classiques que l'on trouve dans le domaine du document. Son principal inconvénient réside dans la détermination manuelle du nombre de classes et des couleurs pour chaque classe. C'est donc une approche efficace mais entièrement supervisée qui nécessite l'intervention de l'opérateur pour chaque document.

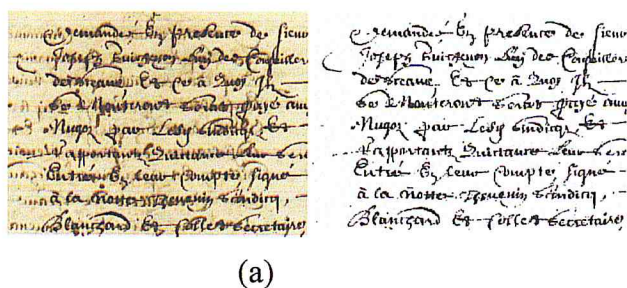


Figure 5.6 : Application de l'approche de [33] : (a) Image originale (b) Résultat pour $K=2$, le nombre de classes

3.5. Restauration des formes par modèle [14, 34,35] : Certaines restaurations des formes se sont orientées vers une définition de modèles pour refléter ces dégradations, mais aucun modèle ne peut décrire toutes les dégradations. On peut citer :

- Le modèle de défaut proposé par H. Baird, inclut un certain nombre de paramètres qui sont reliés aux conditions de numérisation de la page tel que: la taille du texte, la résolution, la rotation, le changement d'échelle, le défaut de jet d'encre.
- Le modèle proposé par Kanungo, suppose que les probabilités d'altération des formes ne dépendent que de la distance à leur frontière et suivent une loi normale. Ils simulent certains défauts, comme le flou, dus à l'impression et la numérisation, par des opérations de fermeture morphologique.
- Le modèle proposé par Zhang et Kanungo est défini à partir de la numérisation d'une image idéale de caractères. Il est basé ainsi sur les paramètres extraits de l'image dégradée produite.
- Baird élabore un modèle de dégradation utilisé pour entraîner un logiciel de reconnaissance optique de caractères et donc améliorer ses performances. La restauration par modèle conduit en général à des résultats intéressants. Mais, elle exige des hypothèses de déformation homogène sur l'ensemble de la collection.

Ces hypothèses figées ne restent applicables que pour des conditions très particulières d'acquisition. Une simple modification des conditions prises en compte (changement de la résolution ou de la taille du texte, mauvais positionnement de la page sur le scanner) met en échec tout le modèle proposé.

- B. ALLIER propose la restauration des caractères cassés par contour actif (snake). Cette restauration consiste à faire évaluer un contour actif, guidé par le modèle d'une

lettre entière sur une lettre cassée(Figure5.7).On ne peut appliquer cette méthode que si on dispose du modèle de la lettre à restaurer. De ce fait cette approche est difficile à utiliser. De plus, cette méthode nécessite une segmentation en caractères et une recherche de la forme entière à partir de la forme cassée.

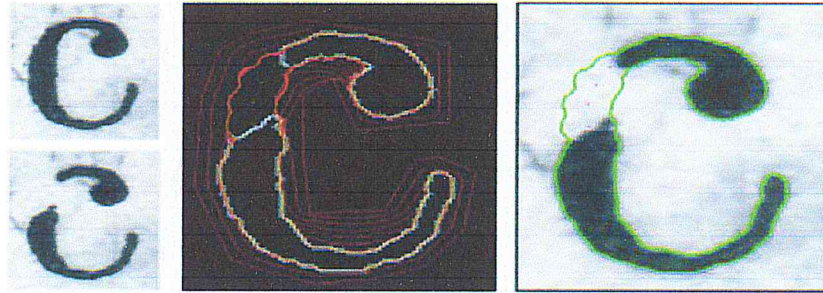
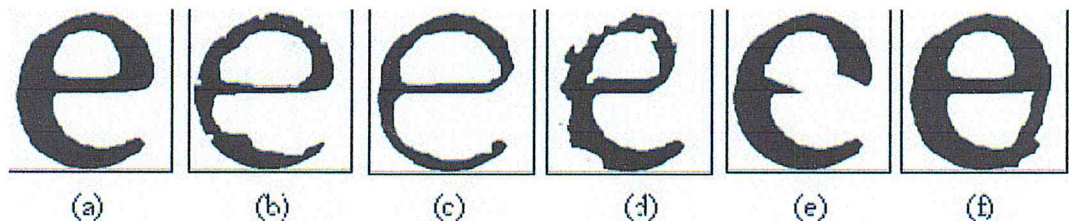


Figure 5.7 : Principe de la restauration d'une lettre cassée par les contours actifs

3.6. Restauration des formes par prototypage [36]:Le prototypage consiste à collectionner toutes les formes similaires dans la page. On l'applique généralement aux documents imprimés. À partir de la superposition de toutes les formes dégradées d'une même lettre, on construit le modèle de la forme idéale de la lettre. La figure 5.8 montre différentes occurrences dégradées du caractère «e». Malgré les différences générées par les dégradations partielles, pour les trois premières occurrences, la structure est bien préservée. Toutefois, la structure n'est plus préservée avec les autres occurrences.



(a) caractère idéal ; (b), (c) et (d) caractères dégradés avec structure du «e» conservée, (e) et (f) caractères dégradés avec structure confuse : (e) «e» ou «c», (f) «e» ou «o»

Figure 5.8 : Cas de dégradations de la lettre «e»

La technique proposée par Hobby et Baird est basée sur l'interpolation du caractère idéal en utilisant le principe du prototypage. Le principe consiste à

superposer plusieurs occurrences du caractère dégradé et à moyenner l'image résultante pour obtenir une approximation de l'image idéale du caractère. Chaque type de caractère dans le document sera ensuite substitué par son prototype (Figure5.9).

**So, we find special So, we find special
term rewriting for term rewriting for**

Figure 5.9 : Restauration par prototypage des formes dégradées de lettres

Cette méthode reste confrontée aux problèmes de la bonne définition d'une similarité de formes pour le prototypage et de l'application de la segmentation. Elle ne reste efficace que pour des documents imprimés qui contiennent de faibles dégradations n'affectant pas la structure générale des formes des caractères ; elles peuvent alors améliorer les taux de reconnaissance des logiciels d'OCR. Mais, cette méthode ne peut pas s'appliquer aux documents manuscrits. En outre, elle ne permet pas de retrouver les parties irrégulières comme les formes anguleuses (Figure5.10).

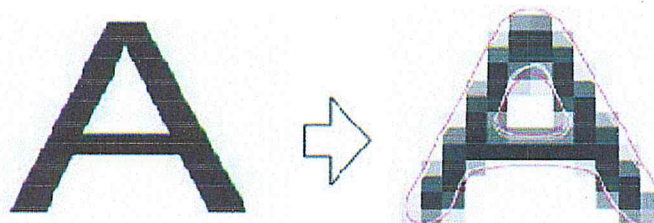


Figure 5.10 : Limite d'une restauration par prototypage

3.7. Restauration géométrique de la courbure [14, 37, 38, 39,40]: Pour la numérisation des livres volumineux reliés apparaissent plusieurs défauts tel que : la présence de la reliure du livre... pour cela un traitement est utilisé pour rendre le document lisible(Figure5.11). Plusieurs méthodes sont alors apparues utilisant une analyse en deux ou trois dimensions de l'orientation des traits ou des lignes de texte qui permettent de construire le relief à partir de l'analyse de la réflexion de la lumière.

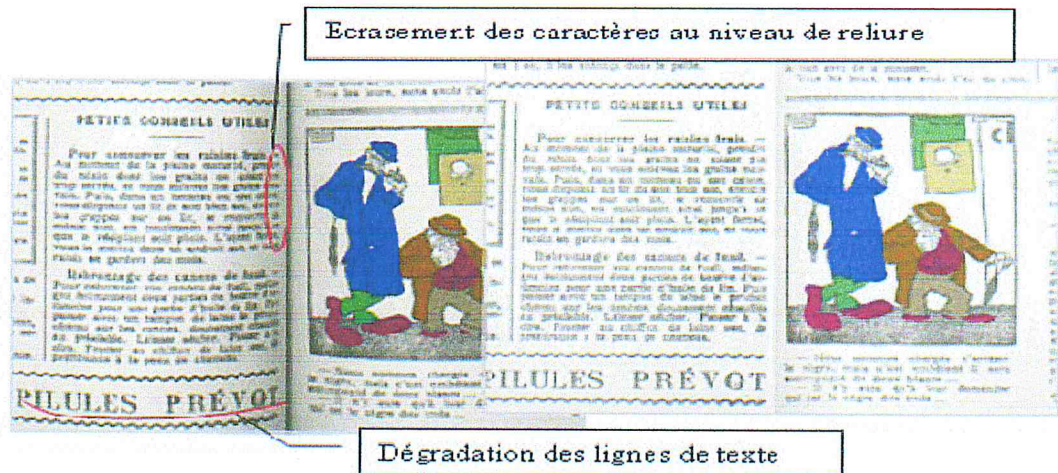


Figure 5.11 : Défauts géométriques de la photographie d'un livre ouvert à gauche et sa version restaurée à droite

3.7.1. Restauration par un modèle 2D : Les méthodes de restauration de la courbure basées sur un modèle 2D procèdent en deux étapes :

1. Calcul de la courbure : Le calcul de la courbure peut être fait à partir des lignes de textes ou par extrapolation des bords de la page.
2. Traitement géométrique pour la corriger.

La modélisation 2D utilise les propriétés d'un document textuel (parallélisme et espacement vertical constant des lignes de texte au sein d'un même paragraphe) afin d'estimer la mise à plat du document. Elle sera limitée pour la segmentation des lignes de textes des documents dégradés ou des manuscrits pour lesquels les alignements ne sont pas naturellement horizontaux. En outre, les dégradations du fond comme la présence de taches ou le passage du verso sur le recto complexifient les opérations de segmentation des lignes de texte.

