

République Algérienne Démocratique et Populaire
Université Saad Dahleb Blida
Faculté des Sciences Département Informatique

Projet de Fin d'Etude Mastère 2 Ingénierie de Logiciel

L'utilisation des Ondelettes dans la Restauration des Anciens Manuscrits

Réalisé par :

- Drazzi Samir
- Djouimaï Sofiane

Promoteur :

- Mr Hadj Yahia Ouahid

Président Jury: N. Boustia

Examinateur: Mr Nahal

Mme Torbaline



MA-004-231-1

2013 / 2014



>> Remerciements <<

Nous remercions « Dieu » de nous avoir accordé les connaissances de la science et de nous avoir aidés à réaliser ce travail.

Au terme de ce modeste travail nous tenons à remercier chaleureusement et respectivement tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste projet de fin d'étude.

Nos vifs remerciements vont tous d'abord à Mr. O. Hadj-Yahia qui nous a aidés à développer notre travail.

Et nous tenons à remercier tous les Enseignants qui nous ont suivis durant notre formation.

Et Merci à Tous...

Dédicaces

*A chers Parents,
Que nulle dédicace ne puisse exprimer ce que je leurs dois,
pour leur bienveillance, leur affection et leur soutien...
Trésors de bonté, de générosité et de tendresse, en
témoignage de mon profond amour et ma grande
reconnaissance.*

*A mon cher frère et mes chères sœurs,
En témoignage de mes sincères reconnaissances pour les
efforts qu'ils ont consentis pour l'accomplissement
de mes études.
Je leur dédie ce modeste travail en témoignage de mon
grand amour et ma gratitude infinie.*

A toute ma Famille

*A tous mes Amis,
Pour leur aide et leur soutien moral durant l'élaboration
du travail de fin d'étude.*

Et A tout le Monde...

Samir

Table des Matières

Remerciement

Dédicace

Résumé

Chapitre I - Introduction aux images numériques

1-	Introduction.....	1
2-	Image et image numérique.....	1
3-	Caractéristique de l'image numérique.....	1
3-1	Pixel.....	1
3-2	La Définition, la Résolution et le Dimension.....	2
3-3	Histogramme.....	3
3-4	La luminance.....	4
3-5	Contraste.....	5
3-6	Niveau de gris.....	5
3-7	Région.....	6
3-8	Contour.....	6
3-9	Bruit (parasite).....	6
3-10	Les représentations de la couleur.....	6
a-	Le codage RVB/RGB.....	6
b-	Le codage CMY.....	6
c-	Le codage Palette (Indexed images).....	7
d-	Le codage TSL/HSL.....	7
e-	Le codage CIE.....	8
f-	Le codage YUV.....	8
g-	Le codage YIQ.....	9
4-	Les types d'images numériques.....	9
4-1-	Images vectorielles.....	10
4-2-	Images Bitmap/Raster.....	10
4-3-	Les formats méta fichiers.....	11
5-	Le traitement d'images.....	11

5-1-	Introduction.....	11
5-2-	Type de traitement d'image.....	12
6-	Conclusion.....	13

Chapitre II - Les manuscrits et leurs dégradations

1-	Introduction.....	14
2-	Définition du manuscrit.....	14
3-	L'évolution du livre.....	14
3-1-	Du volumen au Codex.....	14
3-2-	Du parchemin au papier.....	15
	Les livres manuscrits aujourd'hui ?.....	16
4-	Les supports de l'écriture.....	16
4-1-	La pierre.....	17
4-2-	Ecorces d'arbres, tablettes d'argile.....	17
4-3-	Tablettes de cire.....	17
4-4-	Papyrus.....	17
4-5-	Parchemin.....	18
4-6-	Papier.....	19
5-	Dégradation des manuscrits.....	20
5-1-	Définition.....	20
5-2-	Les Causes.....	20
5-2-1.	Dégradations biologiques.....	20
5-2-2.	Dégradations physico-chimiques.....	21
5-2-3.	Dégradation physique.....	21
5-2-4.	Dégradations chimiques.....	22
5-2-5.	La pollution atmosphérique.....	23
5-2-6.	Dégradations dues à la capture.....	23
5-2-7.	La mauvaise qualité des matériaux constitutifs des documents.....	23
5-2-8.	Les sinistres naturels ou accidentels.....	24
5-2-9.	Les méfaits dus à l'homme.....	24
6-	La Numérisation.....	24

6-1-	Qu'est-ce que la numérisation ?.....	25
6-2-	Pourquoi numériser ?.....	26
6-3-	Les objectifs de la numérisation.....	26
6-4-	Type de scanners utilisés pour la numérisation.....	27
6-4-1.	Scanners à Plat.....	27
6-4-2.	Scanners à défilement.....	27
6-4-3.	Les Scanners à Tambour.....	28
6-4-4.	Les Scanners à Microfilms.....	29
6-4-5.	Scanners à Diapos.....	29
6-4-6.	Appareils Photo Numériques.....	30
6-5-	Les avantages de la numérisation.....	30
6-6-	Les points faibles de la numérisation.....	32
6-7-	Recommandations pour la numérisation de différents supports.....	32
6-8-	Les conséquences de la numérisation.....	34
7-	Conseils pour la consultation des manuscrits.....	35
8-	Conclusion.....	36

Chapitre III - Aperçu sur la transformée en ondelettes

1-	Introduction.....	37
2-	La Transformée de Fourier.....	38
2-1-	Inconvénient de L'analyse de Fourier.....	38
2-2-	Transformée de Fourier à Fenêtre Glissante.....	38
3-	L'Analyse en Ondelette Face à l'Analyse de Fourier.....	40
4-	Les Aspects Théoriques de la Transformée en Ondelette.....	41
4-1	Ondelette Mère.....	41
4-2	La Transformée en Ondelette Continue.....	42
4-2.1	Définition.....	42
4-2.2	Etapas de la Transformée en Ondelette Continue..	42
4-3	La Transformée en Ondelette Discrète.....	43
5-	L'Analyse Multirésolution.....	44

6-	Algorithme de décomposition/reconstruction de S.MALLAT	44
6-1-	Décomposition	44
6-2-	Reconstruction	47
7-	Transformée en Ondelette à deux Dimensions	48
8-	Les Familles d'Ondelettes	51
8-1-	L'ondelette de Haar	51
8-2-	L'ondelette de Daubechies	54
8-3-	L'ondelette Biorthogonale	56
9-	Conclusion	57

Chapitre IV - Démarche de restauration

1-	Introduction	58
2-	Type de restauration	59
2-1-	La restauration physique	59
2-2-	La restauration virtuelle	59
2-3-	La restauration numérique	59
3-	Les méthodes de restauration	60
3-1-	Méthodes de restauration supervisée « Non-Aveugle »	60
3-1-1-	Séparation recto/verso par recalage	60
3-1-2-	Séparation recto/verso par analyse de l'orientation	61
3-1-3-	Séparation recto/verso par analyse de la couleur	62
3-1-4-	Méthode de séparation recto/verso proposée par Wang	63
3-1-5-	Restauration des formes par modèle	69
3-1-6-	Restauration des formes par prototypage	69
3-1-7-	Restauration géométrique de la courbure	70
a.	Restauration par un modèle 2D	70
b.	Restauration par un modèle 3D	70
3-2-	Méthodes de restauration non-supervisées « Aveugle »	71
3-2-1-	Restauration du recto/verso par séparation de sources	71
3-2-2-	Restauration Colorimétrique	72
4-	Conclusion	74

Chapitre V - Application (Méthodes et Outils)

1- Introduction.....	75
2- Nos Méthodes.....	75
3-1 Séparation recto/verso.....	77
3-2 Détection de contours.....	82
3-2.1 Méthode d'élimination de l'image d'approximation (BB).....	82
3-2.2 Méthode de fusion des coefficients (détails).....	84
3-3 Débruitage.....	85
3-3.1 Débruitage des coefficients (détails).....	85
3-3.2 Elimination des coefficients (détails).....	90
4 Conclusion.....	92

Chapitre VI – Implémentation

1- Introduction.....	93
2- Outils de travail.....	93
2-1- Choix du langage.....	93
2-2- Choix de l'environnement de développement.....	94
3- Application.....	95
Conclusion Générale.....	98

Bibliographie

Liste des Figures

- Figure 1-1** : Représentation du pixel
- Figure 1-2** : Rapports d'aspect et résolutions
- Figure 1-3** : Histogramme avec niveau de gris
- Figure 1-4** : 3 histogrammes avec chaque couleur « RVB » d'une image
- Figure 1-5** : Différents réglages de contraste
- Figure 1-6** : Les images en niveaux de gris
- Figure 1-7** : Le codage RVB/RGB
- Figure 1-8** : Le codage CMY
- Figure 1-9** : Le codage Palette
- Figure 1-10** : Le codage TSL/HSL
- Figure 1-11** : Le codage CIE
- Figure 1-12** : Le codage YUV
- Figure 1-13** : Représentation des images vectorielles
- Figure 1-14** : Représentation des images Bitmap/Raster
- Figure 1-15** : Types de traitements d'image
- Figure 2-1** : Tablette de cire
- Figure 2-2** : Le codex
- Figure 2-3** : Tablette d'argile
- Figure 2-4** : Le Papyrus
- Figure 2-5** : Parchemin roulé
- Figure 2-6** : Evolution du support de manuscrits
- Figure 2-7** : Exemples de dégradations biologiques
- Figure 2-8** : Exemple de dégradation physico-chimique
- Figure 2-9** : Exemple de dégradation physique
- Figure 2-10** : Exemple de dégradations chimiques
- Figure 2-11** : Dégradations à cause de la mauvaise qualité des matériaux constitutifs des documents
- Figure 2-12** : Passé vs Future
- Figure 2-13** : Scanner à plat
- Figure 2-14** : Scanner à défilement
- Figure 2-15** : Scanner à Tambour
- Figure 2-16** : Scanner à Microfilm
- Figure 2-17** : Scanner a diapos
- Figure 2-18** : Appareils Photo Numériques
- Figure 3-1** : La Transformée de Fourier

- Figure 3-2** : La Transformée de Fourier à fenêtre glissante
- Figure 3-3** : La différence entre une onde sinusoïdale et une ondelette
- Figure 3-4** : La différence entre une analyse de Fourier (a) et une analyse en ondelette (b)
- Figure 3-5** : Le pavage (découpage) de l'espace temps-fréquence pour la STFT et la TO
- Figure 3-6** : Illustration de la variation du facteur d'échelle
- Figure 3-7** : Algorithme de décomposition
- Figure 3-8** : Schéma d'une itération
- Figure 3-9** : Un niveau de représentation multirésolution d'une image.
- Figure 3-10** : Processus de reconstruction du signal
- Figure 3-11** : Passage d'un niveau à un autre
- Figure 3-12** : Un niveau de reconstruction d'une image
- Figure 3-13** : Modèle de l'analyse multi-résolution 2D
- Figure 3-14** : Représentation de la TOD d'une image.
- Figure 3-15** : L'ondelette de Haar
- Figure 3-16** : L'ondelette de Daubechies 4
- Figure 3-17** : L'ondelette Biorthogonale 1.3
- Figure 4-1** : Suppression des taches en transparence par recalage
- Figure 4-2** : Suppression des taches en transparence par recalage
- Figure 4-3** : Suppression des taches en transparence par analyse de l'orientation
- Figure 4-5** : Séparation recto/verso par analyse de la couleur
- Figure 4-6** : (a) l'image originale avec le texte anglais (b) son histogramme
- Figure 4-7** : (a) l'image ajustée par histogramme et (b) son histogramme
- Figure 4-8** : (a) l'image floue de texte d'origine chinoise et (b) son histogramme
- Figure 4-9** : (a) l'ajustement de l'histogramme de l'image floue et (b) c'est histogramme
- Figure 4-10** : (a) l'image débruitée par le filtre médian et (b) son histogramme
- Figure 4-11** : (a) L'image débruitée par filtre à ondelette et (b) son histogramme
- Figure 4-12** : Principe de la restauration d'une lettre cassée par les contours actifs
- Figure 4-13** : Cas de dégradations de la lettre «e»
- Figure 4-14** : Restauration géométrique de la courbure
- Figure 4-15** : Résultat d'une classification globale en trois classes sur une image de document
- Figure 4-16** : Gauche : Image de document dégradé - Droite : image restaurée par la méthode de
- Figure 4-17** : Restauration de la dégradation colorimétrique par ajustement du contraste
- Figure 5-1** : décomposition avec niveau 1 en utilisant (a) Haar et (b) Biorthogonal9/7.
- Figure 5-2** : Histogramme montré les pics du texte et du fond

- Figure 5-3** : Avant et après le lissage d'un histogramme
- Figure 5-4** : Seuillage d'image d'approximation après la décomposition de (a) Haar et (b) Biorthogonal9/7.
- Figure 5-5** : Résultat final après seuillage en utilisant l'ondelette de (a) Haar, (b) Biorthogonal9/7 et (c) sans ondelettes.
- Figure 5-6** : Le seuillage après l'utilisation de notre méthode (a) Haar, (b) Biorthogonale9/7 et (c) sans ondelettes.
- Figure 5-7** : (a) Image originale, (b) après transformation de (a) avec deux niveaux de décomposition en utilisant l'ondelette de Haar
- Figure 5-8** : Détection des contours avec l'ondelette de (a) Haar, (b) Bior9/7, (c) Daubechies 2.
- Figure 5-9** : Détection des contours par la méthode de fusion des coefficients (détails)
- Figure 5-10** : Graphe de la technique de restauration des documents suggéré
- Figure 5-11** : Deux images manuscrites bruitées
- Figure 5-12** : Des images manuscrites débruitées par la méthode proposée en utilisant les différents types d'ondelettes
- Figure 5-13** : Image après l'élimination les informations
- Figure 5-14** : Résultats après débruitage par élimination des coefficients (détails) avec les différents types d'ondelettes.

Liste des Tableaux

Tableau 2-1 : Recommandations pour la numérisation de différents supports

Tableau 3-1 : 1ere itération de la multirésolution

Tableau 3-2 : 2eme itération de la multirésolution

Tableau 3-3 : Types d'ondelettes de Daubechies

Tableau 4-1 : Nombre de Caractères reconnus et taux de précision
par des méthodes de débruitage différents

Tableau 5-1 : Les résultats de la décomposition par différent types d'ondelettes

ملخص

شهد مجال تكنولوجيا المعلومات في السنوات الأخيرة تطورا كبيرا في جانب الأجهزة أو جانب البرمجيات، يمكن العثور في هذه الأيام على أنظمة تستطيع تنفيذ العمليات الأكثر تعقيدا.

هذا التطور أدى إلى رقمنة المخطوطات القديمة، معالجة الصور هي خطوة أساسية في رقمنة المخطوطات القديمة. وتستخدم أنواع مختلفة من المعالجات حيث تسمح بترميم أو تنظيف الصور واستخراج مختلف الهياكل من الوثيقة (الرسوم التوضيحية والصور والنصوص والخطوط)، وفي بعض الحالات التعرف على الحروف في الصور (الاعتراف البصري للأحرف).

تحويل المويجات هو أسلوب لتحليل الإشارات، التي اقترحت في بداية 1980م، ويستخدم في مجال معالجة الصور للضغط، وتقليل الضوضاء، والكشف عن الحواف... الخ

في بحثنا سوف نركز على استعادة الوثائق القديمة المكتوبة بخط اليد باستخدام التحويل بالمويجات.

مفاتيح الكلمات: معالجة الصور، المسح الضوئي، المخطوطات القديمة، ترميم الوثائق القديمة، التحويل بالمويجات.

Résumé

Le domaine de l'informatique a connu ces dernières années une grande évolution soit dans le côté matériel (Hardware) ou dans le côté logiciel (Software), on peut trouver dans nos jours des systèmes qui réalisent les opérations les plus compliquées qu'on peut voir.

Ce développement a mené à la numérisation des anciens manuscrits, le traitement des images est une étape incontournable de la numérisation des manuscrits anciens. Les différents types de traitements permettent de restaurer ou nettoyer les images, d'extraire les différentes structures du document (illustrations, éléments graphiques et textuels, lignes de texte) et dans certains cas de reconnaître les caractères présents dans l'image (Reconnaissance Optical des caractères « OCR »).

La transformée en ondelette est une méthode d'analyse du signal introduite au début des années 1980, son utilisation dans le domaine de traitement d'image est la compression, le débruitage, la détection des contours...etc.

Dans notre recherche on s'intéressera à la restauration des anciens documents manuscrits Algériens à l'aide de la transformée en ondelettes.

Mots clés : Traitement des images, Numérisation, Manuscrits anciens, Restauration des anciens documents, Transformations en ondelettes.

Abstract

The field of information technology has in recent years known a great evolution in the hardware side or software side, it can be found in our day systems that perform the most complicated operations we can see.

This development led to the digitization of old manuscripts, image processing is an essential step in the digitization of old manuscripts. Different types of treatments are used to restore or clean images, extract the different structures of the document (illustrations, images, text, lines) and in some cases to recognize the characters in the image (Optical Recognition characters "OCR").

The wavelet transform is a signal analysis method introduced in the early 1980s, its use in the field of image processing is compression, denoising, edge detection ... etc.

In our research we will focus on the restoration of old manuscript documents using the wavelet transformation.

Keywords: Image Processing, Scanning, Ancient Manuscripts, Restoration of old documents, Wavelet Transformations.