3.7.2. Restauration par un modèle 3D : Les méthodes de restauration, basées sur un modèle 3D, peuvent être réalisées avec une prise de vue soit

- **normale** (un appareil : les dégradations des feuilles sont généralement modélisées par une surface cylindrique aux abords de la reliure) : Cette modélisation implique que tous les points de la surface ayant la même altitude forment des droites parallèles. La surface est définie par son rayon de courbure. Les lignes de textes sont ensuite corrigées le long de l'axe des abscisses par déplacement horizontal des pixels. Une alternative à cette modélisation est la reconstruction de relief des

documents scannés à partir des niveaux de gris des pixels de l'image. Il s'agit de l'application de la méthode «Shape from shading» qui est une technique bien connue de reconstruction tridimensionnelle à partir des variations de lumière sur une surface de texture homogène. Une fois le relief reconstruit, il suffit d'opérer par la mise à plat sur l'axe des abscisses en fonction de l'altitude du point photographique.

• **Particulière** (deux appareils photographiques ou deux photographies du document prises sous des éclairages différents) : Avec l'utilisation de plusieurs vues sous des angles ou des éclairages différents d'un même document, une mise en correspondance est calculée entre les images pour reconstruire le relief de la page. Ce calcul du relief s'effectue soit par triangulation soit par «Shape from shading». La mise à plat d'un document dégradé par une modélisation 3D offre l'avantage d'utiliser les informations directement liées à la déformation 3D. Néanmoins, sa dépendance vis-à-vis de nombreuses hypothèses (texture, forme a priori de la courbure...) et la complexité du système de capture d'image limitent fortement son usage.

Les méthodes de restauration géométrique de la courbure doivent être suivies d'une restauration des formes. Mais quelque soit la méthode utilisée, l'image issue de cette correction de courbure sera une image interprétée car les caractères situés le long d'une forte courbure ne pourront pas retrouver leurs formes d'origine.

4. Méthodes de restauration non-supervisées :

4.1. Restauration du recto/verso par séparation de sources [14, 41,42]: Les taches présentes sur le texte original du recto n'ont pas la même couleur que celle du verso. Alors la restauration consiste à séparer ces deux couches d'information. A.Tonazzini, L. Bedini, E. Salerno proposent une approche utilisant l'analyse des composantes indépendantes(ACI) et plus particulièrement l'algorithme de Fastl CA pour faire une séparation du recto et du verso. Cette approche est très intéressante à utiliser car elle ne nécessite ni l'information du verso, ni un modèle, ni un apprentissage. Cette méthode résout le problème de transparence pour quelques images de documents dégradées (s'agissant le plus souvent d'images de synthèse). Cependant, les résultats associés à des images réelles de document restent bien modestes (Figure5.12).

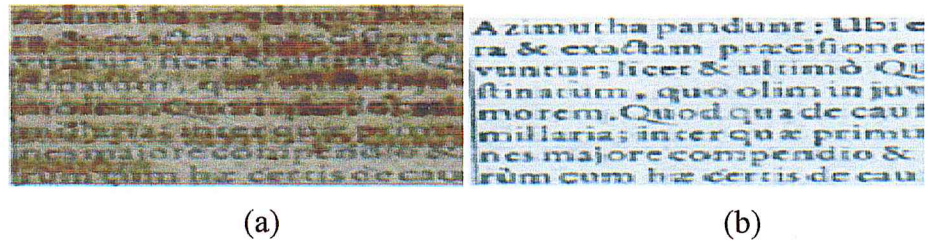


Figure 5.12 : Suppression des taches en transparence par analyse de la couleur

(a) image de synthèse, (b) image restaurée de (a)

4.2. Restauration des formes par morphologie mathématique [14,43] : Ce type de restauration est utilisé pour les petites dégradations qui concernent quelques pixels. Elle permet de combler les trous et réparer les ruptures de traits.

Le principe de cette méthode est d'appliquer sur cette image un ou plusieurs opérateurs morphologiques au moyen d'un masque ou motif géométrique prédéfini (carré, rectangle, cercle...). Le motif géométrique, appelé élément structurant (ES), est déplacé dans l'image de façon à ce que son centre occupe toutes les positions de l'espace (les pixels p). Pour chacune des positions, on applique l'opérateur choisi. Il en résulte une image transformée dans laquelle des éléments X peuvent avoir été éliminés, fusionnés, amincis, etc. (figure 5.13).

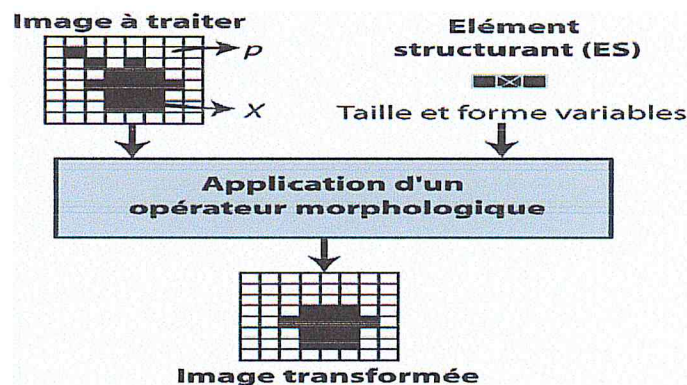


Figure 5.13 : Principe de l'analyse d'image par la morphologie mathématique

La morphologie mathématique ensembliste qui porte sur des images binaires. Elle fait appel à la théorie des ensembles. Pour chaque position du centre de l'élément structurant, on vérifie une relation ensembliste (union, intersection) de l'ES avec les éléments de l'image. Le résultat de la transformation est une nouvelle image binaire composée des points qui satisferont la relation. Pour les images en niveaux de gris, les transformations morphologiques sont généralisées et s'appliquent sur des

fonctions (et non plus des ensembles) en tenant compte cette fois des valeurs prises par les pixels dans le domaine de l'élément structurant. Suivant l'opérateur choisi, on affectera par exemple la valeur maximale des pixels au centre de l'ES (cas d'une dilatation plate) ou la valeur minimale (cas d'une érosion plate).

La restauration par morphologie mathématique n'est applicable que sur un petit voisinage du caractère et ne prend pas en compte des informations structurelles ou globales de la forme. En outre, la morphologie mathématique, par sa définition même, ne présente aucune précision sous-pixellique. La taille de la fenêtre est toujours un nombre entier, ce qui n'est pas toujours le cas lors de la manipulation de traits plus ou moins fins. Enfin, les méthodes de restauration proposées restent très modestes et manquent même d'une évaluation objective des performances.

4.3. Restauration Colorimétrique [14,44]: Toute chaîne de numérisation est exposée à deux types de dégradations que ce soit d'une source externe ou interne. Le support à numériser peut influencer sur la réflexion de la lumière.

- Les contraintes physiques impliquées par la reliure engendrent des dégradations photométriques. Nous observons alors une diminution des intensités lumineuses autour des courbures. E.Trinh propose une méthode de correction des modifications de la couleur et de la luminosité près de la reliure à partir du calcul du profil de l'éclairage. Ce profil utilise l'histogramme de chrominance et de luminosité de toute l'image et des colonnes. Dans le cas d'un changement d'illumination interne ou externe à la source émettrice, il s'avère primordial d'appliquer un prétraitement global identique sur tous les pixels de l'image. Changement de la valeur de luminance, correction gamma et recalage d'histogramme peuvent être des solutions pour résoudre ce problème et améliorer les couleurs de l'image.
- Certaines dégradations photométriques ont tendance à accentuer un ton par rapport à un autre. Pour remédier à ce problème, il suffit de supprimer la tonalité dominante et de mettre en place un équilibrage pour les couleurs. Toutefois, ce traitement doit conserver la balance des couleurs de l'ensemble de l'image pour ne pas en perdre les détails. Une solution simple pour résoudre ce problème consiste à accentuer les contrastes en général. Le contraste est une propriété intrinsèque à une

CHAPITRE 6 :

TEXTURE

INPAINTING

Sommaire :

Parmi les types de dégradations, nous allons nous intéresser à la dégradation de forme d'où nous avons besoin d'étudier dans ce chapitre les différents types de restaurations qui s'intéressent au remplissage de l'information manquante. Nous avons divisé ce chapitre en :

- 1- Introduction
- 2- Problématique
- 3- L'état de l'art
- 4- Conclusion

1. Introduction :

Dans ce chapitre nous allons nous intéresser à la reconstruction des parties manquantes de l'information par texture inpainting et les différentes approches utilisées dans ce domaine.

2. Problématique : Depuis quelques années, on numérise quotidiennement des quantités importantes de documents issus du patrimoine car ils sont les principaux éléments porteurs de l'information depuis la nuit des temps. Cette révolution technologique a ouvert de nouveaux domaines de recherche comme la restauration des images. Nous allons par la suite étudier les différentes technologies existantes pour reproduire l'information textuelle manquante à cause des dégradations.

3. L'état de l'art : Il y'a des siècles, les gens préservaient déjà leurs travail visuel avec soin. Maintenant, grâce à la popularité des ordinateurs, les gens gardent souvent leurs arts, photos, films ...etc dedans. Par la grande puissance des nouveaux ordinateurs, nous pouvons retoucher les images et les redessiner.

3.1. Qu'est ce que l'inpainting ? [17, 19,45] L'inpainting est une technique pour restaurer, compléter la surface d'un objet enlevée. Les techniques produisent une raisonnable bonne qualité des images. Bien qu'il existe des approches antérieures à la suppression des petites zones bien choisies sur les images.

L'inpainting est un processus de reconstruction des parties manquantes détériorées des images et des vidéos, il fait référence aux artistes qui restaurent les tableaux de peintures en repeignant les parties manquantes.

Au monde numérique, l'inpainting fait référence à l'application des algorithmes sophistiqués pour reconstruire les parties manquantes de l'image à partir de l'information de bord de la zone à reconstruire.

3.2. Les méthodes de l'inpainting:

3.2.1. Reconstruction par méthodes variationnelles et équations différentielles partielles (EDP) [14, 45,46] : C'est une méthode de reconstruction des parties manquantes de l'image à partir des informations situées sur les bords de la zone à reconstruire, elle utilise les EDP pour diffuser l'information des bords vers l'intérieur.