Chapitre I - Introduction aux images numériques

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 :

Une image numérique est une image acquise, créée, traitée par un système de numérisation (Appareil photo, Scanner...etc.). Elle est stockée en langage binaire sur système de stockage (mémoire, disque optique...etc.). [1]

3- Caractéristique de l'image numérique :

3-1 Pixel :

Sorti de l'appareil photo une image est composée d'un nombre x de pixels. Un pixel a une couleur exprimée (codée) en langage binaire mathématique (uniquement des 0 et des 1). L'œil humain est capable de discerner aux environs de 220 nuances dans une couleur. On comprend donc tout de suite qu'il faut cette quantité minimum de nuances pour avoir une qualité correcte, appelée parfois qualité photo. Ce seront donc toujours des multiples de 2 en progression géométrique qui donneront les valeurs de travail $2, 2^2, 2^4, 2^8, 2^{12}, 2^{16}...$ C'est 2^8 soit 256 qui se rapproche le plus des 220 niveaux requis pour une bonne séparation des nuances, on comprend donc tout de suite pourquoi on retrouve la majorité des APN réglés sur 256 niveaux de nuances par couleur. On dit que l'image est codée en *8 bits par couche* ($2^8 = 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 256$ niveaux) soit 8 pour le R, 8 pour le V, 8 pour le B au total 24 bits en RVB On appelle cela : la profondeur de couleur ou dynamique de l'image.

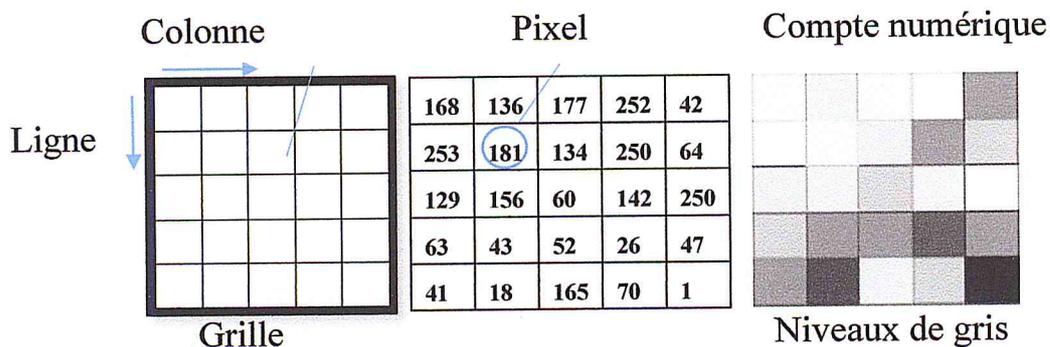


Figure 1-1 : Représentation du pixel

3-2- La Définition, la Résolution et la Dimension :

Les termes Résolution et Définition sont souvent confondus dans le langage de la photo numérique. Ainsi le menu de configuration d'un APN donnera plusieurs choix de réglage de la Résolution d'une image alors que l'on devrait plutôt parler de Définition.

Les mots Résolution et Définition sont liés à la notion de pixels. Le pixel est l'élément de base d'une image ou d'un écran, c'est-à-dire un point.

La résolution est le nombre de points (ou pixels) que comporte une image numérique en largeur et en hauteur exprimé par : (exemple : 1600x1200).

La définition est le nombre de pixels par unité de longueur. La définition est exprimée le plus souvent en ppp (point par pouces) ou en dpi (dots per inch).

Rappel : 1 pouce = 2.54 cm.

Une image de 100 ppp (ou 100 dpi) contient 10 000 points par pouce carré. (100 x 100 = 10 000).

La définition définit la netteté d'une image et sa qualité d'affichage à l'écran. Plus la définition est grande (c'est-à-dire plus il y a de pixels dans une longueur de 1 pouce), plus votre image est précise dans les détails, c'est un critère important lorsque l'on utilise un scanner. [3]

La dimension 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. [4]

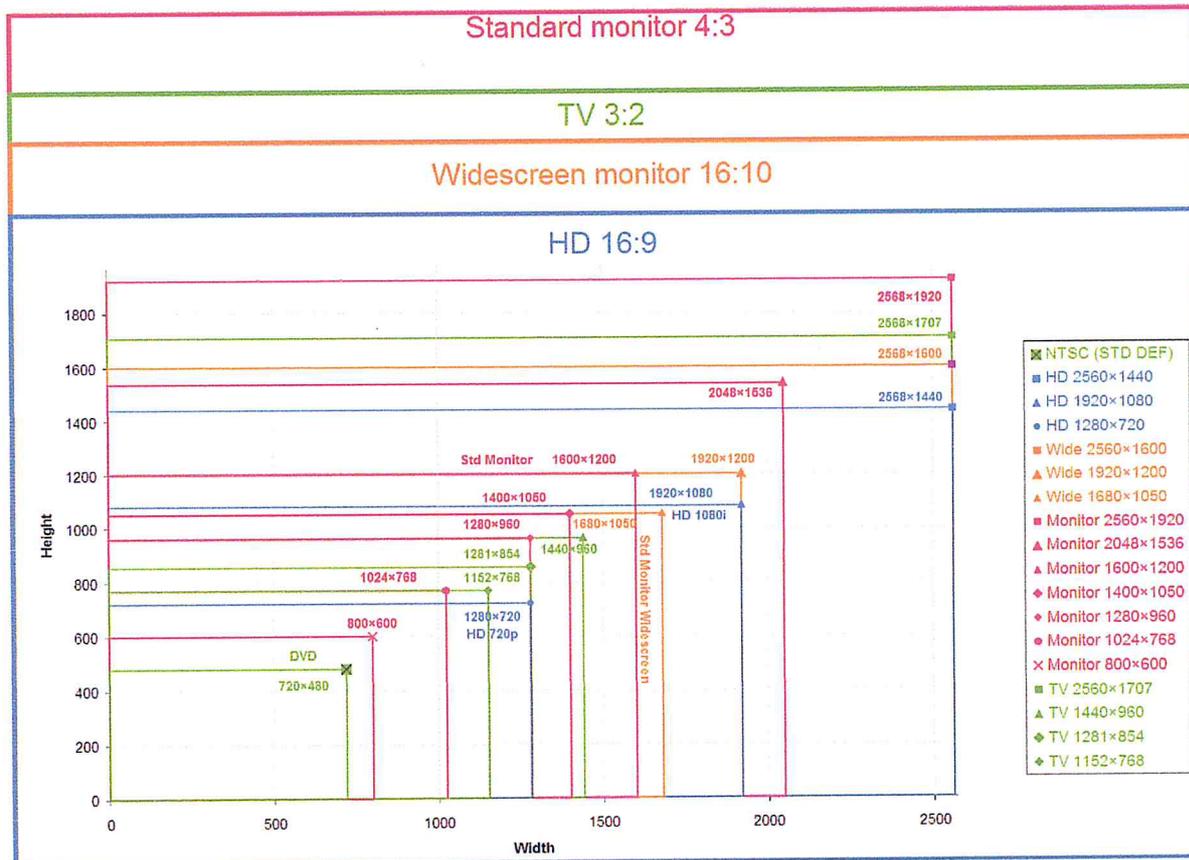


Figure 1-2 : Rapports d'aspect et résolutions [5]

Formules :

Résolution = définition / dimension

Définition = résolution x dimension

Dimension = résolution / définition

3-3- Histogramme :

Un histogramme est un graphe statistique permettant de représenter la distribution des intensités des pixels d'une image, c'est-à-dire le nombre de pixels pour chaque intensité lumineuse. Par convention un histogramme représente le niveau d'intensité en abscisse en allant du plus foncé (à gauche) au plus clair (à droite). Ainsi, l'histogramme d'une image en 256 niveaux de gris sera représenté par un graphe possédant 256 valeurs en abscisses, et le nombre de pixels de l'image. [6]

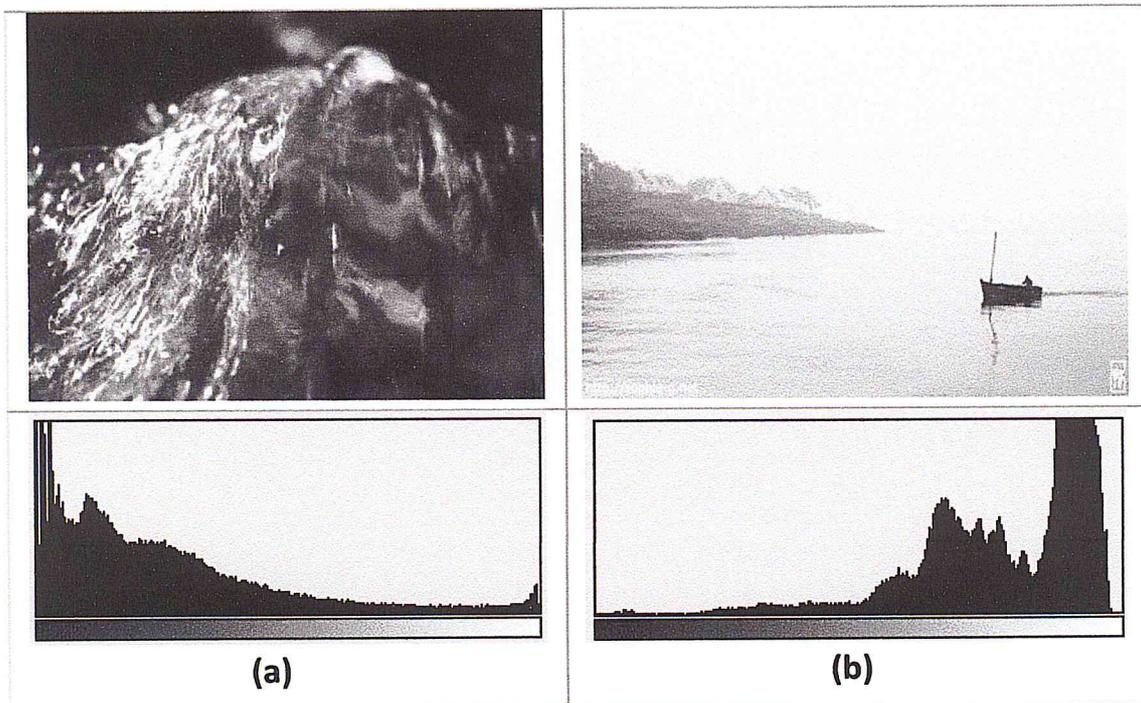


Figure 1-3 : Histogramme avec niveau de gris

(a) Image sombre (sous-exposée) : beaucoup de pixels pour les faibles valeurs de niveau de gris, (b) Image claire (sur-exposée) : beaucoup de pixels pour les valeurs claires de niveau de gris

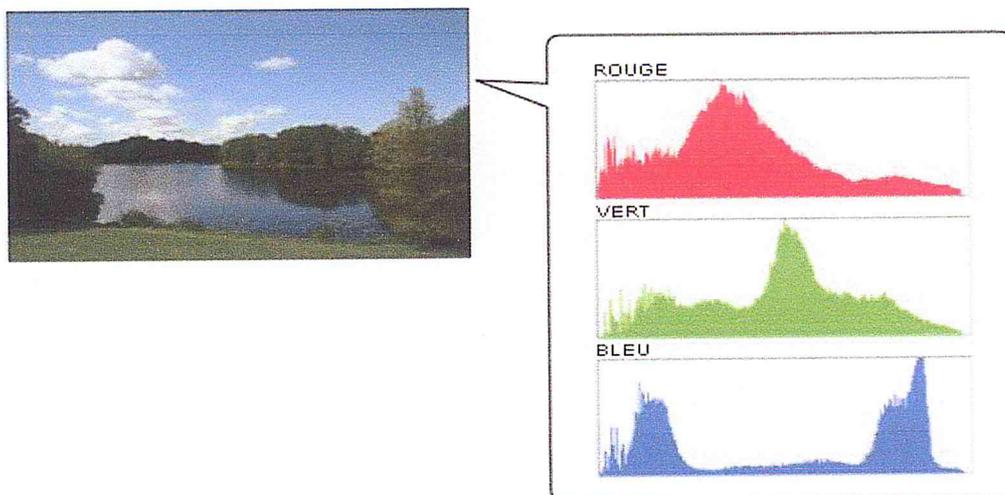


Figure 1-4 : 3 histogrammes avec chaque couleur « RVB » d'une image

3-4- La luminance :

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 :

- 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.

3-5- Contraste :

Le contraste est la différence d'intensité lumineuse entre les zones les plus claires et les plus sombres de l'image. Il faut juste partir du principe que plus c'est important, mieux c'est, car plus le rendu sera fidèle.

Ce taux de contraste est lié à l'option contraste, réglable sur votre téléviseur. Lorsqu'on augmente cette option contraste, l'image devient plus vive : les noirs plus noirs, les teintes claires plus lumineuses. Si le taux de contraste du téléviseur n'est pas assez élevé, les teintes claires deviendront vite blanches et on perdra des informations à l'image (voir l'exemple). Un fort taux de contraste permet donc une image toujours plus vive, pêchue, sans perte de finesse dans les teintes claires. [7]

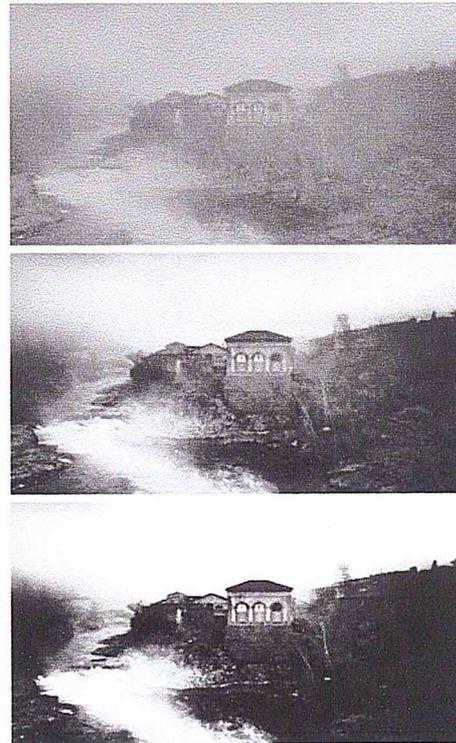


Figure 1-5 : Différents réglages de contraste

3-6- Niveau de gris :

Une image en niveau de gris est une image composée de points gris plus ou moins foncés. (Pour chaque point l'ordinateur enregistre une valeur de gris entre le noir et le blanc). En général on sauvegarde les images à 16 niveaux de gris ou à 256 niveaux. (Noir et blanc compris) [8]



Figure 1-6 : Les images en niveaux de gris occupent 4 fois (16 niveaux) ou 8 fois (256 niveaux) plus de mémoire qu'une image de même taille en noir et blanc. L'effet subjectif d'une image en 256 niveaux est nettement meilleur qu'en 16.

3-7- Région :

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. [4]

3-8- Contour :

C'est un ensemble de pixels formant une frontière entre deux ou plusieurs régions voisines. [4]

3-9- Bruit (parasite) :

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. [9]

3-10- Les représentations de la couleur :**a- Le codage RVB/RGB :**

Pour représenter les couleurs, les écrans utilisent le codage RVB (Rouge, Vert, Bleu). Ce système, basé sur la synthèse additive des couleurs, représente chacune d'entre elles par ses niveaux de rouge, vert et bleu. Chacun de ces niveaux est codé par un nombre allant de 0 à 255. On retrouve donc nos $256 \times 256 \times 256 = 16777216$ couleurs. [10]

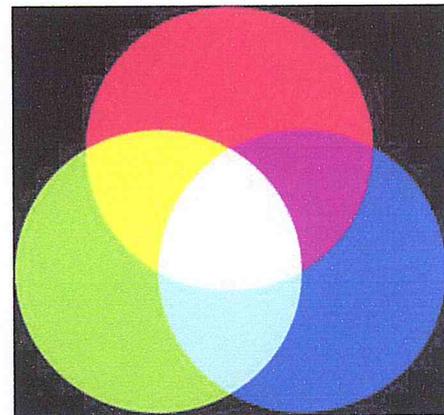


Figure 1-7 : Le codage RVB/RGB

b- Le codage CMY :

Le codage *CMY* (*Cyan, Magenta, Yellow*, ou *Cyan, Magenta, Jaune* en français, soit *CMJ*) est à la synthèse soustractive, ce que le codage *RGB* est à la synthèse additive. 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

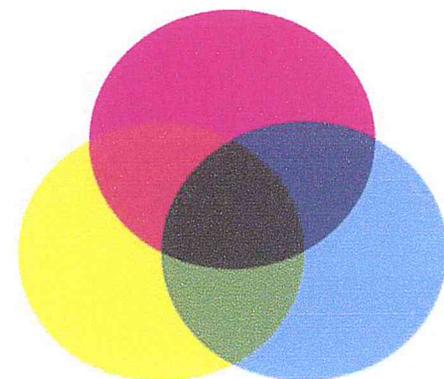


Figure 1-8 : Le codage CMY

noire que l'on appelle *noir pur*. On parle alors de **quadrichromie**, ou **modèle CMYK** (*Cyan, Magenta, Jaune, Noir pur*, ou en français *CMJN*). [11]

c- Le codage Palette (Indexed images) :

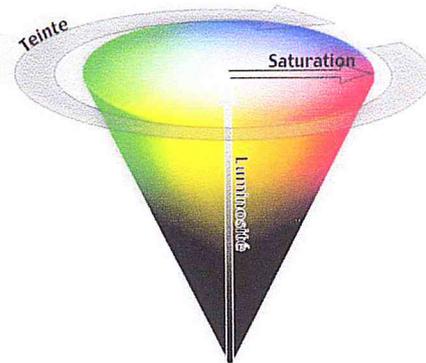
Dans ce type représentation chaque pixel est représenté par le numéro de la palette qui contient les valeurs RGB. [12]

FFFFFF	000000	333333	666666	999999	CCCCCC	CCCC99	9999CC	666699
660000	663300	996633	003300	003333	003399	000066	330066	660066
990000	993300	CC9900	006600	336666	0033FF	000099	660099	990066
CC0000	CC3300	FFCC00	009900	006666	0066FF	0000CC	663399	CC0099
FF0000	FF3300	FFFF00	00CC00	009999	0099FF	0000FF	9900CC	FF0099
CC3333	FF6600	FFFF33	00FF00	00CCCC	00CCFF	3366FF	9933FF	FF00FF
FF6666	FF6633	FFFF66	66FF66	66CCCC	00FFFF	3399FF	9966FF	FF66FF
FF9999	FF9966	FFFF99	99FF99	66FFCC	99FFFF	66CCFF	9999FF	FF99FF
FFCCCC	FFCC99	FFFFCC	CCFFCC	99FFCC	CCFFFF	99CCFF	CCCCFF	FFCCFF

Figure 1-9 : Le codage Palette

d- Le codage TSL/HSL :

Le modèle TSL/HSL est un modèle colorimétrique perceptuel car il se rapproche fortement de la perception physiologique de la couleur par l'œil humain. Dans ce système, les couleurs sont toujours caractérisées par trois dimensions mais qui ont une signification tout autre que dans le modèle RVB, puisqu'elles représentent ici la teinte, la saturation et la luminosité.