L'inpainting consiste à joindre les lignes de même niveau (isophotes) de part et d'autre des bords du domaine à remplir et à poursuivre les contours dans leurs directions initiales (Figure 6.1).

Intuitivement, la meilleure façon de reconstruire la partie manquante consiste à connecter les contours et à conserver des lignes de niveaux les plus courtes et les plus lisses possible.

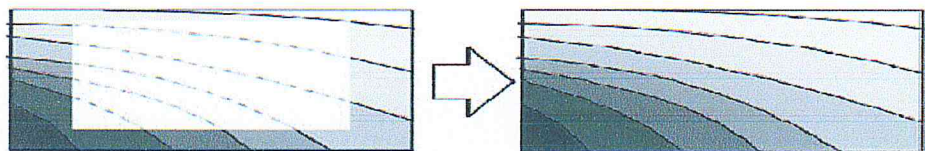


Figure 6.1 : Principe de reconstruction par jonction des isophotes

3.2.1.1. Etapes de l'inpainting : L'inpainting nécessite deux étapes :

1- Repérer l'information manquante et créer un masque qui délimite de façon précise les zones à reconstruire à l'aide d'un masque binaire.

$$M : \Omega \longrightarrow \{0, 1\}$$

$$M(x,y) = \begin{cases} 1 & \text{si le pixel } (x, y) \text{ se situe dans la zone à reconstruire} \\ 0 & \text{sinon.} \end{cases} \quad \text{Equation 6.1}$$

On note ∂M le bord de la zone à reconstruire.

2- Reconstruire la partie manquante de l'image en transportant les valeurs des pixels des bords ∂M à l'intérieur de la zone M à reconstruire.

La diffusion de l'information des bords de la zone à reconstruire à l'intérieur peut se faire par trois approches différentes :

- **Diffusion des intensités de l'image dans la direction des isophotes :** L'inpainting selon Bertalmio se fait par une propagation des informations I des contours du domaine ∂M dans le domaine M suivant la direction N des isophotes (Figure 6.2).



Figure 6.2 : Direction de la propagation des informations situées le long du contour ∂M à l'intérieur du domaine M dans la direction des isophotes N

L'équation $I_t = \nabla I \cdot \vec{N}$ converge et se stabilise lorsque l'information I a été complètement propagée dans la direction N . La meilleure solution consiste à prendre pour N la direction des isophotes. Comme on souhaite un remplissage «lisse», on propage un estimateur de lissage de l'image, par exemple le Laplacien L tel que :

$$L = \Delta I = I_{xx} + I_{yy}$$

L'EDP se stabilisera donc quand le lissage sera constant le long des isophotes. La forme générale de l'équation adaptée à ce processus s'écrit sous la forme suivante (Equation 6.2) :

$$I_t = \begin{cases} \nabla(\Delta I) \cdot \nabla \perp \vec{I} & \text{Si } (x, y) \in M \\ 0 & \text{Sinon} \end{cases} \quad \begin{array}{l} \text{Equation} \\ \mathbf{6.2} \end{array}$$

Au niveau de l'implémentation, dans le modèle de Bertalmio, si on ne fait que diffuser dans le prolongement des isophotes, les lignes de niveaux risquent de se mélanger et de devenir incohérentes. Il faut donc appliquer une diffusion non-linéaire entre chaque itération de l'équation 6.2.

- **Diffusion des intensités suivant le modèle de la totale variation(TV) :** Le modèle de reconstruction par une diffusion des intensités suivant le modèle de la

totale variation (TV) a été introduit par Chan et Shen. Il repose sur la minimisation des variations totales (TV) dans le domaine M à remplir (Equation. 6.3).

$$I_{\text{restauree}} = \text{Min TV}(I) = \int \|\nabla I\| dx dy \quad \text{Equation 6.3}$$

En appliquant le schéma de résolution d'Euler Lagrange, la minimisation de la totale variation donne une EDP équivalente à celle de l'équation 6.4.

$$\begin{cases} I_t = \text{div}\left(\frac{\nabla I}{\|\nabla I\|}\right) & \text{Si } x \in M \\ I_t = 0 & \text{Sinon} \end{cases} \quad \text{Equation 6.4}$$

Cette EDP basée sur la totale variation ne conserve pas la courbure (Figure 6.3.a) et ne peut pas prolonger les isophotes lorsque la zone à reconstruire possède une largeur L supérieure à l'épaisseur d du trait (Figure 6.3.b).

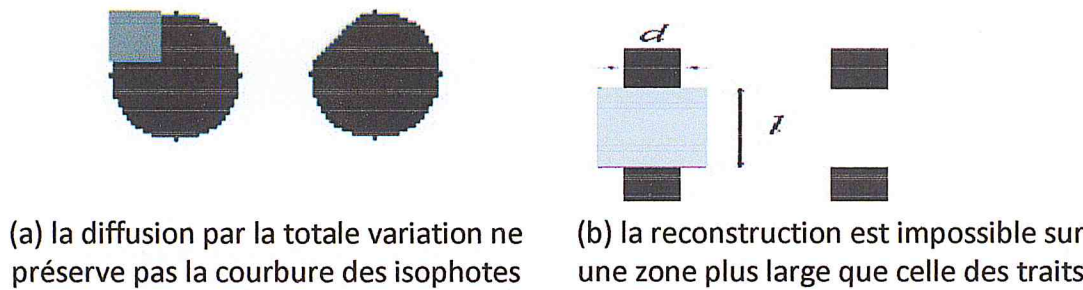


Figure 6.3 : Limites d'une reconstruction par la totale variation

- **Diffusion des intensités dirigée par la courbure (CDD):** Pour remédier aux inconvénients du modèle précédent, cette diffusion propose d'ajouter l'information sur la courbure des isophotes dans l'équation de la diffusion. Cette astuce permet d'encourager la reconstruction de larges zones même si celles-ci ont une dimension supérieure à celle des traits. La CDD (Curvature Driven Diffusion) ou diffusion guidée par la courbure des isophotes (Equation 6.5) prolonge donc les parties de fortes courbures aux extrémités des traits pour renforcer la reconstruction dans des zones larges (Figure 6.5).

$$\begin{cases} I_t = \text{div}\left(K^P \frac{I_t}{\|I_t\|}\right) & \text{Si } (x, y) \in M \\ I_t = 0 & \text{Sinon} \end{cases} \quad \text{Avec} \quad \begin{cases} P \geq 1 \\ k = \text{div}\left(\frac{\nabla I}{\|\nabla I\|}\right) \end{cases} \quad \text{Equation 6.5}$$

Mais la CDD ne préserve toujours pas la courbure car le prolongement des isophotes s'effectue toujours linéairement. C'est avec l'élastique d'Euler que le modèle d'inpainting permet de reconstruire des courbures (Equation 6.6).

$$\begin{cases} It - \text{div} \left((a + bk^2) \frac{\nabla I}{\|\nabla I\|} \right) & \text{Si } x \in M \\ It = 0 & \text{sinon} \end{cases} \quad \begin{array}{l} \text{Equation} \\ \mathbf{6.6} \end{array}$$

Ce nouveau modèle permet aussi de réunir la formulation de la totale variation avec $a = 1$ et $b = 0$ ainsi que la diffusion guidée par la courbure (CDD) avec $a = 0$ et $b = 1$. Ce modèle permet de reconstruire correctement les courbures des formes (Figure 6.4).



(a) Le modèle CDD introduit les courbures en rouge pour reconstruire le trait même si son épaisseur d est inférieure à la largeur l de la zone à reconstruire

(b) Le modèle de l'Elastique d'Euler permet de reconstruire les courbures

Figure 6.4 : Les avantages des modèles de diffusion dirigée par la courbure (CDD) et de l'élastique d'Euler par rapport à la méthode de diffusion basée sur la totale variation (TV)

3.2.1.2. Expérimentation : Nous allons nous intéresser aux images des manuscrits expérimentés par Fadoua Drira dans [14]. La figure 6.5 montre la dégradation avec une grande perte d'informations obtenue par la grille en blanc superposée sur l'image originale.

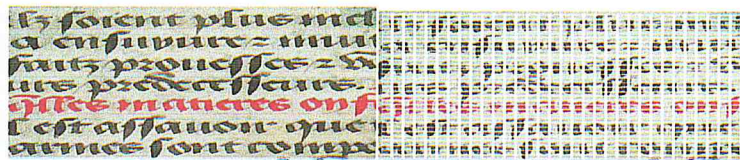


Figure 6.5 : Image originale (IRHT) et superposition du masque en blanc
Le modèle de Chan figure 6.6 et Beltramio figure 6.6 rendent l'image lisible à nouveau.

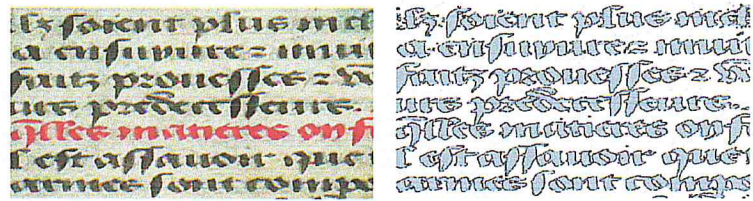


Figure 6.6: Reconstruction par inpainting avec le modèle de Beltramio

De gauche à droite respectivement, l'image reconstruite par Beltramio et l'image reconstruite (en niveau gris) superposée aux contours de l'image originale.

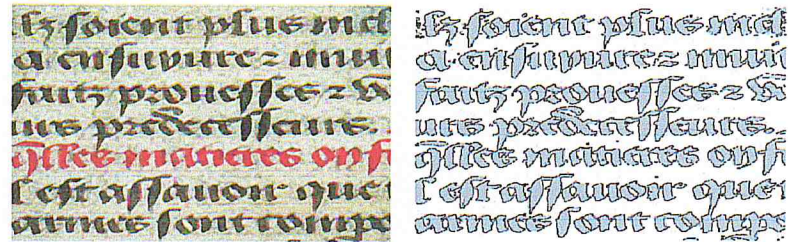


Figure 6.7 : Reconstruction par inpainting avec le modèle de Chan et l'élastique d'Euler

De gauche à droite respectivement, l'image reconstruite par le modèle CDD2 et l'image reconstruite (en niveau gris) superposée aux contours de l'image originale.

3.2.1.3. Discussion sur la méthode de reconstruction par EDP: L'inpainting par EDP est utilisable pour les images des documents car :

- L'inpainting classique ne marche pas dans les images des documents texturées : la texture du papier peut devenir un obstacle pour la reconstruction si elle possède des variations et perturbe le calcul des dérivées.
- Il fonctionne bien sur les images contenant des variations et des contours clairement définis : Il est utilisé pour les images des documents dont leurs traits forment des contours bien marqués pour guider la diffusion.

Néanmoins, cette approche a des limites dans son traitement, les principales sont :

- L'inpainting consiste à reconstruire à travers la continuité des isophotes, mais cette continuité n'existe pas toujours. Alors il faut que la zone à reconstruire ne se situe pas dans les extrémités des lettres ni sur les jonctions entre les traits. (Figure 6.8). Ce n'est pas le cas de l'écriture de la langue arabe car les lettres sont reliées

entre elles, mais il faut aussi que la zone à reconstruire ne se situe pas dans le début ou la fin du mot.



Figure 6.8 : Limite de l'inpainting par manque de continuité dans les isophotes

- La largeur de la zone à reconstruire doit être inférieure à la largeur des traits des caractères : Si le masque est trop large par rapport à l'épaisseur des traits c'est-à-dire que la partie à reconstruire est large par rapport à la lettre, alors il n'y aura pas de réparation (Figure6.9). En fonction de la résolution, l'inpainting est donc limitée à la réparation des rayures ou des craquelures de largeur inférieure à l'épaisseur des traits des caractères.



Figure 6.9 : Limites de l'inpainting pour un masque plus large que l'épaisseur des traits

- L'inpainting de Beltramio et celui de Shan détruisent les parties irrégulières des formes : Pour le modèle de Beltramio, c'est la diffusion anisotropique qui permet de ne pas mélanger les courbes de niveaux dans la zone à repeindre (figure 6.10).

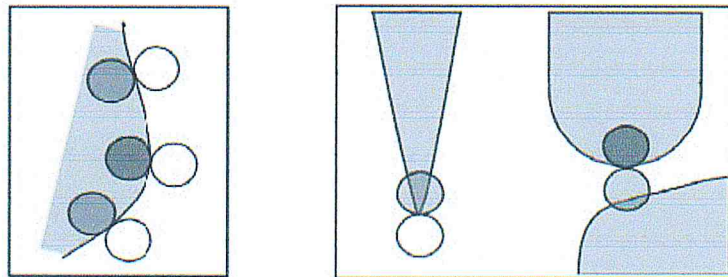


Figure 6.10 : De gauche à droite respectivement : Forme régulière que l'inpainting pourra reconstruire et Formes irrégulières que les modèles actuels ne pourront pas reconstruire

- L'inpainting nécessite la définition d'un masque: Dans le cas de la restauration des images de documents, il faut trouver un traitement qui permette de localiser et segmenter correctement la zone à reconstruire. Si les trous ou les taches peuvent être assez facilement segmentés, les dégradations comme les rayures ou l'effacement partiel de l'encre, sont, en revanche, très difficiles à localiser.

3.2.2. Méthodes basées sur la recherche de similarités (« Exemplar-based ») [19,47] : Les techniques de l'inpainting par exemplar-based sont souvent les plus utilisées pour compléter les zones dégradées. L'algorithme proposé repose sur la typologie de l'inpainting en utilisant une valeur de priorité de chaque patch qui est calculée à l'aide de l'équation 6.7.

$$P(p) = 0,2 * C(p) + 0,8 * D(p) \quad \text{Equation 6.7}$$

Cette priorité utilise le terme data ($D(p)$) qui est basé sur le produit des pixels de la zone à reconstruire et le gradient de l'image original.

La confiance favorise à un pointage de région dans la zone à reconstruire.

D'autre part le terme data donne une haute priorité à la zone à reconstruire qui dispose d'un potentiel de copier l'information structurelle de la région source (image à restaurer hors que la région à reconstruire).

Le terme confiance et data, ensemble permettent de bien choisir le meilleur patch de la région source, on utilise ainsi l'approche de similarité basée sur un nouveau modèle de patch afin de bien sélectionner la forme la plus similaire.

3.2.2.1. Etape de l'inpainting : L'algorithme d'Exemplar-Based se divise en plusieurs étapes, parmi les plus importantes étapes nous pouvons citer :

- 1- Repérer l'information manquante de l'image et on la note par M
- 2- Initialiser la confiance (Equation 6.8).

$$\text{Confidences}[x,y]= \begin{cases} 0 & \text{Si le pixel } (x,y) \in M \\ 1 & \text{Sinon} \end{cases} \quad \text{Equation 6.8}$$

- 3- Initialiser le terme data comme suit : $\text{Data}=0,1$.
- 4- Trouver la zone à reconstruire et trouver les pixels sur le bord de cette dernière.
- 5- Calculer la confiance de chaque pixel sur le bord de la zone à reconstruire.
- 6- Calculer le terme data par le produit du gradient de la région à reconstruire et l'image originale.
- 7- Trouver le patch correspondant avec la priorité maximum. Cette dernière est calculée à l'aide de l'équation 6.7 afin de trouver la forme la plus similaire.
- 8- Remplir les pixels de la zone manquante avec la forme la plus similaire.

3.2.2.2. Discussion sur la méthode de l'inpainting par Exemplar-Based : Ce genre de méthodes est la plus utilisée pour le remplissage des trous et les parties manquantes des documents car :

- Elle est facile à implémenter par ses équations simples.
- Capacité de reconstruction de texture si déjà présente dans l'image.
- Calcul rapide par rapport à l'algorithme précédent.

Par contre cette méthode peut entraîner :

- Une reconstruction difficile si la zone à reconstruire est trop grande (une lettre complète ou plus).
- La propagation des isophotes ne se fait pas si le masque se trouve sur les extrémités de la lettre ou les jonctions.

Par conséquent la figure 6.11 montre la différence entre les deux approches diffusion et exemplar-based:

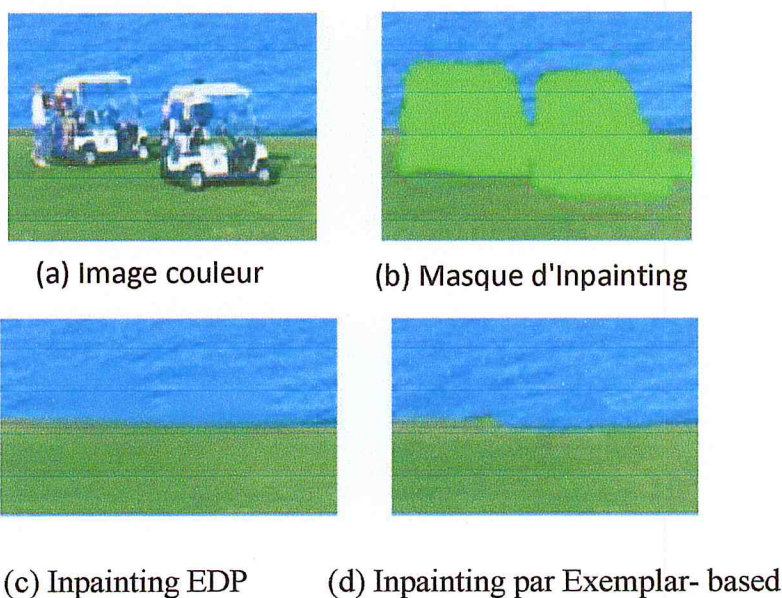


Figure 6.11: Différence entre la restauration par diffusion et par Exemplar-based

L'image restaurée par l'approche EDP a un effet flou par rapport à celle restaurée par Exemplar-based.

3.2.3. Reconstruction par apprentissage : Il s'agit de réaliser un apprentissage des configurations possibles des formes pour prédire les pixels de la zone manquante.

Cette méthode est adaptée aux images de documents qui contiennent des formes continues et répétitives c'est le cas de la langue arabe.

Parmi les méthodes de reconstruction par apprentissage, nous retrouvons :

3.2.3.1. Les champs de Markov (Markov Random Field) [48] : Ils sont utilisés pour réparer les images binaires de caractères.

Cette section fait surtout référence à l'article d'Alexi Efros et de Thomas Leung sur la propagation de texture à partir de textures existantes dans l'image. Cet algorithme propage les textures pixel par pixel et prend comme hypothèse que la modélisation de texture se fait comme un champ de Markov aléatoire. Pour décrire l'algorithme, nous allons en expliquer le principe sur un pixel : Soit I l'image à synthétiser à partir d'un exemple I_{imp} et I_{real} étant l'image réelle de texture infinie soit $p \in I$.

Soit de plus $w(p)$ un patch carré de grosseur w et $d(w_1, w_2)$ la distance entre deux patches. Pour synthétiser les valeurs de p , on construit une approximation de la distribution de probabilité conditionnelle : $P(p|w(p))$

En se basant sur les champs de Markov on assume p indépendant avec $I/w(p)$.

On définit ainsi : $\Omega(p) = \{w' \in I_{\text{real}} : d(w', w(p)) = 0\}$ Contenant toutes les occurrences de $w(p)$.

On cherche alors la meilleure distance pour appliquer une méthode de plus proche voisin : $w^{\text{best}} = \min(w(p), w) \in I_{\text{imp}}$ Une fois celle-ci trouvée pour tous les patches images w' tel que : $\text{distance}(w^{\text{best}}, w') < \varepsilon$ (avec ε étant un seuil) inclus dans Ω' , on peut propager le pixel. On choisit souvent distance comme étant une SSD normalisé (Sum of Square Differences metric) d_{SSD} . Ainsi la valeur du pixel central de Ω' est la valeur donnée à p .

Nous pouvons voir sur des images de structures que cet algorithme marche bien comme le montre la figure 6.12. Cet algorithme ne prend qu'un pixel à la fois et ne s'intéresse qu'à la texture.