La teinte : qui correspond à la perception de la couleur est mesurée sur une échelle circulaire (cercle de chromaticité de Newton) par un angle de 0° à 360°.

La saturation : mesure le degré de pureté d'une couleur, c'est-à-dire la quantité de gris ajoutée à la couleur. Elle est représentée par le rayon d'une section circulaire du cône et varie de 1 (couleur pure ou saturée) à 0 (niveau de gris correspondant).

La luminosité : représente le degré d'éclaircissement ou d'assombrissement d'une couleur. Elle est définie selon une échelle linéaire allant de 0 (noir) à 1 (blanc) en passant par tous les niveaux de gris. Sur la figure ci-dessous, la luminosité est représentée par la droite joignant les deux sommets des cônes et passant par le

Figure 1-10 : Le codage TSL/HSL

centre du cercle chromatique. Les teintes sur le cercle chromatique sont donc toutes au même niveau de luminosité de 50%. [13]

e- Le codage CIE :

Les couleurs peuvent être perçues différemment selon les individus et peuvent être affichées différemment selon les périphériques d'affichage.

La Commission Internationale de l'Eclairage (CIE) a donc défini des standards permettant de définir une couleur indépendamment des périphériques utilisés. A cette fin, la CIE a défini des critères basés sur la perception de la couleur par l'œil humain, grâce à un triple stimulus.

En 1931, la 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) :

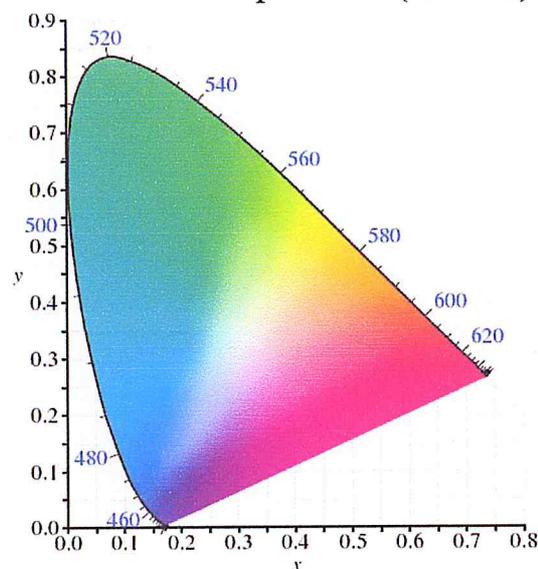


Figure 1-11 : Le codage CIE

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. [14]

f- Le codage YUV :

On utilise un système nommé YUV. Ou Y représente la **luminance** et U , V les composantes Rouge et Bleu (**chrominances**). Ces 3 informations permettent de restituer au final les composantes RVB et apporte l'avantage de permettre une compression facile des couleurs et de fiabiliser le transport du signal vidéo. Mais surtout la partie Y (luminance) permet une totale compatibilité avec les

anciens équipements noir et blanc, ce qui a permis le déploiement de la diffusion couleur alors que la totalité des foyers étaient alors équipés de téléviseurs noir et blanc.

La distinction entre des vidéos personnelles et “broadcast” (qualité diffusable à la télé) vient en grande partie du type de codage YUV utilisé pour le signal vidéo. Ainsi le codec DV (utilisé par les caméscopes numérique grand-public) utilise un YUV de 4:2:0 en PAL et 4:1:1 en NTSC, alors que la qualité “broadcast” du DV (DV50) utilise un YUV de 4:2:2, beaucoup plus fidèle au niveau des couleurs.

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

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

Ainsi U est parfois noté Cr et V noté Cb , d'où la notation $YCrCb$. [15]



Figure 1-12 : Le codage YUV

g- 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 : [16]

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

4- 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 :

4-1- Images vectorielles :

Les formats les plus connus sont PDF, SWF...etc. On utilise un système de coordination (x,y) pour placer les éléments, puis on définit les formes.

Exemple :

```
raw circle
center    0.5, 0.5
radius    0.4
fill-color yellow
stroke-color black
stroke-width 0.05
draw circle
center    0.35, 0.4
radius    0.05
fill-color black
draw circle
center    0.65, 0.4
radius    0.05
fill-color black
draw line
start     0.3, 0.6
end       0.7, 0.6
stroke-color black
stroke-width 0.1
```

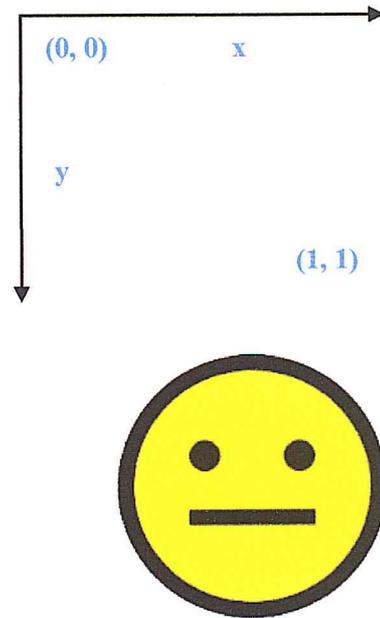


Figure 1-13 : Représentation des images vectorielles

4-2- Images Bitmap/Raster: Les formats les plus connus sont JPG, PNG, GIF...etc. Ce sont des photos digitales dont la représentation est la plus commune des images riches en détails.

C'est une matrice où chaque valeur d'élément (pixel) représente la couleur d'un point dans l'image, la dimension est noté **Hauteur x Largeur**

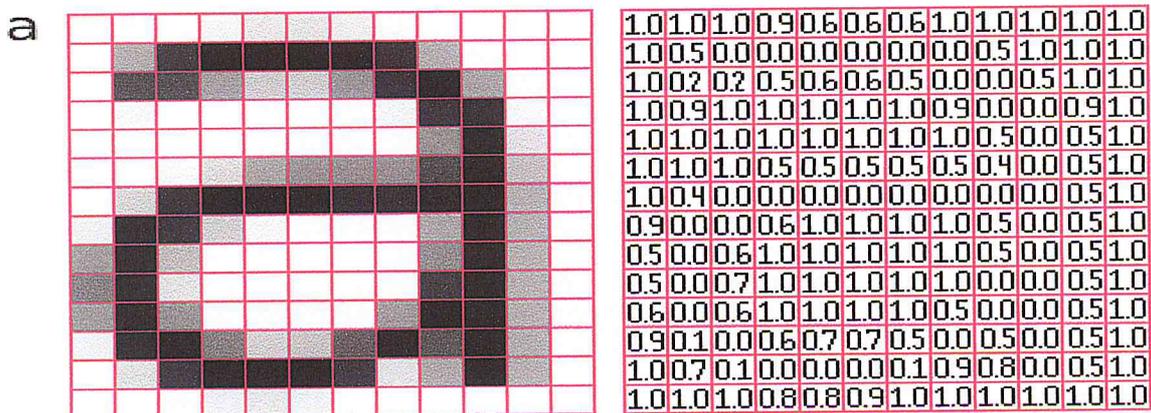


Figure 1-14 : Représentation des images Bitmap/Raster

4-3- Les formats méta fichiers :

Les métras fichiers sont très vite 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 métras fichiers ont la facilité de les porter d'une plateforme à une autre. Les données étant stockées la plupart de temps, en ASCII, les problèmes d'ordre de bit n'apparaissent pas. L'ASCCII offre aussi une plus grande aptitude à la compression des fichiers.[9]

4-3-1- Les avantages :

- Les métras fichiers ont des avantages des fichiers bitmap et des fichiers vectoriels.
- La plupart des formats sont binaires quelques-uns sont orientés, le partage d'un ordinateur à un autre se fait généralement sans encombre.
- Les métras fichiers peuvent généralement être compressés avec un gain important.

4-3-2- Les inconvénients :

- Avec un nouveau format il y a risque de créer de nombreux problèmes lors d'échange de données avec d'autres logiciels.
- L'exemple d'une image en méta fichier en format WMF. Le format WMF est un méta format en effet, il est utilisé pour stocker des images vectorielles, des images bitmap sur disque ou en mémoire afin de les utiliser ultérieurement sous Windows.
- Un fichier contient une suite d'objet chacun décrit par un en-tête. Le format WMF peut contenir 65535 objets au maximum. Le type possible pour chaque objet (cercle, carré, bitmap...) est défini dans la librairie 'windows.h'.

5- Le traitement d'images :

5-1- Introduction :

Le traitement d'images est une discipline de l'informatique et des mathématiques appliquées qui étudie les images numériques et leurs transformations, dans le but d'améliorer leur qualité ou d'en extraire de l'information.

Il s'agit d'un sous-ensemble du traitement du signal dédié aux images et aux données dérivées comme la vidéo (par opposition aux parties du traitement du signal consacrées à d'autres types de données : son et autres signaux monodimensionnels notamment), tout en opérant dans le domaine numérique (par opposition aux techniques analogiques de traitement du signal, comme la photographie ou la télévision traditionnelles).

Dans le contexte de la vision artificielle, le traitement d'images se place après les étapes d'acquisition et de numérisation, assurant les transformations

d'images et la partie de calcul permettant d'aller vers une interprétation des images traitées. Cette phase d'interprétation est d'ailleurs de plus en plus intégrée dans le traitement d'images, en faisant appel notamment à l'intelligence artificielle pour manipuler des connaissances, principalement sur les informations dont on dispose à propos de ce que représentent les images traitées (connaissance du domaine).

La compréhension du traitement d'images commence par la compréhension de ce qu'est une image. Le mode et les conditions d'acquisition et de numérisation des images traitées conditionnent largement les opérations qu'il faudra réaliser pour extraire de l'information. En effet, de nombreux paramètres entrent en compte.

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, écographie...).

5-2- Type de traitement d'image : On distingue deux types de traitement d'image :

5-2-1-Traitement de bas niveau :

Opèrent en général, sur les grandeurs calculées à partir des valeurs attachées à chaque point de l'image sans faire nécessairement la liaison avec la réalité qu'elles représentent. Par exemple, la détection des contours est un traitement de bas niveau qui est effectué «sans comprendre» l'image. Le contour détecté peut très bien ne pas correspondre à un bord d'objet dans la réalité et ne présenter aucun intérêt, Ainsi les traitements de bas niveau opèrent plutôt sur des données de nature numérique.

5-2-2-Traitement de haut niveau :

S'appliquent à des entités de nature symbolique associées à une représentation de la réalité extraite de l'image. Ils sont relatifs à l'interprétation et à la compréhension de l'image et sont exprimés avec des mots du vocabulaire de l'application. Par exemple, des zones d'une image aérienne peuvent être caractérisées par leur forme (rectangulaire, linéique,..), être étiquetées avec les termes : bâtiment, route, bosquet, ombre, etc... Des relations entre ces zones sont exploitées pour comprendre la scène étudiée, par exemple : une route ne peut pas être incluse dans un bosquet.

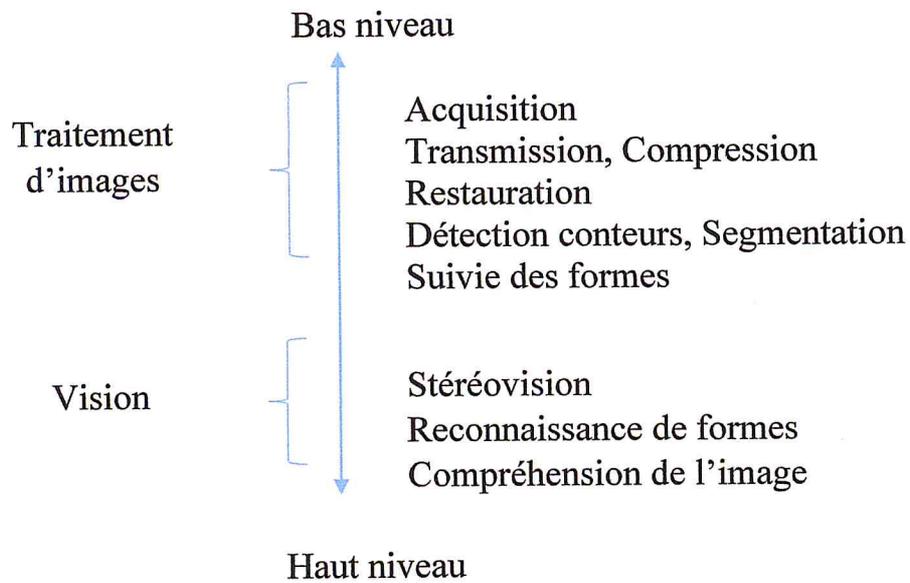


Figure 1-15 : Types de traitements d'image

6- Conclusion :

On a cité dans ce chapitre quelques notions et concepts de base sur le domaine du traitement d'image qui vont nous aider à commencer notre travail, ce domaine est un domaine très actif et très utilisé dans ces dernières années.

Dans le chapitre suivant on va présenter quelques définitions sur les anciens manuscrits et les causes de leurs dégradations.

Chapitre II - Les manuscrits et leurs dégradations

1- Introduction :

Les documents anciens représentent un patrimoine culturel très précieux, témoin de la civilisation passée et utile pour les générations futures. Ces documents encourent une dégradation progressive et sont menacés d'un réel danger de disparition. Ils se heurtent à de multiples problèmes telle que : l'effet de transparence, l'irrégularité de fond et les tâches d'humidité absorbé par le papier...etc.

Pour sauver ce patrimoine il nécessite la numérisation, la numérisation de documents sur support analogique, notamment sur papier, est en croissance dans les organisations depuis plus de 25 ans. Les améliorations techniques de l'équipement et la réduction de ses coûts d'acquisition ont contribué à populariser le transfert d'information analogique sous forme numérique. Longtemps utilisée principalement à des fins de diffusion, la numérisation gagne en popularité comme élément d'un programme de gestion documentaire.

2- Définition du manuscrit :

Un manuscrit (du latin *manu scriptus*) est, littéralement, un texte « écrit à la main », sur un support souple, que ce soit par son auteur « manuscrit autographe » ou par un copiste, avant ou après l'invention de l'imprimerie. Le terme s'applique à différents supports (papyrus, rouleaux, codex de parchemin ou de papier) mais ne s'emploie pas pour des supports rigides, comme les tablettes de cire ou les inscriptions épigraphiques ou rupestres. Avant la mise au point de l'imprimerie, au milieu du XV^e siècle, tous les livres étaient des manuscrits. [17]

3- L'évolution du livre :

L'histoire du livre est si intimement imbriquée à celle des civilisations que les débats sur l'avenir de ce support ne peuvent être que virulents et teintés d'inquiétude. Quand certains parlent de la disparition de l'odeur de l'encre et du papier, pour opposer le livre tel que nous le connaissons aux liseuses, ce n'est pas tant de conservatisme dont il s'agit mais de la crainte de perdre une part de ce qui a construit l'identité des hommes et de l'humanité toute entière.

3-1- Du volumen au Codex :

Cette histoire a commencé tout d'abord avec la civilisation Sumérienne. Les hommes gravaient alors à l'aide d'une tige de roseau des signes cunéiformes sur des tablettes d'argile 3000 ans avant J.-C. Ce fut aussi les Égyptiens qui tracèrent les hiéroglyphes sur des feuilles de

papyrus collées les unes aux autres, constituant ainsi des « volumina », rouleaux de plusieurs mètres, à l’instar du « Papyrus Prisse », recensé comme le plus vieux livre du monde.

Le coût de fabrication du papyrus produit par l’Egypte et la rivalité avec Alexandrie conduira Pergame – ville d’Asie mineure abritant une bibliothèque contenant 200 000 rouleaux à inventer un nouveau support, le parchemin. Ce support peut être utilisé sur deux faces et présente l’énorme avantage de pouvoir être réemployé en grattant le texte précédent.

Le volumen, omniprésent à Rome, sera concurrencé à compter du 1er siècle par une nouvelle forme de livre. Il s’agit de tablettes de cire destinées aux notations d’ordre pratique et reliées entre elles. Cet agencement inspira sans doute le codex, feuilles de parchemin pliées en cahiers et cousues ensemble. Son usage se développera dès le IIIe siècle, avec les débuts de la chrétienté, ce support étant plus commode à consulter et à conserver.



Figure 2-1 : Tablette de cire

En effet, le volumen devant être tenu des deux mains, il était impossible de lire et d’écrire en même temps, à l’inverse du codex. Ce support permet enfin aux lecteurs d’annoter et de se repérer dans le texte à l’aide des numéros de pages qui facilitent la navigation dans le texte (index, table des matières, renvois...). Ainsi, le changement de forme matérielle du livre a changé la façon d’aborder le texte ; la lecture pouvait ne plus être linéaire, mais tabulaire, facilitant ainsi le travail de consultation d’un livre.



Figure 2-2 : Le codex

3-2- Du parchemin au papier :

Une vaste production de manuscrits se développe en France, en Germanie et en Angleterre. Celle-ci dépasse le cadre des monastères et des abbayes ; le livre n’étant plus uniquement un objet de vénération religieuse, mais aussi vecteur d’érudition et d’affirmation du statut social. La création des premières universités suscite une demande importante de la part des

étudiants et, par conséquent, de la société civile. Dans les ateliers, les copistes travaillent alors à la chaîne dans les librairies. Le premier mouvement de démocratisation du livre s'affirmera dès le XVe siècle.

Très vite donc, la nécessité se fait jour de trouver un support moins coûteux et moins long à fabriquer que le parchemin ; c'est ainsi que le papier, inventé en Chine, s'introduit en Europe, mais son usage ne se généralisera qu'à compter de l'invention de l'imprimerie. [18]

Les livres manuscrits aujourd'hui ?

Envié, désiré, commandé, longuement attendu, payé, rangé, conservé avec soin et transmis de génération en génération, le livre manuscrit était un bien d'importance. Son caractère unique, le temps et les matériaux (parchemins, encre ...) nécessaires à sa confection, parfois très coûteux, faisaient du livre un objet cher et précieux.

Aujourd'hui encore, le livre manuscrit est un objet rare et précieux ! Il est une source d'une extrême richesse pour de nombreux chercheurs : historiens, historiens de l'art, philologues, historiographes... C'est un témoignage historique et artistique d'une importance considérable qui nous offre de très nombreux renseignements sur les mœurs et coutumes de la société médiévale.

Il n'est pas étonnant qu'il soit aujourd'hui conservé encore plus précieusement. Placé sous haute surveillance, il est évidemment protégé contre le vol, mais aussi de redoutables ennemis d'un autre genre : la lumière, l'humidité et la température ambiante. À éviter absolument : les variations de températures. Techniquement, il est conservé à 18°C, dans une atmosphère à un taux d'humidité de maximum 60%. Pour le présenter au public, on l'installe dans une vitrine spéciale, éclairée par un faisceau de lumière froide (50 lux maximum) et maintenue à température et à humidité constantes. Le livre manuscrit du Moyen Âge est montré avec prudence et manipulé avec les plus grandes précautions. Le livre manuscrit du Moyen Âge était, reste et sera toujours un trésor !

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. [19]

4- Les supports de l'écriture :

Des dessins sur les murs des grottes préhistoriques aux documents numériques actuels, les supports de l'écriture ont évolué en fonction des techniques, des besoins mais aussi des enjeux sociaux et économiques. La croissance des échanges d'informations et le besoin de les mémoriser ont influencé l'utilisation de tel ou tel support.

4-1- La pierre :

La pierre des grottes préhistoriques accueille les premiers écrits sous forme de dessins, peintures représentant la vie quotidienne (scènes des chasses, animaux, événements particuliers,...). La pierre fut également très utilisée en Egypte et dès l'Antiquité grecque et romaine pour des inscriptions sur les monuments ou les tombeaux. Ces inscriptions avaient essentiellement un but commémoratif.

4-2- Ecorces d'arbres, tablettes d'argile :

Les échanges d'informations et économiques devenant de plus en plus nombreux et volumineux, des supports maniables, d'une utilisation plus pratique et moins coûteuse étaient nécessaires. On utilisait donc des écorces d'arbres ou des tablettes d'argile. L'origine du mot *livre* vient d'ailleurs du latin *liber* qui désigne la pellicule blanchâtre située entre le bois et l'écorce de l'arbre. Sur les tablettes d'argile ont été découvertes des traces de l'écriture cunéiforme (traits en forme de clou) de la civilisation sumérienne (vers 3500 av. JC). L'argile était humidifiée. Le scribe traçait les symboles à l'aide d'un *calame* (pointe de roseau taillée à son extrémité). Ce support fut longtemps utilisé notamment pour des documents comptables, administratifs et relatifs à la propriété. [20]



Figure 2-3 : Tablette d'argile

4-3- Tablettes de cire :

Dès le 13^e s. av. JC apparaissent les tablettes de cire : il s'agissait de planches de bois creusées et recouvertes d'une épaisse couche de cire. On écrivait avec un *stylet*, c'est-à-dire une tige d'os ou de métal avec, à une extrémité, une pointe arrondie et, à l'autre extrémité, une forme aplatie pour effacer. Ces tablettes servaient surtout à prendre des notes, de brouillons, à consigner des listes,... Elles étaient très utilisées dans le bassin méditerranéen de l'Antiquité jusqu'au 15^e s. ap. JC car moins chères que le papyrus ou le parchemin. Elles pouvaient être reliées entre elles par des lanières de cuir et formaient ainsi de véritables livres de bois. La Musée St Raymond, ph. Daniel Martin couche de cire malléable permettait de les utiliser plusieurs fois.

4-4- Papyrus :

Vers 3000 ans av. JC, les Egyptiens développèrent la fabrication du papyrus à partir d'une plante du même nom poussant au bord du Nil. Des

bandelettes étaient découpées dans la tige, puis disposées en deux couches perpendiculaires et collées entre elles par une colle végétale.

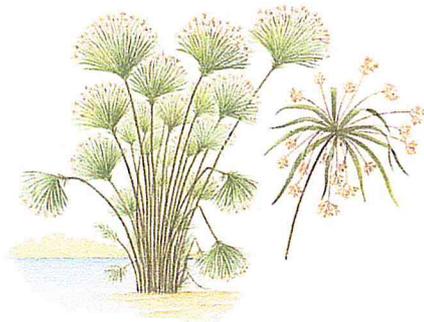


Figure 2-4 : Le Papyrus

Elles étaient ensuite pressées et frappées avec un maillet. Enfin, elles étaient séchées au soleil. On obtenait ainsi de grandes feuilles, vendues sous forme de rouleaux pouvant atteindre 20 à 30 cm de haut et jusqu'à 10 m de long. L'écriture, à l'encre, se faisait en colonnes. On écrivait en général sur une seule face. Pour lire, on déroulait horizontalement le rouleau (*rotulus*) ou *volumen* d'un côté et on le réenroulait de l'autre. La manipulation était facilitée par la présence de deux baguettes de bois fixées à chaque extrémité. La lecture était continue et non sélective. Il était difficile de revenir en arrière, de sauter des passages, de feuilleter.

Les textes étaient surtout destinés à être mémorisés ou lus à haute voix. L'usage du papyrus se répand jusqu'au 9e s. ap. JC dans tout le bassin méditerranéen.

4-5- Parchemin :

La peau d'animal était utilisée depuis le 3e millénaire av. JC mais sous forme de cuir tanné et utilisé d'un seul côté. Le parchemin était connu dès le 3e s. av. JC mais peu utilisé. La légende veut que le parchemin ait été inventé par le roi de Pergame au 2e s. av. JC car sa ville n'était plus approvisionnée en papyrus du fait de la rivalité entre la bibliothèque de Pergame et celle d'Alexandrie. Mais c'est plutôt grâce aux progrès des techniques de travail du cuir permettant l'écriture des deux côtés de la peau que sa production se développa dès le 1er s. ap. JC. *Parchemin* vient du grec "pergamênê" signifiant *peau de Pergame*. Il s'agit d'une peau d'animal (chèvre, mouton, veau...)

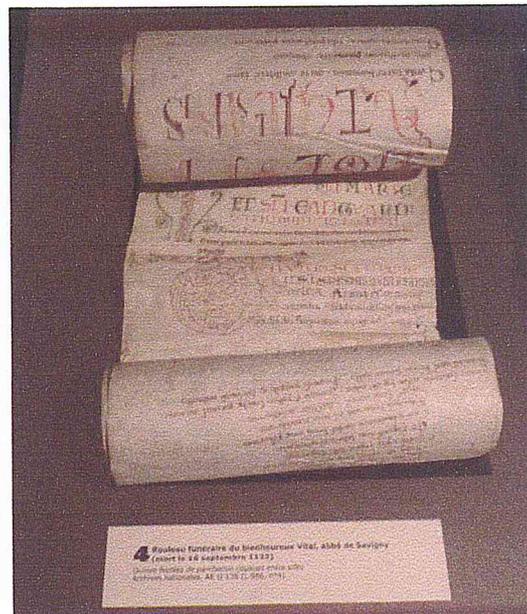


Figure 2-5 : Parchemin roulé

travaillée de façon lisse et fine pour recevoir l'écriture. [21]

Autre avantage : il pouvait être produit partout en Occident alors que l'approvisionnement en papyrus dépendait étroitement des relations internationales entretenues avec le principal producteur, l'Égypte.

4-6- Papier :

On sait maintenant que le secret de fabrication du papier était certainement connu et préservé par les Chinois depuis le 3^e millénaire av. JC. C'est lors de leurs conquêtes vers l'Orient au 8^e s. ap. JC que les Arabes obtinrent ce secret de prisonniers chinois. La fabrication du papier se répandit alors peu à peu dans toute l'Europe et son utilisation se généralisa en France au 13^e s. ap. JC.

Le papier était fabriqué essentiellement à base de chiffons (lin, chanvre) macérés dans l'eau chaude puis broyés. C'est pourquoi on parle de "*papier chiffon*". On obtenait ainsi de longues fibres. Après cette macération, le mélange était passé dans une *forme* (sorte de tamis avec un cadre en bois). La fibre était ensuite mise sous presse pour en extraire l'eau. Il restait alors une fine feuille de fibres que l'on faisait sécher sur une corde dans une pièce ventilée ou étendoir.

Une fois le papier sec, il fallait l'enduire d'un *apprêt* pour qu'il soit plus rigide et puisse recevoir l'encre. Cet encollage était fait à base de rognures de peaux ou de chutes de parchemin.

Au 19^e s. la fabrication de papier se transforma de façon radicale. Pour faire face à la pénurie de l'approvisionnement en chiffon et à la demande toujours croissante des besoins en papier, il fallut trouver une nouvelle matière première : le bois. La pâte de bois avait comme propriété d'être plus facile à manipuler et donc de se prêter à une exploitation mécanique intensive. Les industriels développèrent les machines à papier, se présentant sous forme de grandes rotatives et permettant de produire de grandes longueurs de papier à une vitesse très rapide.

Le 19^e s. fut le siècle du développement de la presse périodique (journaux). Mais ces papiers à base de pâte de bois vieillissent mal. Lors de leur fabrication, ils ont subi des apports chimiques. Le bois contient aussi naturellement des éléments organiques tels que les lignines réagissant de façon négative à la progression du temps et à la lumière. Trop exposées, les fibres se cassent, le papier jaunit, perd de sa solidité et devient friable. On dit que le papier est "*acide*". [23]

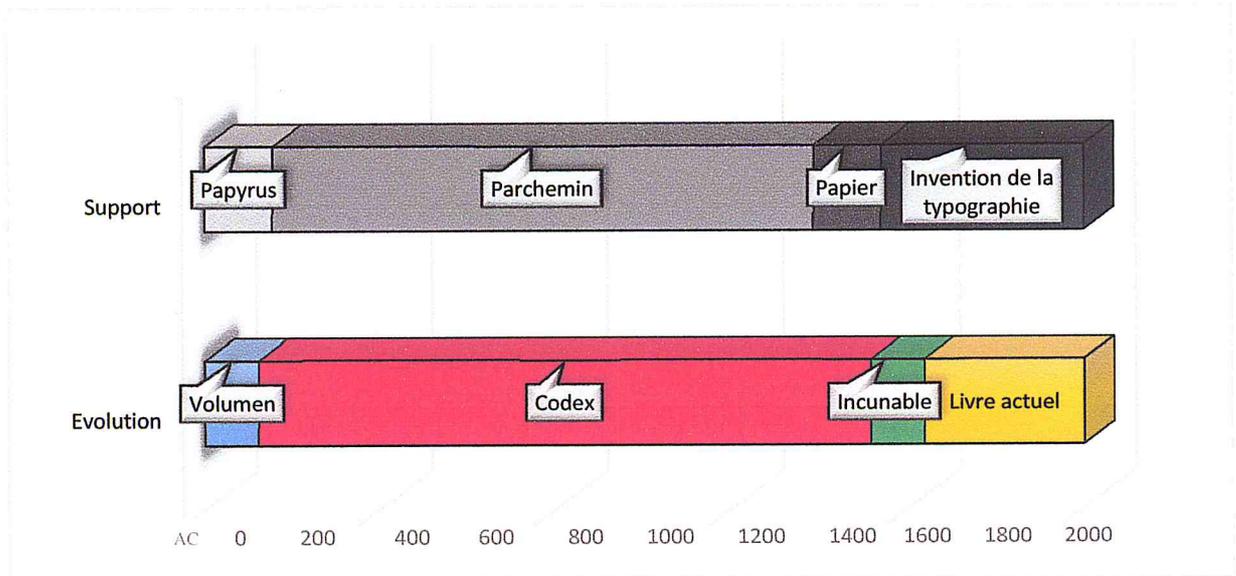


Figure 2-6 : Evolution du support de manuscrits [22]

5- Dégradation des manuscrits :

5-1- Définition :

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

5-2- Les Causes [4] :

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.

5-2-1 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. 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 2-7 : Exemples de dégradations biologiques

Exemples : Les champignons (papyriques, moisissure, mэрule), les Bactéries, Les insectes, Les rongeurs ... etc.

5-2-2 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.

Exemples : La lumière, L'humidité et la température... etc.

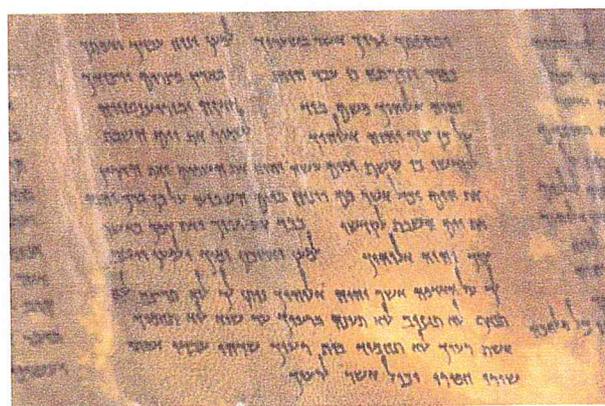


Figure 2-8 : Exemple de dégradation physico-chimique

5-2-3 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.

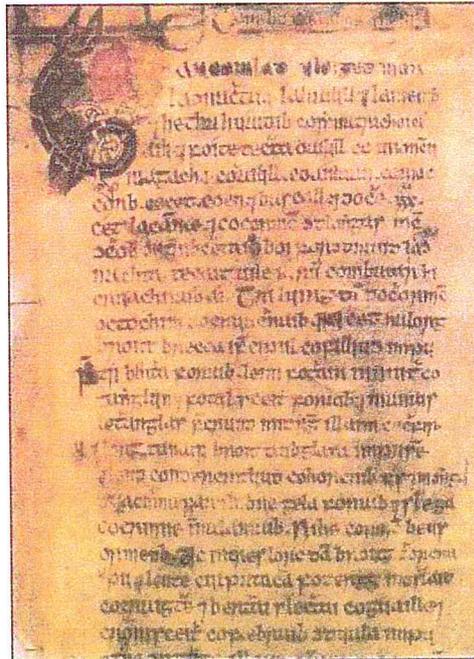


Figure 2-9 : Exemple de dégradation physique

5-2-4 Dégradations chimiques :

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 2-10 : Exemple de dégradations chimiques

5-2-5 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.

5-2-6 Dégradations dues à la capture (Dégradations 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.

5-2-7 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.



Figure 2-11 : Dégradations à cause de la mauvaise qualité des matériaux constitutifs des documents

5-2-8 Les sinistres naturels ou accidentels :

Tels que les inondations, cyclones, vents de sable, le feu, l'eau...etc.

5-2-9 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.

Exemples : Le vandalisme, le manque de soin des lecteurs ou des visiteurs, Les manipulations maladroites, etc.

6 La Numérisation :

Le sentiment que le livre est le moyen le plus sûr de transmettre un héritage, un savoir, remonte à des époques lointaines. Or il existe quelque 25 millions d'ouvrages et documents anciens conservés dans les bibliothèques et les archives qui sont menacés de disparaître, rongés par l'acidité du papier et par l'humidité. Il est capital de préserver ce patrimoine, mémoire du millénaire qui s'achève.

Avec Gutenberg, l'humanité a vécu la révolution de l'imprimerie, microformes et photographies ont permis depuis des années d'accéder à la consultation. La numérisation constitue désormais une alternative, et ouvre de nouveaux horizons pour le livre et l'écrit. [24]

La bibliothéconomie est, depuis le milieu des années 1990, marquée par la double révolution du numérique et du développement des réseaux. Depuis lors, l'utilisation d'Internet, ajoutée à la généralisation et à l'apparition des livres numériques, a amené de profonds changements dans notre rapport avec l'écrit.

La technologie numérique, le multimédia et les réseaux perturbent les points de repère du monde du livre, lequel est basé depuis plus de cinq siècles sur les valeurs de l'imprimerie. Comme lors des précédentes évolutions (culture orale vers la culture écrite, puis culture écrite se confondant avec la culture du livre), l'adoption du numérique suscite de vifs débats. Au cœur des polémiques entretenues, on retrouve la crainte manifestée des dépositaires de savoir de perdre leur identité.

Pourtant, les avantages de la numérisation sont incontestables. Non réductible à un simple jeu de conservation, celle-ci réduit la consultation des documents fragiles, prolonge ainsi leur « durée de vie », offre une grande capacité de stockage, permet la conduite de différentes opérations de recherche et d'intervention multiforme sur le document, et enfin, démultiplie l'accès aux documents, en particulier l'accès à distance.