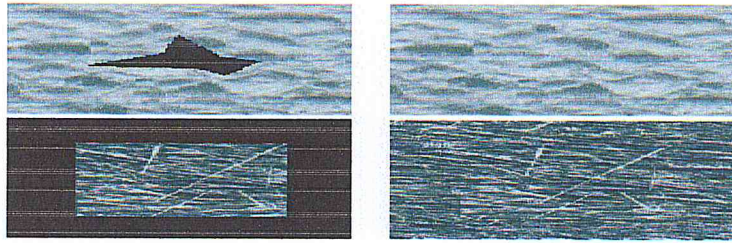


Figure 6.12: Résultat de l'algorithme de synthèse de texture

3.2.3.2. L'approche par les contours actifs guidée par un modèle [34]:

- Zheng et Kanungo proposent d'utiliser un modèle de dégradation (à 6 paramètres, un état de l'art des modèles de dégradations est présenté dans le chapitre 5) pour designer une catégorie de filtres particulièrement adaptés à la restauration d'un ensemble d'images de document acquises dans les mêmes conditions et donc souffrant des mêmes défauts. Pour cela, ils utilisent une image idéale qui leur sert à évaluer les paramètres correspondant à l'image dégradée qu'ils ont à traiter par maximum de vraisemblance; cela permet d'élaborer un filtre morphologique de reconstruction parfaitement adapté. Cette solution fonctionne pour les caractères binaires imprimés.
- Une autre approche est exposée par Baird, dont il élabore un modèle de dégradation utilisé pour entraîner un logiciel de reconnaissance optique de caractères et donc améliorer ses performances (un exemple est présenté sur la Figure 6.13).

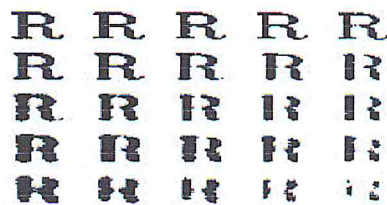


Figure 6.13 : Exemples de dégradations obtenues à partir d'un caractère 'R' écrit en Romain

3.2.3.3. Discussion de la méthode d'inpainting par apprentissage : Cette méthode est adaptée mieux aux documents imprimés car on aura une bonne restauration pour chaque modèle de lettre d'après l'écriture de cette dernière.

Néanmoins la complexité de calcul est très élevée, elle est liée aux nombre de configurations possibles et à la taille de la base d'apprentissage.

4. Conclusion :

Dans ce chapitre nous avons présenté les différentes approches trouvées sur la restauration d'inpainting. L'approche basée sur la recherche de similarités (« Exemplar-based ») nous paraît plus intéressante à mettre en œuvre par sa simplicité à implémenter et son temps de calcul, nous voulons mentionner que l'inpainting par les trois approches ne permet pas la reconstruction d'une grande partie dégradée (un mot ou plus) car cette reconstruction a besoin des informations externes telles que le thème du manuscrit, l'auteur qui a écrit ce dernier ... etc pour remplir les lignes en se basant sur ses informations externes.

Par la suite nous allons essayer de trouver une méthodologie fiable pour l'approche choisie et de l'implémenter.

LA TROISIÈME PARTIE :
METHODOLOGIE ET
IMPLEMENTATION

CHAPITRE 7 :

MÉTHODOLOGIE

Sommaire :

Dans ce chapitre nous présentons la méthodologie utilisée pour résoudre le problème de l'inpainting. Notre chapitre se divise en :

- 1- Introduction
- 2- Méthodologie
- 3- Expérimentation et évaluation
- 4- Discussion sur notre approche
- 5- Conclusion

1. Introduction :

Les documents anciens du patrimoine algériens qui se trouvent dans la bibliothèque nationale et le centre d'Adrar courent un grand danger de disparition à cause des agents de dégradation.

La restauration de ces documents prend une grande ampleur ces dernières années concernant le traitement automatique de ces documents, car elle permet de rendre au document sa richesse qui a disparue au fil du temps, permettant l'exploitation des informations qu'il contient.

2. Méthodologie : Le manque dans un document, ou la craquelure d'une lettre sont causés par plusieurs facteurs, nous pouvons citer comme exemples : les agents biologiques ou physicochimiques mentionnées dans le chapitre 4. Notre travail se nomme « texture inpainting » et consiste à remplir les parties manquantes dans les documents anciens.

2.1. Principe : Nous avons opté pour la méthode d'inpainting par exemplar-based. Elle débute comme le montre la figure 7.1 par une phase de sélection de la partie manquante, une phase de repérage des pixels des bords et enfin une phase d'inpainting par Exemplar-Based pour arriver à une image restaurée à la fin.

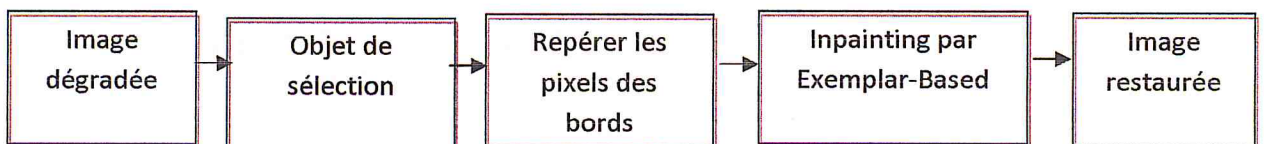


Figure 7.1 : Diagramme d'inpainting

2.1.1. Phase 1 : Sélection de la partie manquante.

La sélection de la partie dégradée se fait de manière automatique ou bien manuelle en cas de la restauration non performante c'est-à-dire quand cette dernière n'est pas restaurée complètement.

Pour la sélection de la partie manquante automatiquement nous avons créé un masque gris ou bleu sur les images des manuscrits algériens, car nous n'avons pas trouvé des documents contenant des dégradations de forme. Si la couleur du pixel est grise ou bleue, nous l'ajoutons à la partie manquante.

2.1.2. Phase 2 : Repérer les pixels des bords.

Pour avoir les pixels des bords de la zone à reconstruire, ceci est fait par la convolution de la zone à reconstruire avec le filtre Laplacien, Sobel, Kerneï selon la méthode précisée dans le chapitre 1. Voici l'algorithme qui montre la convolution entre deux matrices.

```
Entier nbreligneA; /*nombre de ligne de la matrice a*/
Entier nbreColA ; /*nombre de colonne de la matrice a*/
Entier nbreligneB; /*nombre de ligne de la matrice b*/
Entier nbreColB; /*nombre de colonne de la matrice
a*/
Entier conv[][] ;
Fonction Convolution (entrée a [] [], b [] [] : Entier) ;
Sortie : conv [] [] : entier ;
Debut
Pour i allant de 0 à nbreligneB
Debut
    Pour j allant de 0 à nbreColB
    Debut
        Pour k allant de i à i+nbreligneA-1
        Debut
            Pour L allant de j à nbreColA-1
            Debut
                Conv[k][L]=Conv[K][L]+B[i][j]*a[k-
i+1][l-j+1];
            Fin
        Fin
    Fin
Fin
Fin
```

L'algorithme de convolution

2.1.3. Phase 3: Inpainting par Exemplar-Based.

Après avoir trouvé les pixels des bords, débute le travail essentiel:

- Avoir le patch pour chaque pixel de la zone à reconstruire qui se trouve sur le bord de cette dernière. Voici l'algorithme qui retourne le patch autour du pixel donné.

```
Entier x, y, tempo1, tempo2, tempo3, tempo4, N[][] ,ih;
/*ih:longueur d'image*/
avoirPatch (entrée pixelmap[][] ,p:Entier)
/*pixelmap: Matrice des pixels*/
/*p: la position du pixel autour du
patch*/
Sortie : N [] []
/*retourne les coordonnées des pixels du patch qui a la
Longueur 6*/
Debut
    x = (p-1)%ih +1;
    y = (p-1/ih);
    tempo1 = Math.max(x-3, 0);
    tempo2 = Math.min(x+3, ih);
    tempo3 = Math.max (y-3, 0);
    tempo4 = Math.min (y+3, ih);
    Pour i allant de 0 à tempo4-tempo3+1
        Pour j allant de 0 à temp2-temp1+1
            N[i] [j] = temp1+j + (temp3+i)*ih;
        Fin
    Fin
Fin
```

L'algorithme de la création du patch

- Avoir le gradient des pixels des bords de la zone à reconstruire qui se trouvent dans le patch avec la méthode de la différence centrale qui consiste à calculer la moyenne de la différence entre chaque la composante rouge, bleue et vert et de chaque pixel et celui du son voisin dans les directions x et y.


```
Entier : pixelmap [][] ,ih,iw
Pour i allant de 0 à ih-1
    Pour j allant de 0 à iw-1
        ExtraireRGB(pixelmap,i+1,j) ;
        ExtraireRGB(pixelmap,i,j-1) ;
        t1= (r2-r1)/2 ;
        t2= (g2-g1)/2 ;
        t3= (b2-b1)/2 ;
        Gradientx[i] [j]= (t1+t2+t3)/3*255 ;
    Fin
Fin
Pour i allant de 0 à ih-1
    Pour j allant de 0 à iw-1
        ExtraireRGB(pixelmap,i,j+1) ;
        ExtraireRGB(pixelmap,i,j-1) ;
        t1= (r2-r1)/2 ;
        t2= (g2-g1)/2 ;
        t3= (b2-b1)/2 ;
        Gradientx[i] [j]= (t1+t2+t3)/3*255 ;
    Fin
Fin
```

L'algorithme du gradient avec la méthode de la différence centrale

- Calculer le terme data pour les pixels du patch avec l'équation suivante :

$$\text{Data}[i][j] = \text{GradientX}[i][j]*N[i][0] + \text{GradientY}[i][j]*N[i][1]) \quad \text{Equation 7.1}$$

Avec

$N[i][j]$: Le gradient de la région ou se trouve le patch(c'est-à-dire les bords de la zone manquante).

GradientX, GradientY : le gradient de l'image originale.

- Calculer la confiance des pixels du patch selon l'équation suivante :

$$\text{Confidence}[x][y] = \text{confidence}[x][y] + \frac{\text{confidence}[x][y]}{\text{longueur}(\text{patch})} \quad \text{Equation 7.2}$$

- Avoir le patch qui la plus grande priorité.
- Trouver la forme la plus similaire, si l'erreur du patch sélectionné est favorable (c'est-à-dire qu'on a défini une erreur maximum, et si cette dernière est supérieur de l'erreur du patch) alors le patch est trouvée, et l'erreur maximum devient l'erreur du ce patch.

Alors le pixel dégradé sera rempli par la valeur du pixel qui se trouve dans le patch avec la procédure de diffusion des isophotes (c'est-à-dire qu'on suit le pixel qui a une distance plus proche du dégradé et qui se trouve sur la continuité des contours).

3. Expérimentation et évaluation :

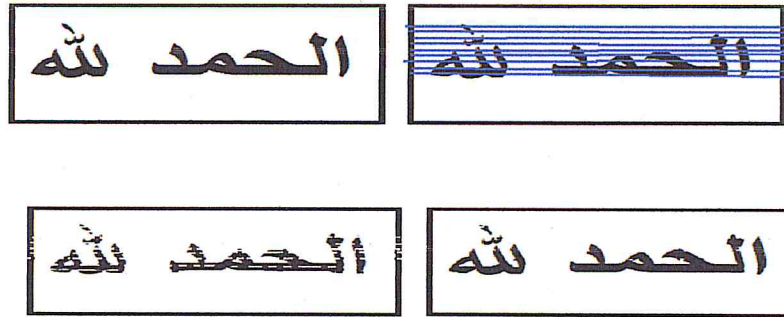
3.1. Expérimentation sur des images de synthèse : L'inpainting a d'abord été appliqué pour effacer les craquelures ou remplir les traits des textes incrustés dans les images ou les vidéos. L'information sur tout le contour du masque est utilisée pour remplir la zone à reconstruire. Dans le cas des images de documents, seule l'information autour des traits sur le contour du masque est utilisable pour la reconstruction. L'application de l'inpainting pour la reconstruction des lettres dans les images de documents donne des résultats beaucoup moins impressionnants que pour les autres applications où l'inpainting a généralement été utilisé.

En effet, dans le cas de photos ou de peintures, les défauts de reconstruction sont difficilement perceptibles. En revanche, pour la reconstruction de formes de caractères, le moindre défaut de reconstruction devient très visible.

Dans cette partie nous allons faire des tests et expérimentations sur différents types d'écritures arabes.

Pour évaluer les performances de la reconstruction, nous avons créé plusieurs images contenant des caractères imprimés de polices très différentes. Nous avons synthétisé des masques artificiels à partir de grilles aléatoires de différentes tailles. Nous avons augmenté graduellement les pertes d'informations en augmentant la finesse de la grille et l'épaisseur des traits jusqu'à ce que la reconstruction devienne difficile. Avec la perte d'information la plus élevée possible, le texte reconstruit par inpainting reste lisible. Cela s'explique par le fait qu'il existe dans l'alphabet arabe de nombreuses redondances d'information et des symétries qui permettent de conserver la lisibilité malgré de nombreuses altérations.

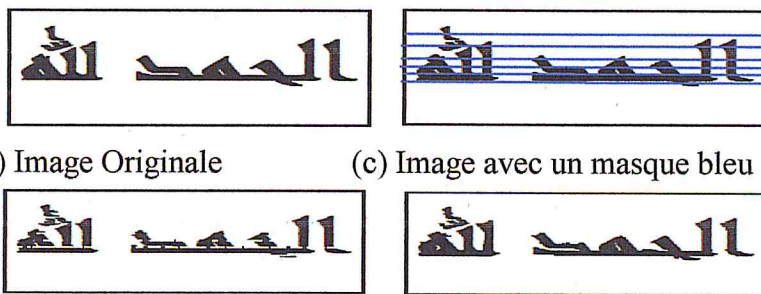
Tout d'abord nous avons appliqué le filtre laplacien pour la convolution. Des différents tests sont réalisés (figure7.2, figure7.3, figure7.4)



(c) Image reconstruire
par un patch de taille 12

(d) Image reconstruire
par un patch de taille 6

Figure 7.2 : Comparaison du modèle d'inpainting par notre approche sur une police Arial



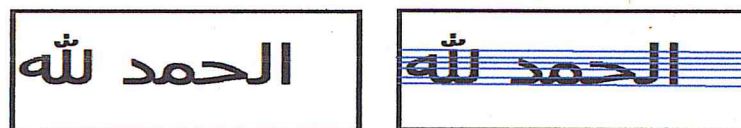
(c) Image Originale

(c) Image avec un masque bleu

(c) Image reconstruire par un
patch de taille 12

(d) Image reconstruire
par un patch de taille 6

Figure 7.3 : Comparaison du modèle d'inpainting par notre approche sur une police
Andalouse



(a) Image originale

(b) Image avec un masque bleu



(c) Image reconstruire
par un patch de taille 12

(d) Image reconstruire
par un patch de taille 6

Figure 7.4 : Comparaison du modèle d'inpainting par notre approche sur une police
Microsoft Sans Serif

La reconstruction par ce modèle avec une taille de patch 12 ne peut pas reconstruire le texte en revanche, si la taille du patch est 6 alors la reconstruction est très favorable, nous pouvons bien distinguer la forme des lettres et le mot dans la police Arial et Microsoft Sans Serif mais un peu moins performante dans la police andalouse. De plus si la taille du patch est grande alors la courbure ne peut pas être reconstruite facilement.

De même, cette méthode d'inpainting est utilisable pour supprimer des formes indiscernables sur les images comme le montre l'exemple 7.5.



(a) Image originale

(b) Image restaurée

Figure 7.5: Suppression avec le modèle « exemplar-based »

3.2. Expérimentation sur des images réelles de documents :

Nous avons appliqué la même expérimentation (avec le filtre laplacien) sur des images en couleurs des manuscrits du patrimoine algérien. Les figures 7.6, 7.7 montrent la grille obtenue en gris superposée à l'image originale et sa reconstruction avec notre modèle respectivement.



Figure 7.6 : Image originale et superposition du masque en gris



(a) Image reconstruite par un patch de taille 12

(b) Image reconstruite par un patch de taille 6

Figure 7.7 : Reconstruction avec notre approche

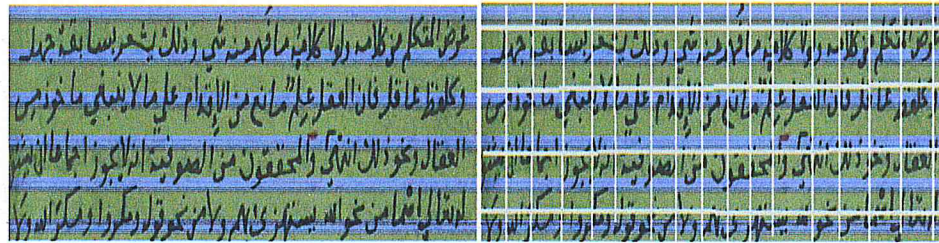
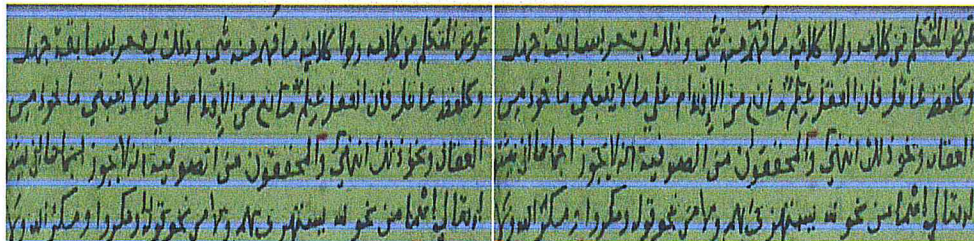


Figure 7.8 : Image originale2 et superposition du masque en gris



(a) Image reconstruire par un patch de taille 12

(b) Image reconstruire par un patch de taille 6

Figure 7.9 : Reconstruction avec notre approche de l'exemple 2

Malgré les pertes d'information, la reconstruction par ce modèle avec la taille du patch 6 est mieux que celle avec la taille 12. Mais sans oublier de mentionner qu'avec ce modèle la reconstruction des courbures est parfois difficile.

- Nous avons fait d'autres expériences avec les filtres Sobel et Kernel des images précédentes (figure 7.2, figure 7.8) avec la taille 6 du patch, et nous avons obtenus les résultats suivants :



(a)

(b)

(a) La reconstruction avec le filtre Sobel pour la détection des contours

(b) La reconstruction avec le filtre Kernel pour la détection des contours

Figure 7.10 : Reconstruction avec notre approche avec les filtres Sobel et Kernel de l'exemple de la figure 7.2



(a) : La reconstruction avec le filtre Sobel pour la détection des contours
 (b) : La reconstruction avec le filtre Kernel pour la détection des contours

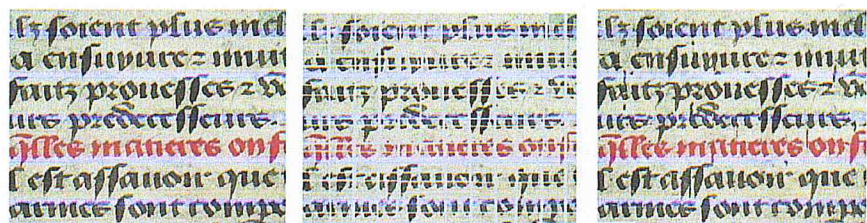
Figure 7.11 : Reconstruction avec notre approche avec les filtres Sobel et Kernel de l'exemple de la figure 7.8

Nous constatons que la performance de la reconstruction avec le filtre Sobel est presque la même qu'avec le filtre Laplacian par contre avec le filtre Kernel c'est un peu moins performant.