Les documents numérisés ne devraient pas être destinés à la consommation exclusive des chercheurs. Il importe de cibler aussi le grand public. Une pareille ouverture induit la définition et l'adoption de mesures juridiques appropriées. Elles porteront, entre autres domaines, sur la propriété intellectuelle. En somme, il va falloir enrichir le dispositif des droits avec les droits de propriété intellectuelle. [25]

6-1 Qu'est-ce que la numérisation ?

Numériser, c'est simplement transférer un document d'un support analogique ou inerte (parchemin, papier, métal, cire, magnétique...) sur un support informatique. Pour les images fixes ou les textes, la numérisation passe par un scanner. Mais on peut aussi faire directement des prises de vue numériques ; (par l'intermédiaire d'un appareil photographique ou d'une caméra numérique), ou enregistrer du son numérique (via un magnétophone).

Dans les bibliothèques, les archives et les centres de documentation, la numérisation peut concerner les documents les plus divers :

- Documents écrits anciens, rares, précieux ou fragiles de tous types (manuscrits, chartes, livres, périodiques...)
- Images fixes (photographique, cartes et plans, cartes postales, estampes...)
- Documents sonores et audiovisuels.

Elle peut aussi permettre de réaliser des produits d'actualité, comme des revues de presses ou des dossiers documentaires. [26]

Le Réseau canadien d'Information sur le Patrimoine (RCIP) définit la numérisation « comme le processus par lequel on crée une image numérique (c'est-à-dire contenue dans un ordinateur) à partir d'un objet (document, photographie ou artefact en trois dimensions)».

La numérisation est donc une technologie de conversion en images électroniques des documents originellement disponibles sur support papier ou microfilm. Les documents peuvent être numérisés en mode image (type fac-similé) ou en mode texte intégral en utilisant des logiciels de reconnaissance optique de caractères. [25].

6-2-1 Pourquoi numériser ?

Pierre-Yves Duchemin (1995) répond à cette interrogation dans sa communication présentée lors de la 61^{ème} conférence générale de l'IFLA. Il les décline en ces termes : permettre la conservation, la préservation et la valorisation d'un patrimoine ; améliorer et multiplier les possibilités d'accès aux documents ; permettre leur communication grâce à des media optiques ou électroniques et des réseaux en fournissant un accès à distance ; favoriser la mise en valeur de collections de documents rares ; inciter à la recherche par la découverte de documents méconnus et originaux de par leur contenu, etc.

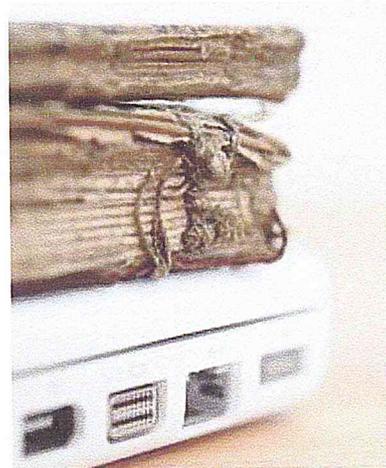


Figure 2-12 : Passé vs Future

6-3 Les objectifs de la numérisation :

- **Sauvegarder les originaux** d'une détérioration graduelle liée à une fréquence de consultation importante et en diminuer la communication « numériser pour sauvegarder ».
- **Faciliter l'accès au patrimoine archivistique**
 - ♣ Mettre à disposition des publics présents ou distants, amateurs ou chercheurs des images comme substituts des originaux
 - ♣ démultiplier les possibilités de consultation en rendant possible l'accès simultané à un même document.
- **Multiplication des copies**

Offrir la possibilité de reproduire aisément des documents qui n'existent la plupart du temps qu'en un seul exemplaire.
- **Aider et susciter la recherche**
- **Alléger la charge du personnel** liée e à la gestion du fonds en supprimant les tâches répétitives.

6-4 Type de scanners utilisés pour la numérisation [27]

6-4-1 Scanners à Plat :

Le scanner à plat est le type de scanner le plus connu et le plus répandu, et ce pour de bonnes raisons. Il convient à de nombreuses utilisations, est facile à utiliser, et largement disponible. Sa popularité pour la publication web a ouvert les portes d'un marché énorme, en faisant chuter les prix des entrées de gamme à moins de \$100. A l'autre bout, les unités professionnelles destinées au marché des arts graphiques rivalisent en qualité avec les scanners à tambour. Tous utilisent la même technologie de base,



Figure 2-13 : Scanner à plat

dans laquelle un capteur lumineux (généralement un CCD) et une source lumineuse, tous deux montés sur un bras mobile, balayent le document immobile placé sur une plaque en verre. Les ADH (Automatic document handlers, manipulateurs de documents automatisés) sont disponibles sur certains modèles, et peuvent augmenter le rythme de production tout en réduisant la fatigue de l'opérateur lors de la manipulation d'ensembles de documents uniformes en état général convenable. Le scanner vertical est une variante spécialisée du scanner à plat dans laquelle la source de lumière du scanner, le dispositif de capteurs et les optiques sont déplacés vers un ensemble d'appui sous lequel un volume relié peut être placé face vers le haut pour la scannérisation.

6-4-2 Scanners à défilement :

Les scanners à défilement utilisent la même technologie de base que les scanners à plat mais maximisent le débit, aux dépens de la qualité, en général. Normalement conçus pour des environnements professionnels de traitement de grands volumes, ils scannérisent généralement en noir et blanc ou en niveaux de gris à des résolutions relativement basses. Les documents doivent être de tailles égales et assez solides pour résister à une manipulation un peu rude, bien que les mécanismes de transport sur les modèles plus récents permettent de réduire l'effort. Que le transport soit effectué par rouleau, ceinture, tambour ou vide, le capteur et la source lumineuse restent immobiles tandis que le document est passé au-dessus. Une sous-classe importante de scanners à défilement sont des

modèles verticaux spécialement conçus pour les documents de taille supérieure tels que les cartes ou les plans d'architecture.

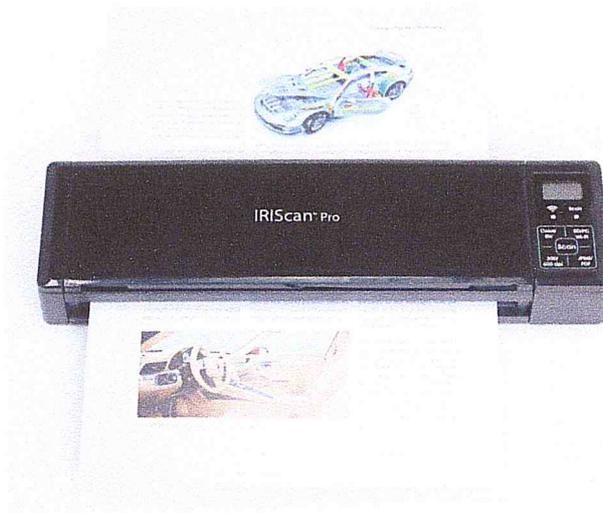


Figure 2-14 : Scanner à défilement

6-4-3 Les Scanners à Tambour :

Les scanners à tambour offrent la résolution la plus élevée, la numérisation la meilleure de tous les types de scanners, mais à un certain prix. En dehors de leur coût élevé, les scanners à tambour sont lents, non adaptés aux documents fragiles et exigent des compétences élevées de manipulation. Par conséquent, on les retrouve en général dans les bureaux de prestataires de services pour le marché pré-presse.



Figure 2-15 : Scanner à Tambour

6-4-4 Les Scanners à Microfilms :

Les scanners à microfilms sont des appareils pointus pour la numérisation des films en bobine, microfiches et cartes à microfilm. Une bonne qualité cohérente à partir d'un scanner à microfilms peut être difficile à obtenir car ils peuvent être complexes opérationnellement, la qualité et la condition du film peuvent varier, et parce qu'ils offrent des possibilités d'amélioration limitées. Seules quelques entreprises fabriquent des scanners à microfilms, et le manque de concurrence contribue à maintenir les prix très élevés de ces appareils.

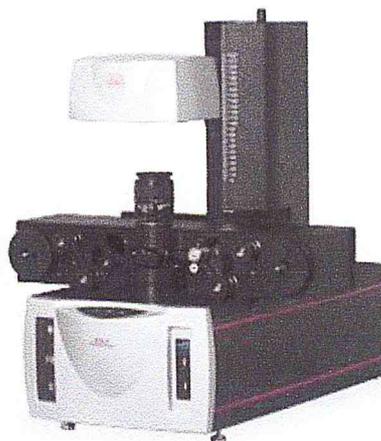


Figure 2-16 : Scanner à Microfilm

6-4-5 Scanners à Diapos :

Les scanners à diapos sont utilisés pour numériser des bibliothèques de diapositives existantes ainsi que les intermédiaires photos d'objets tridimensionnels et les documents non adaptés à la numérisation directe. L'utilisation de médias transparents produit en général une image avec une plage dynamique élevée, mais en fonction de la taille de l'original, la résolution peut être insuffisante pour certains besoins.



Figure 2-17 : Scanner a diapos

6-4-6 Appareils Photo Numériques :

Les appareils photo numériques combinent un scanner avec les optiques de l'appareil afin de former un outil versatile pouvant produire des images de haute qualité. Bien qu'ils soient plus lents et plus difficiles à utiliser qu'un scanner à plat, les appareils numériques s'adaptent parfaitement à tous types de documents et d'objets. Les objets les plus délicats et fragiles peuvent être capturés en toute sécurité, bien que les besoins d'éclairage pour la prise de vues puissent impliquer des précautions concernant la protection de l'œuvre à la lumière. La technologie des

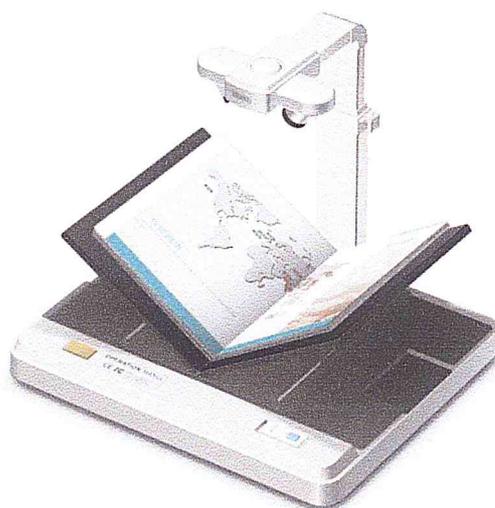


Figure 2-18 : Appareils Photo Numériques

appareils numériques évolue sans cesse, aidée en cela par la demande grandissante des consommateurs.

6-5 Les avantages de la numérisation :

Les avantages de la numérisation sont nombreux. Les plus importants sont :

- la facilité de consultation : les moteurs de recherche permettent en effet en quelques secondes d'aller chercher, au sein de textes de plusieurs milliers de pages, la citation recherchée ou d'en extraire systématiquement toutes les phrases dans lesquelles, par exemple, l'auteur a utilisé un de ses mots préférés. Les gestionnaires de fichiers permettent, quant à eux, de classer toutes les œuvres numérisées par genre, type et catégorie.
- Une plus grande accessibilité et une meilleure promotion du document : l'utilisation de systèmes de stockage robotisés de grande capacité rend la collection numérisée accessible à n'importe quel moment du jour avec un minimum de personnel. La numérisation offre donc un accès distant et par plusieurs personnes en même temps, à des documents peu accessibles ou uniques, leur assurant ainsi une valorisation.

- Une recherche plus performante : la numérisation facilite et rend plus efficace la recherche. En effet, elle offre la possibilité d'effectuer des recherches dans les catalogues aussi bien locaux que lointains et de créer un tissu (web) de liens ou de pointeurs renvoyant à des documents connexes. Des recherches multicritères ou de texte intégral peuvent être effectuées. Des techniques de recherche analogues sont actuellement en cours d'élaboration pour les documents sonores et les images. Une diffusion élargie d'images de collections locales ou uniques encourage l'utilisation savante de ressources enrichies.
- La possibilité d'utiliser toutes les fonctions d'édition qui correspondent à des opérations de couper, copier et coller le document ciblé. Dans le cadre d'études comparatives, on peut ainsi envisager d'analyser un grand nombre de documents ou d'extraits de documents.
- La reproduction : elle peut se faire sous forme de copie de l'information déjà numérisée en utilisant soit le même format de stockage soit un autre format numérique, sans perte de qualité. La reproduction à l'identique se fait sans faire subir des dégradations à l'original. On peut aussi obtenir une reproduction en utilisant une imprimante. Les techniques de reproduction ne se limitent pas à l'écrit, mais s'appliquent également au son et à l'image. Le « copier/coller » s'applique aussi bien à un morceau de musique qu'à une photographie ou à une séquence de film, et la reproduction de ces œuvres de l'esprit, qui supposait autrefois un travail parfois long et fastidieux, se fait aujourd'hui, grâce au numérique, avec une facilité déconcertante.
- La préservation et la conservation des documents originaux : les copies numériques facilitent la sauvegarde du document, en réduisant la fréquence de consultation et donc des facteurs d'usure. À l'heure actuelle, tous les documents ne se prêtent pas à la réalisation d'une copie numérique de qualité suffisante.
- L'utilisation comme support pédagogique : le chercheur ou l'enseignant est souvent appelé à communiquer les résultats de ses recherches ou encore son savoir-faire pratique à des collègues ou à des étudiants.
- La gestion des collections : le choix des objets à numériser exige une définition des priorités, et donc une étude de la valeur relative des objets, d'où une meilleure connaissance des collections au sein de l'établissement. La sécurité des collections est améliorée pendant que leur accès et les services au public sont accrus.
- L'édition numérique : la numérisation a permis à l'édition électronique d'ouvrir une nouvelle fenêtre sur le marché, notamment, en limitant considérablement les délais et les coûts de fabrication, de diffusion et de distribution. Cela permet par exemple aux éditeurs électroniques de faire des offres éditoriales ou d'organiser des manifestations littéraires suivant l'actualité du moment, et ce, à moindre coût.

- En bref, la numérisation nous paraît rentable du point de vue de ses avantages : la facilité de recherche, de consultation et de conservation des documents, même s'il se pose la question de la propriété juridique. La perte de temps et d'énergie occasionnée par la numérisation est rapidement amortie par la convivialité de l'utilisation, la possibilité d'enquêtes comparatives sur de larges corpus, les avantages qu'offrent les fonctions d'édition et les facilités pédagogiques que permettent les outils multimédias. [25]

6-6 Les points faibles de la numérisation :

Le coût est perçu comme le principal inconvénient perçu par de nombreuses personnes. Les oppositions que suscite ce handicap de la numérisation insistent sur trois rubriques :

- l'investissement initial onéreux : le matériel nécessaire pour effectuer la numérisation peut être coûteux et demande souvent à être utilisé par des opérateurs qualifiés pour obtenir des résultats probants ;
- les 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 exigerait en outre un investissement important ;
- les coûts de préparation : la préparation en matière de saisie numérique est d'abord d'ordre physique requise. Elle comprend aussi les opérations onéreuses ci-après : classer et indexer le contenu initial et introduire les références textuelles dans la base de données. Toutes ces tâches doivent être exécutées par un personnel spécialisé. La copie de l'information des supports existants sur de nouveaux supports impliquerait d'importantes contraintes de main-d'œuvre.

Ces problèmes de coût ont été abordés par Pierre-Yves Duchemin qui disait à ce propos que « le coût total d'une campagne de numérisation peut être très élevé, tout particulièrement si l'on a choisi des taux de résolution très importants, même si la numérisation aujourd'hui coûte deux fois moins cher qu'il y a deux ans. L'on doit garder à l'esprit que le coût de la numérisation ne comprend pas la seule numérisation proprement dite, mais couvre également les éventuels coûts de restauration, les éventuelles campagnes photographiques, le catalogage, l'indexation, les différents supports physiques, les éventuels postes de travail spécifiques munis d'écrans graphiques "haute définition", etc.». [25]

6-7 Recommandations pour la numérisation de différents supports :

Les tableaux ci-dessous résument les choix de traitement, les résolutions d'usage et les formats de conservation pour différents types de documents. Retenez que l'objectif qui sous-tend ces recommandations est

d'offrir aux organismes publics des balises visant à assurer la création de documents numérisés de qualité dont l'intégrité peut être assurée. L'application de ces balises peut varier en fonction de la qualité et des caractéristiques physiques des documents et de l'équipement utilisé.

Rappelons qu'une résolution de 300 PPP est considérée comme un minimum suffisant pour reproduire un document à sa taille originale.