4. Discussion sur notre approche :

L'inpainting par « exemplar-based » est applicable aux images de documents pour deux raisons :

- L'approche utilisée est une méthode simple et universelle qui ne dépend pas du contenu du manuscrit. En effet quelque soit le manuscrit et sa couleur, et quelle que soit la langue utilisée nous pouvons le restaurer comme le montre la figure 7.10.
- Le modèle de diffusion, et plus spécialement le modèle de beltramio comme nous avons mentionné dans le chapitre 6, ne peut pas reconstruire si la taille de la partie dégradée est supérieure à la l'épaisseur des traits. Par contre cette approche est utilisée pour ce type de problème.



(a) Image originale (b) Superposition du masque gris sur l'image (c) Image restaurée

Figure 7.12 : Exemple de reconstruction de manuscrit écrit avec le latin

Cependant, cette approche présente des limites dont les principales sont les suivantes :

- Si la partie dégradée possède beaucoup de courbures, la reconstruction de cette dernière est difficile.
- Le temps d'exécution de cette approche est un peu long, car nous avons créé des manuscrits avec la superposition du masque gris ou bleu (en absence de documents réels dégradés pour l'inpainting) de ce fait beaucoup de dégradations sont apparues. Par contre si le document est dégradé, les parties dégradées ne seront pas beaucoup.

Nous pouvons remédier à ce problème l'utilisation d'un ordinateur puissant, malgré que notre objectif c'est la reconstruction ce n'est pas le temps d'exécution.

5. Conclusion :

Dans notre travail nous traitons le problème de l'inpainting envers les trous et les craquelures des lettres dans les documents anciens.

Dans notre solution, nous avons combiné deux méthodes: la première « exemplar based » pour trouver la forme la plus similaire à la partie dégradée, et la deuxième c'est la continuité des contours pour le remplissage des pixels qui se trouvent dans les parties manquantes.

CHAPITRE 8 :

IMPLÉMENTATION

Cependant, cette approche présente des limites dont les principales sont les suivantes :

- Si la partie dégradée possède beaucoup de courbures, la reconstruction de cette dernière est difficile.
- Le temps d'exécution de cette approche est un peu long, car nous avons créé des manuscrits avec la superposition du masque gris ou bleu (en absence de documents réels dégradés pour l'inpainting) de ce fait beaucoup de dégradations sont apparues. Par contre si le document est dégradé, les parties dégradées ne seront pas beaucoup.

Nous pouvons remédier à ce problème l'utilisation d'un ordinateur puissant, malgré que notre objectif c'est la reconstruction ce n'est pas le temps d'exécution.

4. Conclusion :

Dans notre travail nous traitons le problème de l'inpainting envers les trous et les craquelures des lettres dans les documents anciens.

Dans notre solution, nous avons combiné deux méthodes: la première « exemplar based » pour trouver la forme la plus similaire à la partie dégradée, et la deuxième c'est la continuité des contours pour le remplissage des pixels qui se trouvent dans les parties manquantes.

Sommaire :

Le chapitre d'implémentation contient la description de la partie implémentation et application. Il est divisé en :

- 1- Introduction
- 2- Outil de travail
- 3- L'application
- 4- Conclusion

1. Introduction:

Après le chapitre de la méthodologie, nous entamons le chapitre implémentation qui décrit le choix de l'environnement et l'outil utilisé pour développer notre application.

2. Outils de travail: Pour implémenter notre système, nous avons utilisé les outils suivants :

2.1. Choix du langage: Nous avons opté pour le langage java pour l'implémentation car il est le plus utilisé ces dernières années.

2.1.1. Aperçu [49] : Java est un langage de programmation orienté objet créé en 1982, présenté officiellement le 23 Mai 1995 au SunWorld.

Il peut être porté et utilisé sur plusieurs systèmes d'exploitation comme l'Unix, Windows, Mac OS ... etc.



Ce type de langage permet de développer des applications client-serveur, il s'est imposé au monde de l'entreprise grâce aux servelet, JSP (JavaServer Pages).

2.1.2. Avantages de java [49] :

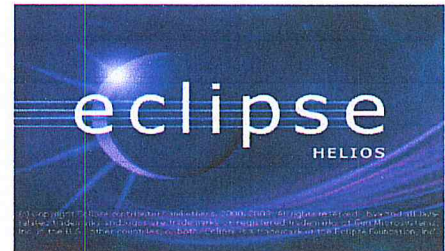
- La mémoire en java est allouée et libérée automatiquement, c'est pas comme les autres langages, il faut qu'on fasse l'allocation manuellement. Nous n'avons pas besoin de se préoccuper des pertes de mémoire.
- Le langage java n'utilise pas l'arithmétique des pointeurs, ils utilisent une bonne gestion de tableau.

- Il utilise la notion d'interface pour remplacer l'héritage multiple. Les interfaces offrent tout ce que nous pouvons obtenir à partir de l'héritage multiple.

2.2. Choix de l'environnement de développement : Nous avons choisis Eclipse 3.6 pour le développement de notre système, car il est simple à utiliser et que nous avons travaillé avant avec cet environnement avant.

2.2.1. Aperçu [50] : Eclipse est un environnement initié par IBM en utilisant le langage java.

Dés l'origine du projet IBM a voulu offrir une solution multiplate-forme pouvant être exécuté sur différents système d'exploitation.



2.2.2. Objectif d'Eclipse [50] : Son objectif est de produire et fournir divers outils gravitant autour de la réalisation de logiciel, englobant les activités de codage logiciel proprement dites mais aussi de modélisation, de conception, de test ... etc.

Son environnement de développement vise à la généricité pour lui permettre de supporter n'importe quel langage de programmation.

3. L'application : L'application est conçue de sorte que la partie dégradée peut être sélectionnée automatiquement ou manuellement. Et de plus chaque changement d'un pixel de l'image peut être suivi par l'œil pour voir le déroulement du processus depuis l'image originale jusqu'au résultat final.

La figure 8.1 nous montre un aperçu sur la fenêtre principale.



Figure8.1 : Aperçu de l'application

Nous avons divisé l'application en 4 menus :

1- Le premier menu c'est Fichier : Comme la montre la figure 8.2 comporte les commandes suivantes :

- Ouvrir Image : Pour le chargement de l'image à restaurer.
- Enregistrer : Pour l'enregistrement de l'image après restauration
- Enregistrer Sous : Pour l'enregistrement de l'image dans un nouveau espace mémoire.
- Sortir : Pour sortir de l'image.



Figure8.2 : Le 1^{er} menu de l'application

2- Le deuxième menu est Edition : Il contient les commandes suivantes :

- Retour Arrière : Pour le retour arrière de la dernière action.
- Répéter la dernière action : Pour répéter la dernière action faite.

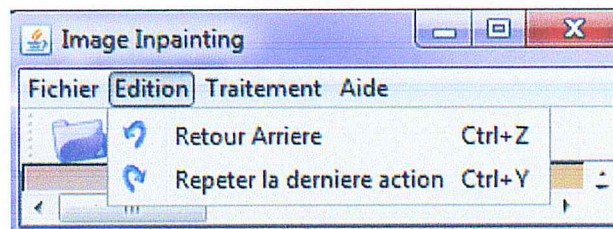


Figure8.3 : Le 2eme menu de l'application

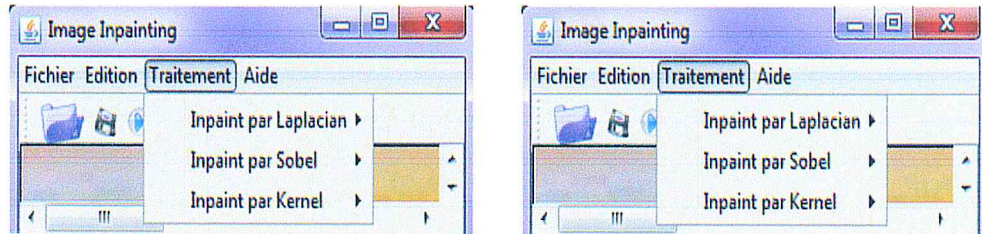
3-Le troisième menu est Traitement : Il contient les commandes d'inpainting par :

- Le filtre Laplacian
- Le filtre Sobel
- Le filtre Kernel

Et tous les trois contiennent les fonctions suivantes :

- Run : Pour l'inpainting par taille du patch au début c'est toute l'image.

➤ Fast Inpainting : Pour la taille du patch au début est inférieure à la taille de l'image c'est-à-dire une partie très petite autour du pixel dégradé.



(a)

(b)

(a) Le contenu du menu traitement

(b) Le contenu de chaque sous menu du menu traitement

Figure8.4 : Le 3^{eme} menu de l'application

4-Le quatrième menu c'est Aide : Pour avoir un petit résumé sur l'application

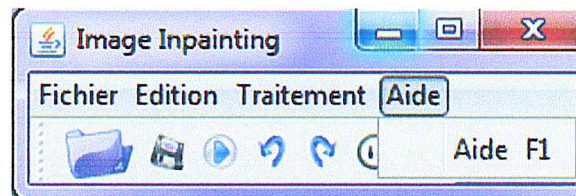


Figure8.5 : Le 3^{eme} menu de l'application

Sans oublier de mentionner que nous avons fait une barre d'outils qui contient des boutons des actions les plus utilisé dans l'application.

4. Conclusion :

Dans ce chapitre nous avons fait un petit résumé sur la partie implémentation pour définir les différentes outils et fonctionnalités de notre application.

CONCLUSION GENERALE

La restauration numérique des documents dégradés est un art qui remonte à plusieurs décennies, et qui s'est propagé ces dernières années, mais ces méthodes restent limitées à des types de dégradations bien définies.

Notre travail s'intitule la restauration des documents anciens : texture inpainting, qui consiste à remplir le vide dans un manuscrit et plus précisément de remplir le vide qui s'apporte à une lettre comme les craquelures de cette dernière. Car quelque soit l'outil de restauration ou l'intelligence de l'être humain, ils ne peuvent pas remplir ce type de dégradations (plus d'un mot dégradé) sauf si nous connaissons le thème autour du quel le manuscrit est écrit.

L'idée consiste à utiliser une méthode qui permet de trouver le patch qui convient le mieux au pixel dégradé et cela autour des bords de ce dernier, et par la suite trouver la forme la plus similaire à l'aide de la continuation des contours. Nous avons fait plusieurs expériences sur des images de synthèse et des manuscrits issus du patrimoine algérien. Malgré le temps d'exécution qui était un peu long, les résultats étaient convainquant dans la majorité de ces expériences.

Des améliorations de notre travail peuvent se faire avec l'utilisation quelques techniques de traitement d'images, et de reconnaissance des formes qui ont pour objectifs d'améliorer la lisibilité de ces manuscrits.

Comme perspectives à notre travail, nous avons identifié quelques pistes possibles :

- Utiliser une base de connaissances des différentes lettres de la langue arabe dans des différents modèles et comparer les caractères du manuscrit restauré a chaque lettre de la base et puis remplacé ces derniers par la lettre qui convient le mieux.
- Combiner les deux méthodes de reconstruction: inpainting par formes similaires plus celle de diffusion afin de bien améliorer la lisibilité des manuscrits.

BIBLIOGRAPHIE

- [1]Eyrolles 1987, Introduction aux techniques de traitement d'images.
- [2]Anne R. Kenney, Bibliothécaire Assistante de l'Université Oya Y. Rieger, Coordinateur de l'Enseignement Dispensé, De la théorie à la pratique didacticiel d'imagerie numérique.
- [3]MICROSOFT, « Encyclopédie Encarta », 2005.
- [4]Jean-François Pillou, Juillet 2007, Imagerie numérique.
- [5]R.C. GONZALES, P. WINTZ, Digital Image Processing, Eddition Wessley, 1997.
- [6]SAMAMBA Tony, Reconnaissance des formes comme outil d'aide aux traitements d'image. Cas des empreintes digitales, mémoire ISP/Bukavu, 2005-2006.
- [7]A. d'HARDANCOURT, « Fou du multimédia » Sybex 1995.
- [8]KAMPEMPE Kilimali, séparation de l'image et du fond, TFC/ISP Bukavu, 2005-2006.
- [9]JDN Developpeur, Les formats d'images bitmap et vectorielles, disponible sur le web à l'adresse : www.journaldunet.com/developpeur/tutoriel/gra/050625-panorama-Format-fichier-graphique.shtm, 2005, consulté le 13/12/2011.
- [10]Compression d'image, disponible sur le web à l'adresse : Laurent, www.esiee.fr/~info/a2si/rapport1999-2003/node161.html - 14k - URL valide le 02/03/2008, consulté le 13/12/2011.
- [11]Jean-marie, MUNGUAKONKWA BIRINGANINE La liaison automatique des plusieurs images perçues sur un scanner, SP(Institut Supérieur Pédagogique de Bukavu) - licencié en pédagogie; Option : Informatique de Gestion 2008.
- [12]J.P Cocquerez et philip,Analyse d'image: filtre et segmentation édition Masson paris,1995.
- [13]Yaovi GAGOU, 2007-2008 cours de Licence de Physique S6 à l'Universite de Picardie Jules Verne Année Académique.

[14]Fadoua Drira, thèse de contribution a la restauration des images de documents anciens, École Doctorale Informatique et Information pour la Société (EDIIS).

[15]Numérisation, disponible sur le web à l'adresse: <http://fr.wikipedia.org/wiki/Num%C3%A9risation>, consulté le 15/12/2011.

[16]Environnement et conservation des documents, disponible sur le web à l'adresse: http://webworld.unesco.org/safeguarding/fr/txt_envi.htm#01volvandalisme.

[17]Inpainting, disponible sur le web à l'adresse: <http://en.wikipedia.org/wiki/Inpainting>, consulté le 8/1/2012.

[18]Banza, Helmut. Conservation and restoration workshops working for libraries. IFLA journal. vol. 1, no 3, 1975.

[19]Jason C. Hung¹, JOURNAL OF SOFTWARE, VOL. 3, NO. 8, NOVEMBER 2008Exemplar-based Image Inpainting base on Structure Construction.

Department of Information Technology, The Overseas Chinese Institute of Technology, Taichung, Taiwan.

[20]Thierry Dambermont, Une formation technique et plastique à la Numérisation/Scanning et traitement chromatique RVB, Première année du graduat en Techniques Infographiques de la Haute Ecole de Namur «Albert Jacquard» Document de formation 2003-2004.

[21]Scanner (informatique), disponible sur le web à l'adresse: [http://fr.wikipedia.org/wiki/Scanner_\(informatique\)](http://fr.wikipedia.org/wiki/Scanner_(informatique)), consulté le 13/1/2012.

[22]Jean-Michel Mermet, Cours licence BDAN – IUT2 Grenoble –Techniques de Numérisation, 2009-2010.

[23]Manuscrit, disponible sur le web à l'adresse: <http://fr.wikipedia.org/wiki/Manuscrit>, Consulté le 14/1/2012.

[24]Codex, disponible sur le web à l'adresse: <http://fr.wikipedia.org/wiki/Codex>, consulté le 14/1/2012.

[25]Livret du stage d'initiation au manuscrit médiéval, réalisé par l'Institut de Recherche en Histoire et du Texte (IRHT), Lyon – France. Département des manuscrits arabes - disponible en ligne sur

<http://aedilis.irht.cnrs.fr/stage/descriptionmanuscrit.html> édité le 20/09/2006 par l'IRHT, consulté le 14/1/2012.

[26]ADRAR : Les manuscrits contre l'oubli article écrit par Bentouba Saïd source www.quotidien-oran.com disponible sur internet à l'adresse : http://www.vitamedz.com/adrar-les-manuscrits-contre-l-oubli/Articles_15641_2599_1_1.html, consulté le 15/1/2012.

[27] HAMMOUM Zoubida, les services en lignes liés aux activités culturelles, disponible sur le web à l'adresse: <http://www.m-culture.gov.dz>, consulté le 15/1/2012.

[28] G. Sharma. Cancellation of show-through in duplex scanning, Appears in Proceedings IEEE International Conference on Image processing, vol. 2, 2000.

[29] E. Dubois, A. Pathak. Reduction of bleed-through in scanned manuscripts documents, Dans Proceeding IS&T Image processing, Image Quality, Image capture Systems, pp. 177-180, Montreal, Canada, 2001.

[30] Q. Wang, T. Xia, C. L. Tan, L. Li. Directional Wavelet Approach to Remove Document Image Interference, ICDAR 2003, pp. 736-740, Edinburgh, 2003.

[30] C. Wolf. Document Ink bleed-through removal with two hidden Markov random fields and a single observation field, Technical Report RRLIRIS-2006-019. Lyon : LIRIS, INSA Lyon, 2006.

[31] 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.

[32]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.

[33]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.

[34]H. S. Baird. Document Image Defect Models, dans Structured Document Image Analysis. New York : Springer, pp. 546-556, 1992.

[35]B. ALLIER. Contribution à la numérisation des collections : apports des contours actifs, Thèse de doctorat, Institut National des Sciences Appliquées de Lyon, 2003.

[36]John D. Hobby Tin Kam Ho Bell Labs, Lucent Technologies Bell Labs, Lucent Technologies Murray Hill, New Jersey 07974 Murray Hill, New Jersey 07974, Enhancing Degraded Document Images via Bitmap Clustering and Averaging.

[37]M. Bergounioux, cours de master2 2008,2009Quelques méthodes mathématiques pour le traitement d'image, disponible sur le web à l'adresse :

<http://cel.archives-ouvertes.fr/docs/00/34/97/75/PDF/Coursimage09.pdf>, consulté le 20/1/2012.

[38]Jérôme Massot, thèse Implémentation de méthodes de restauration équilibrée 3D, Institut National Polytechnique de Lorraine.

[39]Carole Fritz, Gilles Tosello, Marc Azéma, Olivier Moreau, Guy Perazio, Restauration virtuelle de l'art pariétal paléolithique : le cas de la grotte de

Marsoulas, disponible sur le web à l'adresse: <http://www.insitu.culture.fr/pdf/fritzrestau-894.pdf>, consulté le 2/2/2012.

[40] H. Cao, X. Ding, C. Liu. Rectifying the Bound Document Image Captured by the Camera: A Model Based Approach, ICDAR , pp. 71-75, Edinburgh, Scotland, 2003.

[41]Gradient, disponible sur le web à l'adresse <http://fr.wikipedia.org/wiki/Gradient>.

[42] Anna Tonazzini, Luigi Bedini, Emanuele Salerno, Independent component analysis for document restoration.

[43] David Sheeren, Sébastien Lefèvre, Jonathan Weber, La morphologie mathématique binaire pour l'extraction automatique des bâtiments dans les images THRS, disponible sur le web à l'adresse : <http://dpt-info.u-strasbg.fr/~lefevre/publis/rig2007.pdf>, consulté le 25/2/2012.

[44] E. Trinh, De la numérisation à la consultation des documents anciens, Thèse de doctorat dans le cadre d'un projet européen Dépora et d'un projet industriel, Book Restorer, INSA de Lyon, France, 2003.

[45] T.Chan, J.Shen, Mathematical models for local non texture inpaintings, SIAM J.Appl. Math, 2002.

[46] C. Ballester, M. Bertalmio, V. Caselles, J. Verdera. Filling-in by joint interpolation of vector fields and gray levels. IMA Preprint Series, April 2000, Article 1706.

[47] A. Criminisi, P. Perez and K. Toyama, Region Filling and Object Removal by Exemplar-Based Image Inpainting, IEEE Trans. on Image Processing, vol. 13, Sept. 2004.

[48] Alexi Efros et Thomas Leung, Texture Synthesis by Non-Parametric Sampling.

[49] Java (langage), disponible sur le web à l'adresse : http://fr.wikipedia.org/wiki/Java_%28langage%29#Aper.C3.A7u, consulté le 10/06/2012.

[50] Eclipse (logiciel), disponible sur le web à l'adresse : http://fr.wikipedia.org/wiki/Eclipse_%28logiciel%29#Architecture, consulté le 10/06/2012.

[51] Bibliothèque nationale d'Algérie, disponible sur le web à l'adresse: <http://www.biblionat.dz/>