[28][29]

Support d'origine	Choix de traitement	Résolution	Formats de conservation
Documents textuels sur papier	Recommandés : 8 bits ton de gris 8 bits couleurs	300 PPP	Recommandés : TIFF on compressé
Documents textuels sur microfilms et sur microfiches	Recommandé : 8 bits tons de gris	Le taux varie en fonction du facteur de réduction utilisé (exemples : 18 x, 24 x).	Recommandés : TIFF non compressé
Cartes et plans	Recommandés : 8 bits tons de gris pour les documents en noir et blanc 24 bits couleurs pour les documents en couleurs	300 PPP pour les formats plus grands que 11 x 17, 600 PPP pour les formats plus petits que 11 x 17 ou pour le géoréférencement des plans	Recommandés : TIFF non compressé GeoTIFF pour les plans géoréférencés
Épreuves photographiques	Recommandés : 8 bits tons de gris pour les épreuves en noir et blanc 24 bits couleurs pour les épreuves couleurs, incluant les épreuves sépia	300 PPP pour les formats 8 x 10 et les plus grands formats 600 PPP pour les formats plus petits que 8 x 10	Recommandé : TIFF non compressé

<p align="center">Négatifs</p>	<p>Recommandés : 8 bits tons de gris 24 bits couleurs</p>	<p>2400 PPP pour du 35 mm 1200 PPP pour le format 2 1/2 x 2 1/2 600 PPP pour les formats 4 x 5 et 5 x 7 300 PPP pour le format 8 x 10</p>	<p>Recommandé : TIFF non compressé</p>
<p align="center">Diapositives</p>	<p>Recommandés : 8 bits tons de gris pour les diapositives en noir et blanc 24 bits couleurs pour les diapositives en couleurs</p>	<p>2400 PPP pour le format 35 mm 1200 PPP pour le format 2 1/4 x 2 1/4</p>	<p>Recommandé : TIFF non compressé</p>

Tableau 2-1 : Recommandations pour la numérisation de différents supports

6-8 Les conséquences de la numérisation :

L'intégration des documents numériques dans le système d'information de la bibliothèque ou du service d'archives a un impact fort, à la fois pour leur production et pour leur consultation par le public. D'une façon générale, il est nécessaire correspondre à une base de donnée, existante ou à créer. Cela nécessite souvent l'achat de nouveaux logiciels (imageurs), l'ajout de mémoire et/ou de lecteurs de CD, un interfaçage entre le catalogue et la base d'images. Les coûts de conception, de développement, d'acquisition de matériel et de logiciel ne sont donc pas à négliger. Il faut donc envisager, chaque fois que c'est possible :

- La mutualisation des équipements
 - L'offre payante de nouveaux services (ex : fourniture de disquettes ou de CD-ROMs, cession de droits d'exploitation aux éditeurs), ce qui implique de faire voter des tarifs spécifiques par la collectivité.
- [28]

7 Conseils pour la consultation des manuscrits :

Le manuscrit est un objet unique, fragile et vénérable. Comme tout livre, il souffre d'une consultation fréquente ou sans précaution (pensez par exemple à votre dictionnaire, annuaire ou agenda). Les consignes suivantes - dont celles marquées d'un astérisque étaient déjà préconisées au Moyen Âge — s'appliquent autant au livre imprimé qu'au livre manuscrit.

1. Avant de consulter le manuscrit, utiliser le plus possible le microfilm ou les photographies pour établir votre relevé, transcription, description. Consulter également les références bibliographiques et vérifier s'il n'existe pas déjà des notices détaillées du manuscrit.

2. Consulter avec les mains propres. Certaines bibliothèques exigent qu'on porte des gants.

3. Utiliser uniquement le crayon pour prendre des notes. Les taches d'encre moderne sont indélébiles. Poser le crayon lorsqu'il faut tourner le feuillet.

4. Consulter le livre sur un pupitre ou à plat sur un bureau. Soutenir l'ouverture sans la forcer.

Manipuler avec soin les anciennes reliures : ouvrir et fermer doucement. Ne jamais poser un livre sur un autre livre ouvert.

5. Tourner les feuillets doucement un à un, entre le pouce et l'index, depuis le coin inférieur.

6. Ne jamais toucher la surface enluminée ou écrite, même avec une règle. Afin de suivre un texte ligne à ligne, utiliser une feuille de papier A4 pliée en deux.

7. Ne jamais approcher le visage trop près. Le parchemin réagit considérablement aux changements d'humidité (éternuement, postillon ou même haleine). C'est pourquoi il est déconseillé de parler près d'un manuscrit.

8. Pour relever les filigranes ou pour collationner les cahiers d'un manuscrit dont la reliure est serrée ou en mauvais état, demander l'aide

d'un conservateur. Aucune étude, édition, information n'est plus importante que la conservation du manuscrit.

9. Fermer le livre si vous faites une pause ou si vous vous absentez de votre place.

10. Déposer un exemplaire ou au moins une photocopie de votre publication à la bibliothèque qui détient le ou les manuscrits consultés par vous. [29]

8- Conclusion :

Nous avons cité dans ce chapitre quelques notions sur les manuscrits anciens, les éléments déclencheurs des dégradations 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. Enfin, nous concluons dans ce chapitre qu'afin de faire face à ces dégradations, des techniques de restauration particulières s'imposent.

Chapitre III - Aperçu sur la transformée en ondelettes

1- Introduction :

Le traitement de signal nécessite un ensemble de technique ou de recettes mais, comme toute autre discipline scientifique, le traitement de signal suppose une certaine organisation préalable de la pensée scientifique, car il consiste à dégager des informations contenues dans un signal qui se déroule au cours du temps, l'objet du traitement du signal est d'analyser avec soin, de coder efficacement, de transmettre rapidement et enfin de reconstruire soigneusement à la réception les signaux transmis de l'origine.

Un problème très général en traitement du signal (ou image) est celui posé par l'analyse, c'est-à-dire l'extraction à partir des données un ensemble d'informations pertinentes qui rendent compte tant du contenu spectral du signal que de son organisation temporelle. [30]

Parmi les méthodes de traitement des signaux, il y'a l'analyse par ondelettes qui a été introduite au début des années 1980. Il s'agit de donner une représentation des signaux permettant la mise en valeur simultanée des informations temporelles et fréquentielles (localisation temps-fréquence). C'est ainsi que le besoin des ondelettes, famille de fonctions déduites d'une même fonction, appelée ondelette mère, par translation et dilatation, s'est fait ressentir en remarquant que la transformée de Fourier (TF) qui a dominé dès le début du 19^{ième} siècle perd, lors de la projection, le contrôle de la variable temporelle et reste toujours incapable de décrire localement (en temps ou espace) le comportement fréquentiel des signaux.

L'idée originale sur laquelle sont basées les ondelettes est apparue vers les années 1940 grâce au physicien Denis Gabor, qui a introduit la notion de la TF à fenêtre glissante dans le but de remédier au problème de localisation temps-fréquence en proposant de multiplier le signal par une fonction localisée dans le temps (fenêtre), et ensuite appliquer la TF. Il s'agit de segmenter en tranches de temps fixe le signal à analyser pour appliquer par la suite la TF à chaque tranche. L'inconvénient de cette méthode est que la taille de la fenêtre reste inchangée au cours de l'analyse. Or pour étudier un signal qui a en général une allure irrégulière, il est potentiellement intéressant de pouvoir changer la taille de la fenêtre analysante en temps. C'est à partir de là que sont nées les ondelettes qui s'adaptent d'elles-mêmes à la taille et aux caractéristiques recherchées. [31]

2- La Transformée de Fourier :

La plus célèbre et la plus ancienne des transformations est la TF. Lors de cette transformation, le signal est décomposé sur un ensemble de fonctions de base qui sont le cosinus, le sinus ou l'exponentielle imaginaire.

La TF d'un signal $x(t)$ est exprimée par [32] [33]

$$X(f) = \int_{-\infty}^{+\infty} X(t)e^{-j2\pi ft} dt \quad (2.1)$$

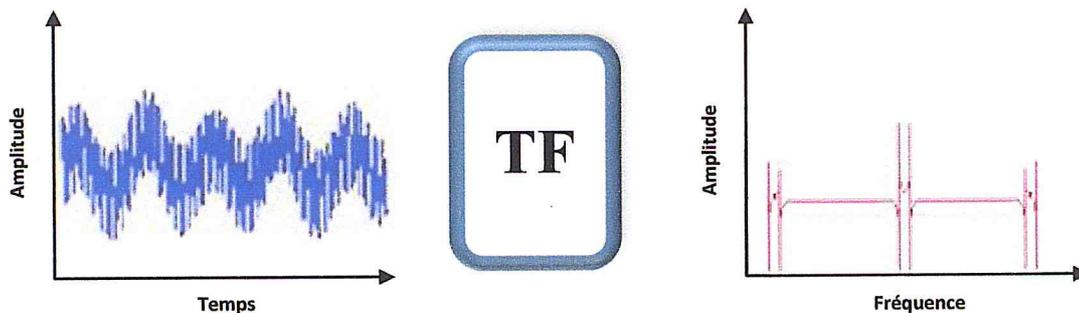


Figure 3-1 : La Transformée de Fourier.

2-1- Inconvénient de L'analyse de Fourier :

Malgré son immense succès, cette technique présente un inconvénient majeur :

- Manque de localisation temporelle. En effet, l'analyse de Fourier permet de connaître les différentes fréquences existant dans un signal, c'est à dire son spectre, mais ne permet pas de savoir à quels instants ces fréquences ont été émises. Cette analyse donne une information globale et non locale, car les fonctions d'analyse utilisées sont des sinusoides qui oscillent indéfiniment sans s'amortir. Cette perte de localité n'est pas un inconvénient pour analyser des signaux stationnaires, mais le devient pour des signaux non stationnaires.

2-2- Transformée de Fourier à Fenêtre Glissante :

Pour pallier au manque d'information sur le temps dans la TF, une nouvelle méthode d'analyse a été introduite qui utilise une *fenêtre glissante* : il s'agit de la TF à fenêtre glissante (en anglais Short Time Fourier Transform, STFT).

Cette méthode, pouvant être adaptée aux signaux non-stationnaires, est très proche de l'analyse spectrale. On définit une fenêtre qui sera utilisée comme masque sur le signal et dans laquelle on considère le signal comme localement stationnaire, puis on décale cette fenêtre le long du signal afin de l'analyser entièrement. La STFT est exprimée par :

$$X_{s,f}(s, f) = \int x(t)g^*(t - s)e^{-j2\pi ft} dt$$

$x(t)$ est le signal lui-même, $g(t)$ est la fonction fenêtre et g^* son complexe conjugué, et f joue le rôle d'une fréquence localisée autour de l'abscisse s du signal temporel

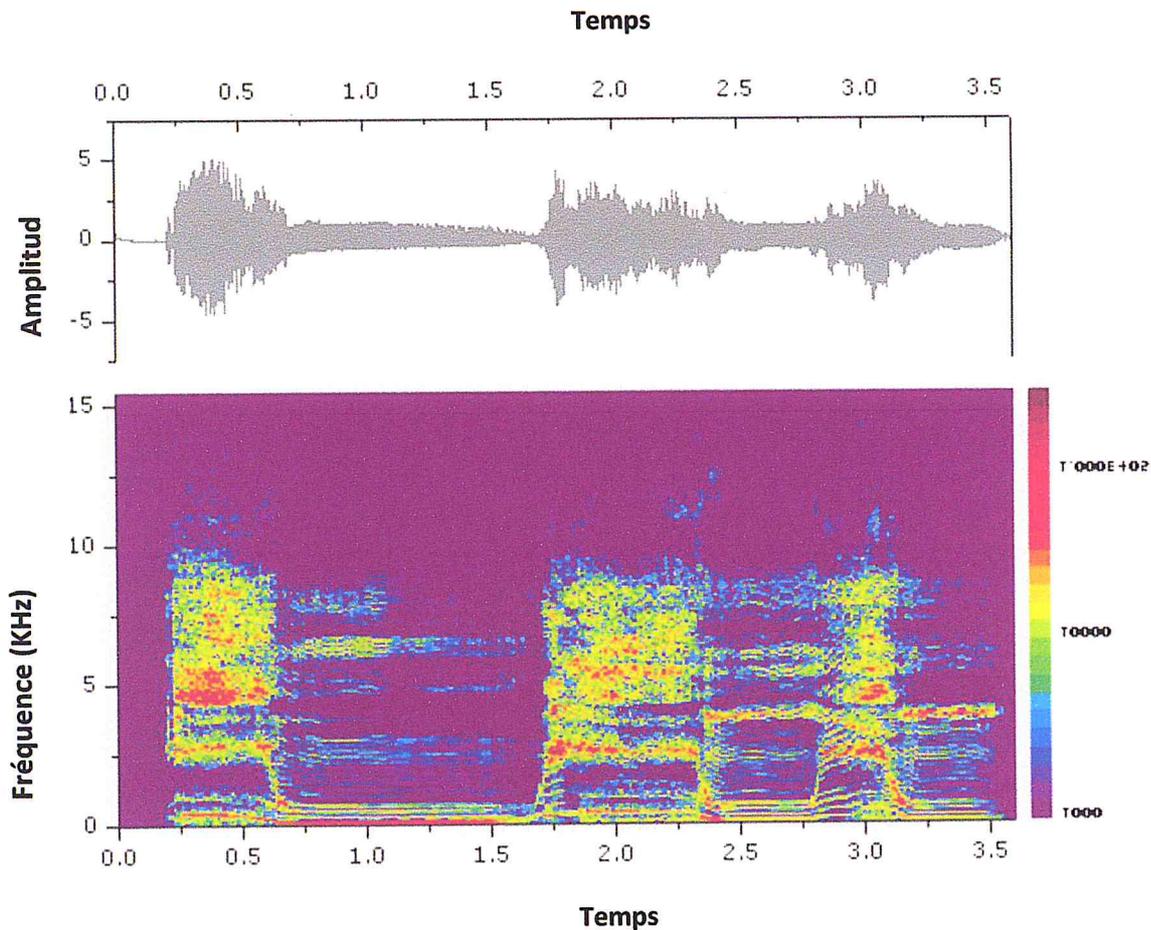


Figure 3-2 : La Transformée de Fourier à fenêtre glissante. [34]

D'après (figure 3-2), on calcule un nouveau coefficient de la STFT pour chaque valeur de f et de s .

On constate que la fenêtre g est indépendante de l'abscisse s , ce qui signifie que l'enveloppe de la fenêtre glissante sera constante : on aura donc une résolution fixe sur toute la durée du signal.

Ainsi, l'étude d'un signal avec la STFT permet d'obtenir à la fois une information sur le temps et sur la fréquence, mais la résolution d'analyse est fixée par le choix de la taille de l'enveloppe :

- Si la fenêtre est trop petite, les basses fréquences n'y seront pas contenues.
- Si la fenêtre est trop grande, l'information sur les hautes fréquences est noyée dans l'information concernant la totalité de l'intervalle contenu dans la fenêtre.

Donc la taille fixe de la fenêtre est un gros inconvénient. L'outil idéal serait une fenêtre qui s'adapte aux variations de fréquence dans le signal à analyser. Cet outil existe, il s'agit de la récente analyse en ondelettes.

1- L'Analyse en Ondelette Face à l'Analyse de Fourier

Une ondelette est une forme d'onde qui a une valeur moyenne nulle et une durée limitée. En regardant des images des ondelettes et des ondes sinusoïdales (Figure 3-3), on voit intuitivement que des signaux avec les changements pointus pourraient être mieux analysés avec une ondelette irrégulière qu'avec une sinusoïde douce. [35]

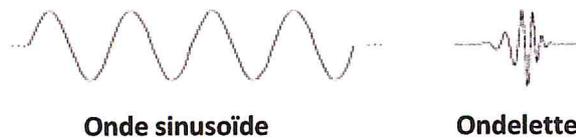


Figure 3-3 : La différence entre une onde sinusoïdale et une ondelette.

La transformée en ondelette (TO) (en anglais wavelet transform, WT) est un outil mathématique qui décompose un signal en fréquences en conservant une localisation temporelle. Le signal de départ est projeté sur un ensemble de fonctions de base qui varient en fréquence et en temps. Ces fonctions de base s'adaptent aux fréquences du signal à analyser. Les figures 3-4 et 3-5 illustrent visuellement la différence entre une analyse de Fourier et celle par ondelette, d'un signal quelconque. Pour l'analyse de Fourier, la fenêtre est de taille constante, pour l'analyse en ondelette, la fenêtre est de surface constante mais sa taille varie en fonction de la fréquence à analyser. On remarque que la largeur de la fenêtre pour en ondelette est modifiée à mesure qu'on calcule la transformée pour chacune des composantes spectrales individuellement. Ce dernier point représente la caractéristique la plus significative de la TO.

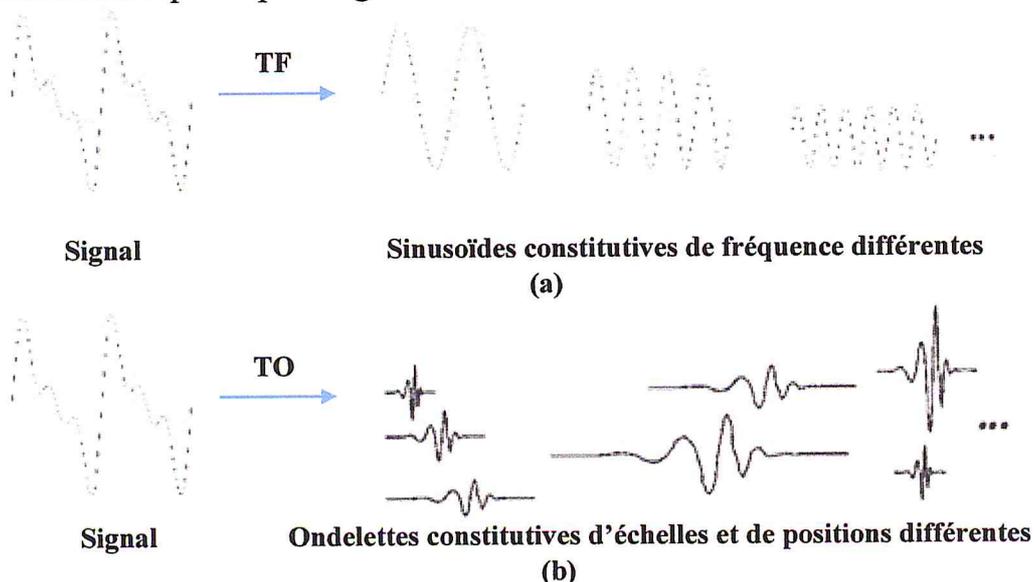


Figure 3-4 : La différence entre une analyse de Fourier (a) et une analyse en ondelette (b).

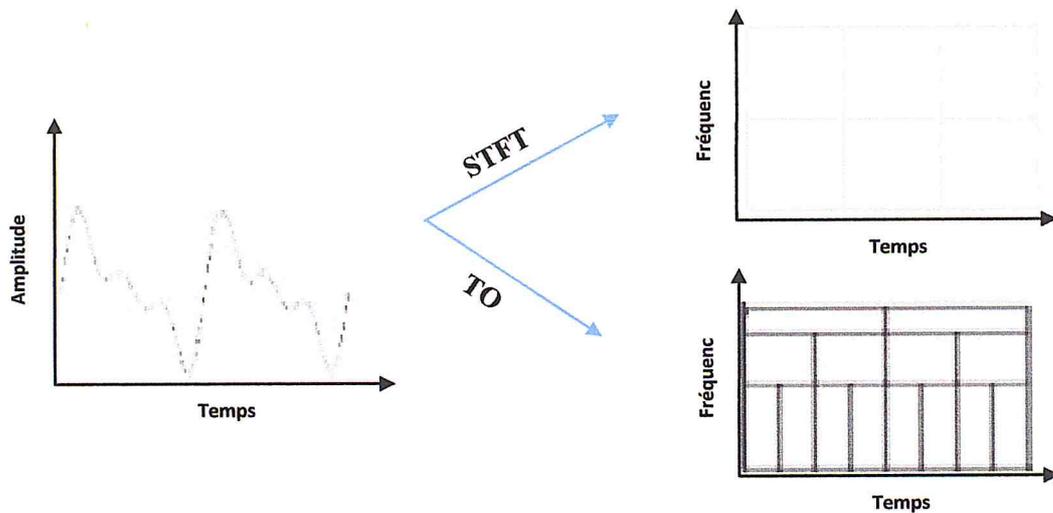


Figure 3-5 : Le pavage (découpage) de l'espace temps-fréquence pour la STFT et la TO.

2- Les Aspects Théoriques de la Transformée en Ondelette

4-1 Ondelette Mère

L'analyse en ondelette adopte une fonction prototype d'ondelettes connue sous le nom d'*ondelette mère*. On aboutira à représenter n'importe quel signal de L^2 par une base d'ondelettes qui ne sont que des versions dilatées et translatées de l'ondelette mère.

La représentation d'un signal continu $x(t)$ par une base d'ondelettes est donnée par :

$$x(t) = \sum C(s, \tau) \varphi_{s\tau}(t) \quad (2.3)$$

où $C(s, \tau)$ est le poids de la contribution de l'ondelette $\varphi_{s\tau}$ liée à l'ondelette mère par :

$$\varphi_{s\tau}(t) = \frac{1}{\sqrt{s}} \varphi\left(\frac{t-\tau}{s}\right) \quad (2.4)$$

où : τ : facteur de translation ;

s : facteur d'échelle ;

φ : ondelette mère ;

$\frac{1}{\sqrt{s}}$: Facteur de normalisation de l'énergie afin que le signal transformé ait la même énergie à toutes les échelles.

La TO a la capacité de changer la taille de la fenêtre à différentes résolutions de temps-fréquence. Grâce au changement d'échelle (Figure 3-6), les fonctions de base d'ondelette peuvent être étirées ou comprimées dans la fenêtre d'analyse. Ceci aide à isoler des discontinuités du signal parce que de courtes fonctions de base sont employées pour

obtenir l'information détaillée de fréquence tandis que de longues fonctions de base sont employées pour rechercher l'information approximative de fréquence du signal.

La translation est l'opération de glissement de l'ondelette tout au long du signal. C'est la technique similaire à celle de la STFT de GABOR.

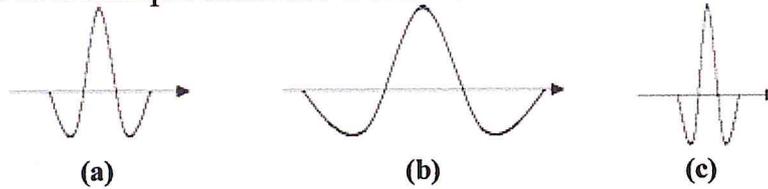


Figure 3-6 : Illustration de la variation du facteur d'échelle.

(a) L'ondelette mère. (b) L'ondelette pour $0 < s < 1$.

(c) L'ondelette pour $s > 1$.

4-2 La Transformée en Ondelette Continue :

4-2.1 Définition :

La transformée en ondelette continue (TOC) utilise des translations et des dilatations de la fonction ondelette mère durant tout l'intervalle temporel de manière continue. Elle est définie par :

$$W[x(t)] = C(s, \tau) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} x(t) \varphi\left(\frac{t-\tau}{s}\right) dt \quad (2.7)$$

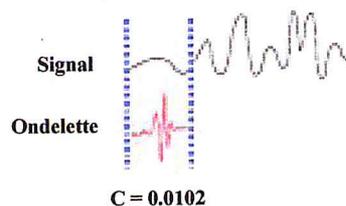
Ainsi donc, la pondération $C(s, \tau)$, appelée aussi coefficients en ondelette et définie par (2.3), n'est autre que la transformée en ondelette de la fonction $x(t)$. Cette transformation est en théorie infiniment redondante puisque l'ondelette est translatée de manière continue.

4-2.2 Étapes de la Transformée en Ondelette Continue :

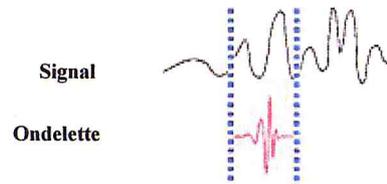
L'implémentation de la TOC peut se résumer comme suit :

1. Prendre une ondelette et la comparer à une section au début du signal original.

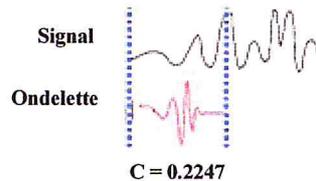
2. Calculer le coefficient $C(s, \tau)$. Par exemple :



3. Translater l'ondelette à droite et répéter l'étape 1 et 2 sur tout le signal. Par exemple :



4. Dilater ou compresser l'ondelette et répéter les étapes 1 à 3. Par exemple :



5. Répéter les étapes 1 à 4 pour toutes les échelles choisies.

4-3 La Transformée en Ondelette Discrète

Il est à peu près évident que ni la TF, ni la STFT, ni la TOC ne peuvent être manuellement calculées en utilisant les équations analytiques, le calcul intégral, etc. C'est aux ordinateurs qu'on confie la plupart des calculs aujourd'hui, c'est donc aux ordinateurs de calculer ces transformées. Il devient donc nécessaire de discrétiser les transformées. Dans le cas de la TOC, on peut utiliser la variation de l'échelle pour réduire la fréquence d'échantillonnage.

La transformée en ondelettes discrète (TOD) est produite pour surmonter le problème de redondance de la TOC. Cette redondance mobilise une grande quantité de ressources de calculs. La TOD, au contraire, fournit suffisamment d'information, tant pour l'analyse que pour la reconstruction du signal original, en un temps de calcul notablement réduit. La TOD est considérablement plus simple à implémenter que la TOC :

$$W_{a,b} = \frac{1}{\sqrt{a}} \psi \left[\frac{(t-b)}{a} \right]; \quad a, b \in R$$

Où « a » est le facteur de graduation qui permet de compresser l'ondelette dans la faible graduation et l'étendre dans la haute graduation, et « b » est le facteur de déplacement qui permet de changer le placement temporel de l'ondelette, $\psi(a,b)$ s'appelle ondelette mère

Pour beaucoup de signaux, le contenu de basse fréquence est la pièce la plus importante. Il est ce qui donne au signal son identité. Le contenu à haute fréquence, d'autre part, nous donne la saveur ou la nuance. Par exemple, avec la voix humaine si l'on enlève les composantes à haute fréquence, c'est à dire les bruits de voix différentes, on peut comprendre ce qui est dit. Cependant, si l'on enlève aussi les composantes de basse fréquence, on entendra le baragouin.

Dans la transformation en ondelette, on parle souvent d'*approximation* et de *détail*. L'approximation est à haute échelle, les composantes de basse fréquence du signal. Les détails sont à basses échelles, les composantes de hautes fréquences. Notons qu'approximation et détail émergent comme deux signaux lorsque le signal original traverse deux filtres complémentaires. Dans le cas d'une analyse multi-résolution, le processus du filtrage sera illustré dans ce qui suit.

3- L'Analyse Multi-Résolution :

Plusieurs approches d'une analyse en ondelettes discrète ont été élaborées. Par exemple celle associée à la décomposition de Littlewood-Paley, la décomposition pyramidale de Burt et Adelson, l'algorithme à trou, et plus récemment l'analyse multi-résolution [36] [37].

La théorie de l'analyse multi-résolution développée par Y. Meyer et S. Mallat, permet d'exprimer une fonction x de L^2 comme une suite d'approximations successives. Ces approximations n'ont pas la même résolution. L'analyse s'effectue alors en calculant ce qui diffère d'une résolution à l'autre, c'est-à-dire les détails qui permettent d'accéder à une représentation d'une qualité meilleure. Cette décomposition conduit à un algorithme général permettant une reconstruction qui conserve le nombre des échantillons du signal [38] [39].

Définition :

Une analyse multi-résolution dyadique de $L^2(\mathbb{R})$ est une suite croissante de sous-espaces linéaires $(V_2^j)_{j \in \mathbb{Z}}$ de L^2 , tels que le passage d'un espace à un autre soit le résultat d'un changement d'échelle.

4- Algorithme de décomposition/reconstruction de S.MALLAT :

C'est un algorithme en cascade à structure pyramidale procédant par filtrages successifs, puis sous l'échantillonnage à une cadence de 2, conformément au théorème de Shannon. [30]

6-1- Décomposition :

Schématiquement l'algorithme se présente comme suit :

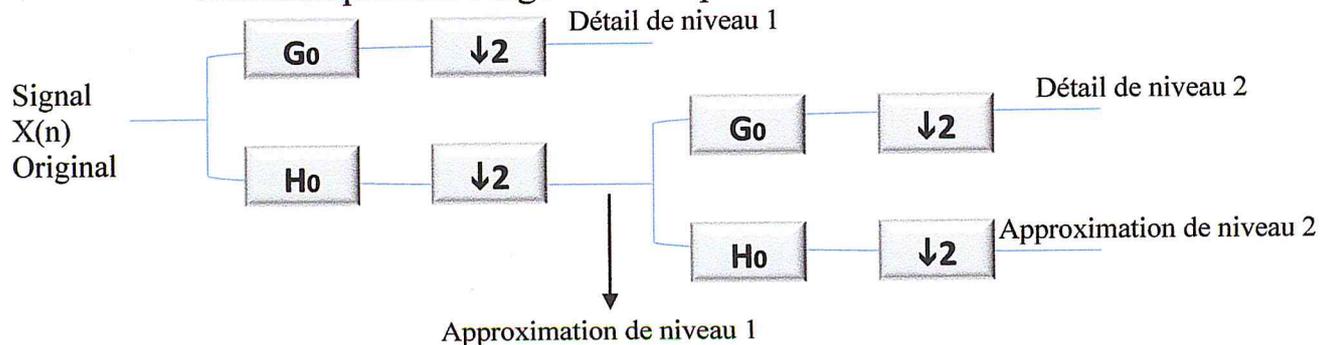


Figure 3-7 : Algorithme de décomposition

$\downarrow 2$: Décimation par deux, cette opération consiste à prendre un point (échantillon) sur deux du signal, c'est-à-dire prendre les points pairs ou impairs.

x : Convolution avec le filtre x (x représente H ou G)

On remarque que l'opération de décomposition passe par plusieurs itérations, une itération est présentée par le schéma suivant :

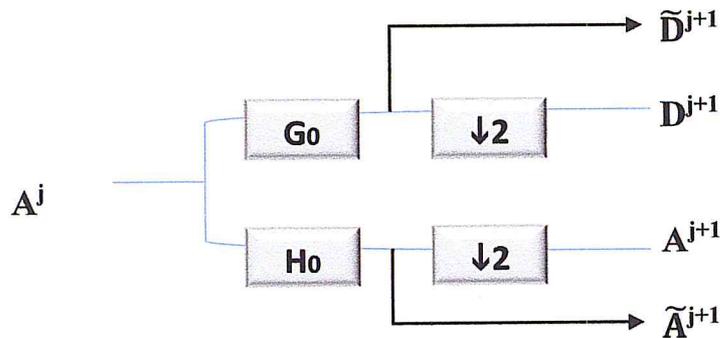


Figure 3-8 : Schéma d'une itération

A^j : Signal d'approximation de niveau j .

A^{j+1} : Signal d'approximation de niveau $j+1$.

\tilde{A}^{j+1} : Signal issu de la convolution de A^j par le filtre H_0 .

D^{j+1} : Signal de détail de niveau $j+1$.

\tilde{D}^{j+1} : Signal issu de la convolution de A^j par le filtre G_0 .

Avec :

$$\tilde{A}^{j+1}(n) = \sum h_0(k-n) A^j(k)$$

$$\tilde{D}^{j+1}(n) = \sum g_0(k-n) A^j(k)$$

On choisit les points pairs on aura :

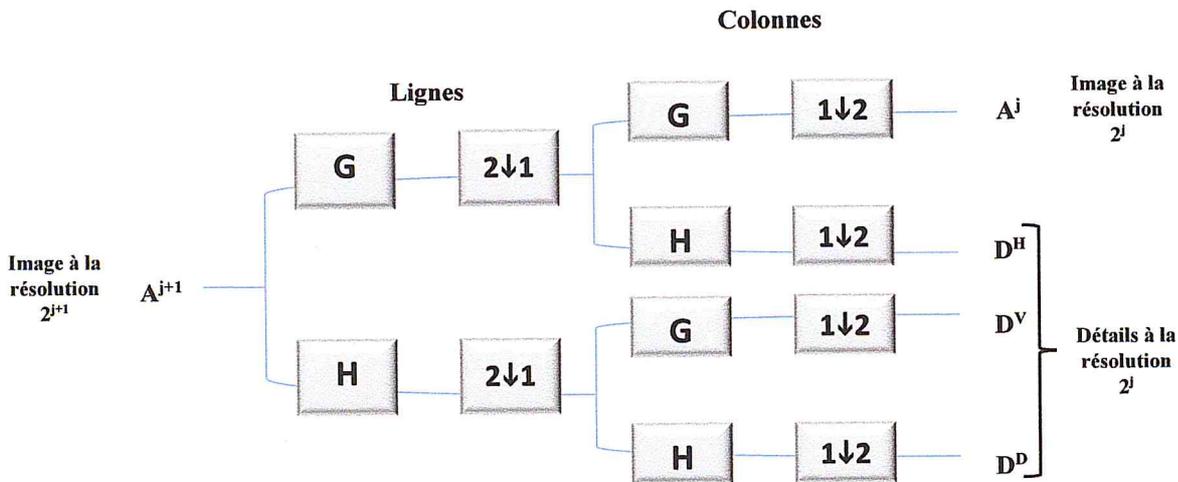
$$A^{j+1}(n) = \tilde{A}^{j+1}(2n)$$

$$D^{j+1}(n) = \tilde{D}^{j+1}(2n)$$

Les relations précédentes donnent :

$$A^{j+1}(n) = \tilde{A}^{j+1}(2n) = \sum h_0(k-2n) A^j(k)$$

$$D^{j+1}(n) = \tilde{D}^{j+1}(2n) = \sum g_0(k-2n) A^j(k)$$



X : Convolution avec filtre X.

$2 \downarrow 1$: Ne conserve qu'une colonne sur deux.

$1 \downarrow 2$: Ne conserve qu'une ligne sur deux.

Figure 3-9 : Un niveau de représentation multirésolution d'une image.

Remarque :

Si le signal est constitué de 2^p échantillons, le nombre d'itérations maximums est P. en effet, à chaque approximation le nombre de points est divisé par deux, après « p » itérations, le signal est réduit à un seul échantillon, d'où la fin de décomposition.

Puisque la décimation par 2 diminue le nombre de points, les résolutions sont donc de plus en plus faibles, le signal A^{j+1} représente une version lissée du signal A^j alors que D^{j+1} est leur différence d'information, les coefficients des signaux D^j représentent les coefficients d'ondelettes.

Si $x(n)$ (qui représente le signal original) à initialement N entrées (on peut dire N échantillons) alors le nombre total d'entrée non nulles de la décomposition est $(D^1, D^2, D^3, \dots, D^p, A^p)$.

$$\text{Soit : } \frac{N}{2} + \frac{N}{4} + \dots + \frac{N}{2^{p-1}} + \frac{N}{2^p} = N$$

On constate qu'à partir d'un vecteur initial contenant 2^p échantillon on crée deux sous vecteurs de 2^{p-1} échantillons, l'un est obtenu par filtrage passe bas (H_0) et l'autre par filtrage passe haut (G_0). A chaque niveau d'algorithme de S.MALLAT préserver le nombre d'échantillons non nuls.

Le passage d'un niveau de décomposition à un autre plus élevé ne donne pas de redondance d'information.

6-2- Reconstruction :

La reconstruction du signal original se fait à partir des signaux de détails et le dernier signal d'approximation à travers plusieurs itérations, le processus est illustré par la figure suivante :

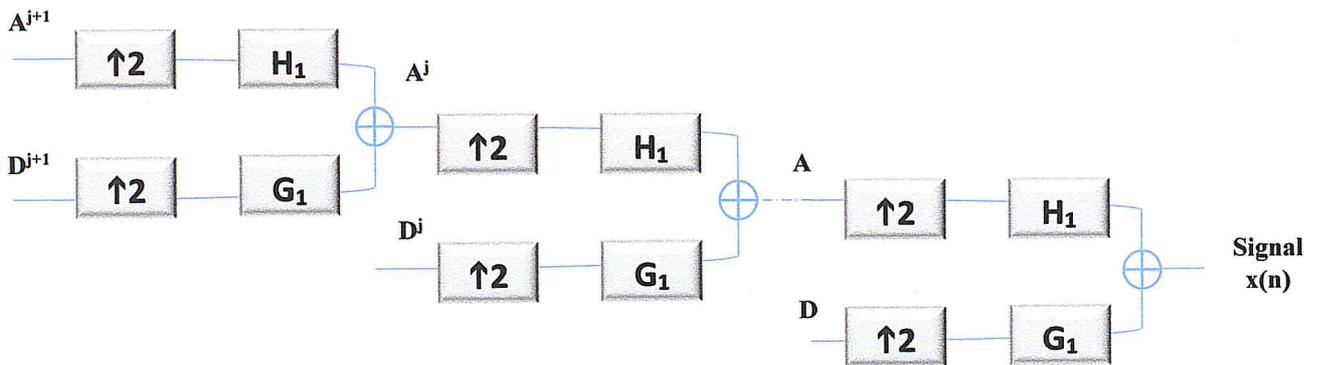


Figure 3-10 : Processus de reconstruction du signal

Avec :

↑2 : Interpolation par des zéros, qui consiste à insérer des zéros entre des points successifs

Le passage d'un niveau à un autre est représenté par le schéma ci-dessous :

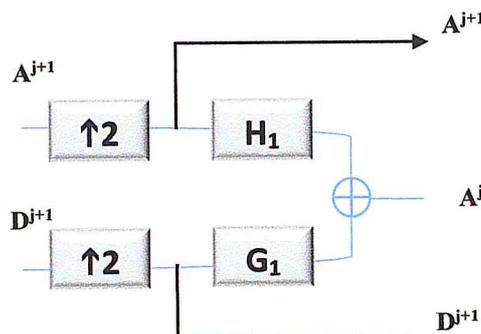


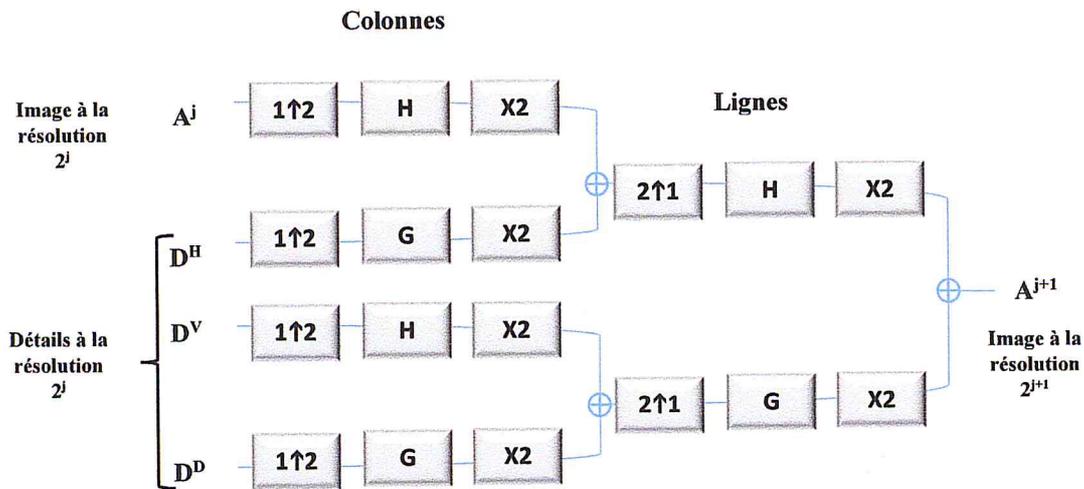
Figure 3-11 : Passage d'un niveau à un autre.

On choisit les échantillons impairs comme étant nuls :

$$\tilde{A}^{j+1}(2n+1)=0$$

Or la formule de reconstruction est :

$$A^j(n) = \sum h_1(n-k) \tilde{A}^{j+1}(k) + \sum g_1(n-k) \check{D}^{j+1}(k)$$



- X : Convolution avec filtre X.
- $1\uparrow 2$: Met une colonne de zéros entre deux colonnes.
- $2\uparrow 1$: Met une ligne de zéros entre deux lignes.
- $X2$: Multiplication par 2.

Figure 3-12 : Un niveau de reconstruction d'une image.

7- Transformée en Ondelette à deux Dimensions :

L'extension à deux dimensions de la TOD (TOD-2D) est essentielle pour la transformation des signaux à deux dimensions (2D), comme une image numérique. L'approche simple pour la mise en œuvre de la TOD-2D est d'effectuer la TOD unidimensionnelle (TOD-1D) ligne par ligne pour produire un résultat intermédiaire et ensuite effectuer la même TOD-1D en colonne sur ce résultat intermédiaire pour obtenir le résultat final (Figure 3-13). La TOD-2D s'effectue de manière naturelle par séparation des variables. Le sous espace $V_2^j(x, y)$ correspondant à la résolution j , le produit tensoriel des sous espaces $V_2^j(x)$ et $V_2^j(y)$:

$$V_2^j(x, y) = V_2^j(x) \otimes V_2^j(y) \quad (2.19)$$

La base d'interpolation est sous la forme :

$$\phi(x, y) = \phi(x) \otimes \phi(y) \quad (2.20)$$

L'information de différence entre deux approximations de l'image originale se calcule à partir de trois ondelettes bidimensionnelles séparables, définies par :

$$\varphi^H(x, y) = \varphi(x)\varphi(y)$$

$$\begin{aligned}\varphi^V(x, y) &= \varphi(x)\varphi(y) \\ \varphi^D(x, y) &= \varphi(x)\varphi(y)\end{aligned}\quad (2.21)$$

La TOD-2D fournit à chaque échelle les quatre sous-images suivantes (Fig.2.8) :

- une image de basse résolution : LL
- une image de détails horizontaux : LH
- une image de détails verticaux : HL
- une image de détails diagonaux : HH

La TOD-2D peut être interprétée comme la décomposition sur un ensemble de voies fréquentielles orientées spatialement.

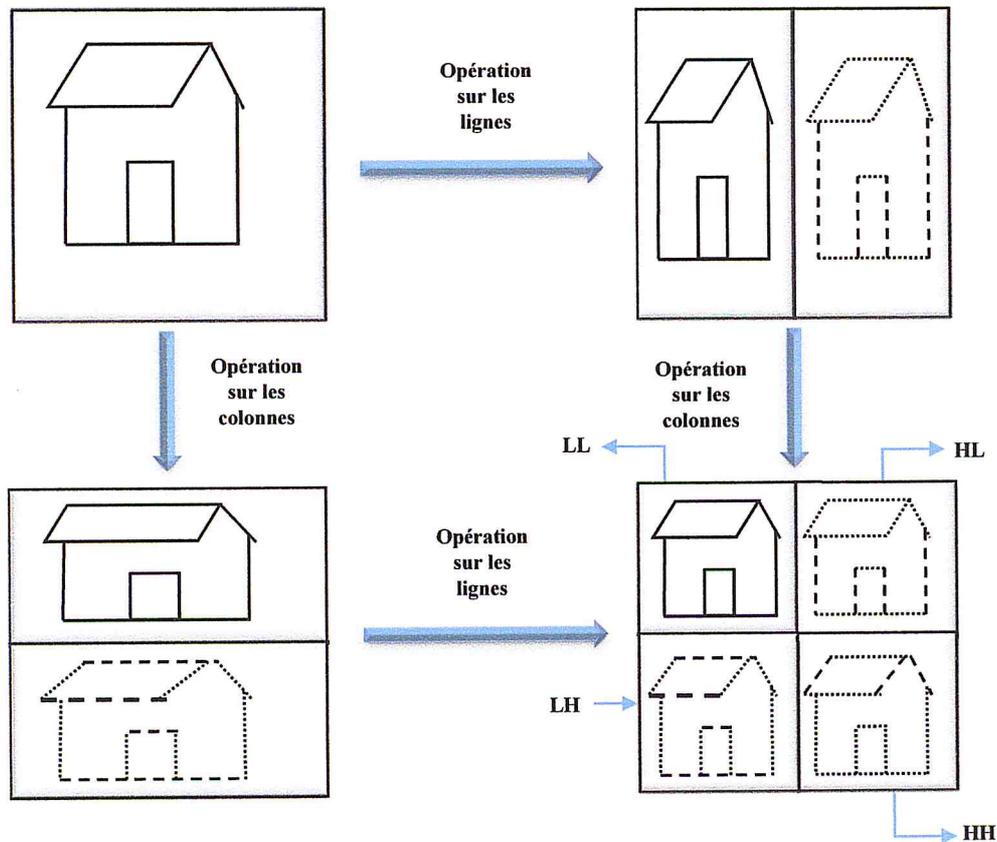
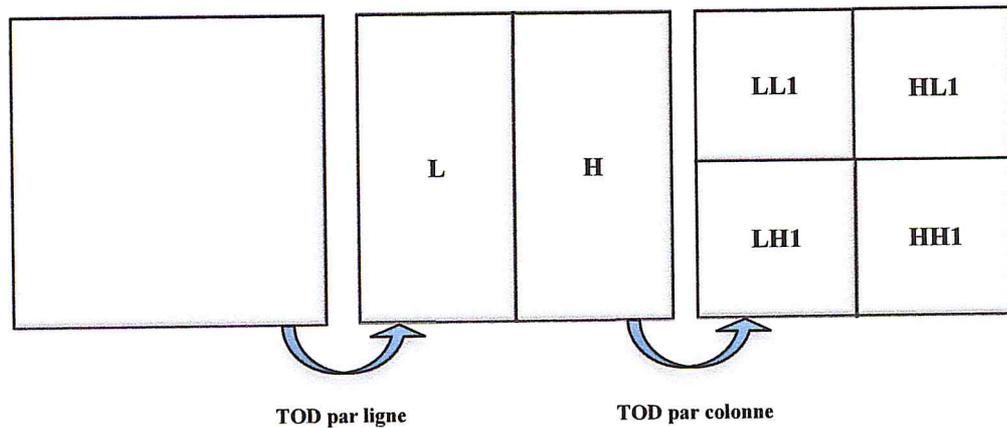


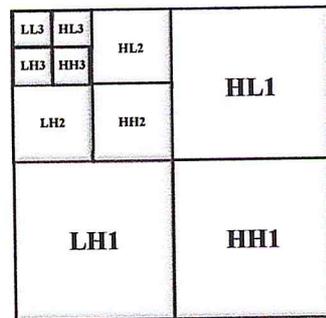
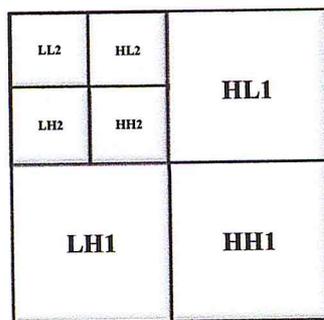
Figure 3-13 : Modèle de l'analyse multi-résolution 2D.

Lorsqu'on itère sur les signaux basse fréquence, on obtient plusieurs niveaux de décomposition en ondelette. Après le premier niveau de décomposition dans la figure 3-14, par exemple, on génère quatre sous-bandes LL1, LH1, HL1, et HH1 qui sont considérées comme sous-échantillonnées 2 : 1 à la fois horizontalement et verticalement (figure 3-14 (a)). Les trois sous-bandes LH1, HL1 et HH1 contiennent davantage d'information détaillée de fréquences orientées dans l'espace horizontal, vertical ou diagonale. La majeure partie de l'énergie dans chacune de ces trois sous-bandes est concentrée dans le voisinage des zones correspondant

à des activités de bord de l'image originale. LL1 est une version grossière (approximation) de l'entrée, et a les mêmes caractéristiques spatiales et statistiques que l'image originale. En conséquence, elle peut être décomposée en quatre sous-bandes LL2, LH2, HL2 et HH2 conformément au principe de l'analyse multi-résolution (figure 3-14 (b)). Finalement, l'image est décomposée en 10 sous-bandes LL3, LH3, HL3, HH3, LH2, HL2, HH2, LH1, HL1 et HH1 après trois niveaux de décomposition multi-résolution pyramidale, comme le montre la figure 3-14 (c).



(a) Premier niveau de décomposition



(b) Deuxième niveau de décomposition (c) Troisième niveau de décomposition

Figure 3-14 : Représentation de la TOD d'une image.

Similairement à la TOD-1D, les sous-bandes approximation et détails sont obtenues par convolution avec les filtres h et g telles que :

Le passage de la résolution $j+1$ à la résolution j se fait par :

$$A_2^j(k_x, k_y) = \sum_{l_x=-\infty}^{+\infty} \sum_{l_y=-\infty}^{+\infty} \bar{h}(2k_x - l_x) \bar{h}(2k_y - l_y) A_2^{j+1}(l_x - l_y) \quad (2.22)$$

Le signal détail appartient au sous-espace $O_2^j(x, y)$ complémentaire de $V_2^j(x, y)$ dans $V_2^{j+1}(x, y)$ et est contenu dans les trois sous images

suivantes, correspondant respectivement aux ondelettes horizontale, verticale et diagonale.

$$D_j^H(k_x, k_y) = \sum_{l_x=-\infty}^{+\infty} \sum_{l_y=-\infty}^{+\infty} \bar{h}(2k_x - l_x) \bar{g}(2k_y - l_y) A_2^{j+1}(l_x - l_y) \quad (2.23)$$

$$D_j^V(k_x, k_y) = \sum_{l_x=-\infty}^{+\infty} \sum_{l_y=-\infty}^{+\infty} \bar{g}(2k_x - l_x) \bar{h}(2k_y - l_y) A_2^{j+1}(l_x - l_y) \quad (2.24)$$

$$D_j^D(k_x, k_y) = \sum_{l_x=-\infty}^{+\infty} \sum_{l_y=-\infty}^{+\infty} \bar{g}(2k_x - l_x) \bar{g}(2k_y - l_y) A_2^{j+1}(l_x - l_y) \quad (2.25)$$

La restauration d'effectuee par :

$$\begin{aligned} A_2^{j+1}(k_x, k_y) = & 4 \sum_{l_x=-\infty}^{+\infty} \sum_{l_y=-\infty}^{+\infty} A_2^j(l_x - l_y) h(l_x - 2k_x) h(l_y - 2k_y) \\ & + D_j^H(l_x - l_y) h(l_x - 2k_x) g(l_y - 2k_y) \\ & + D_j^V(l_x - l_y) g(l_x - 2k_x) h(l_y - 2k_y) \\ & + D_j^D(l_x - l_y) g(l_x - 2k_x) g(l_y - 2k_y) \end{aligned} \quad (2.26)$$

Les procédures de filtrage pour la décomposition en ondelette 2D et reconstruction sont représentées sur les figures 2.10 et 2.11 respectivement.

8- Les Familles d'Ondelettes

Il existe une infinité de fonctions d'ondelettes parce que toute fonction oscillante localisée est une ondelette mère possible. Toutefois, elles ne possèdent pas toutes des propriétés intéressantes. Aussi, de nombreux spécialistes des ondelettes ont construit des familles d'ondelettes possédant certaines propriétés remarquables. Parmi ces familles, les ondelettes de Haar sont les plus simples, mais elles ne sont pas bien localisées. Ingrid Daubechies a construit des ondelettes à support compact qui permettent d'utiliser des filtres de taille finie.

Une autre famille d'ondelettes est la famille des Splines dont la réponse fréquentielle est bien localisée. Les différentes familles d'ondelettes sont utilisées selon leurs propriétés en fonction du problème à résoudre. Voici quelques familles d'ondelettes :

8-1- L'ondelette de Haar :

La première transformation en ondelette discrète a été inventé par le mathématicien hongrois Alfréd Haar en 1909. Pour une entrée représentée par une liste de numéros 2^n , la transformée en ondelettes de Haar peut être considéré simplement à coupler les valeurs d'entrée, le stockage de la différence et la somme passage. Ce processus est répété de façon récursive,

couplant les sommes pour fournir la prochaine échelle : pour aboutir finalement à 2^{n-1} différences et une somme finale. [40]

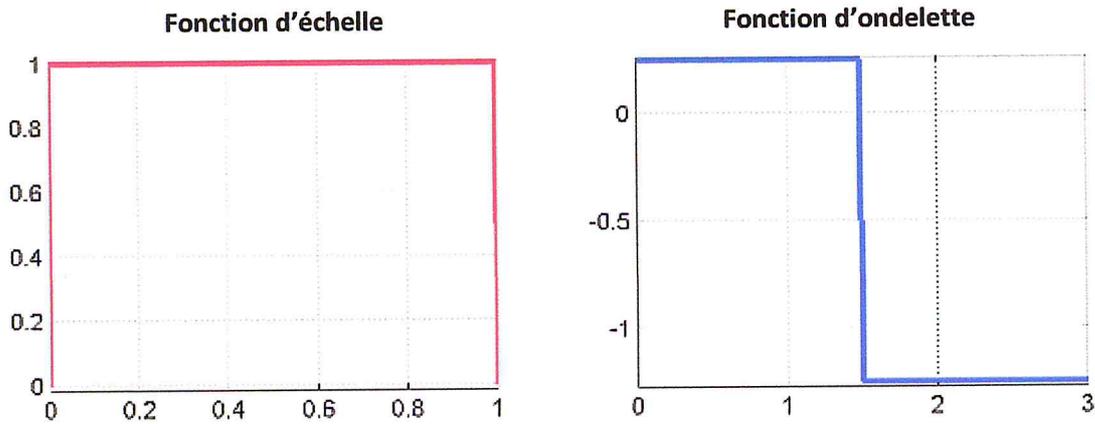


Figure 3-15: L'ondelette de Haar. [41]

Supposons N valeurs :

$$X = (x_1, x_2, \dots, x_N)$$

où N est pair.

On prend par couple la moyenne des nombres :

$$s_k = (x_{2k} + x_{2k+1}) / 2 \text{ pour } k = 0, \dots, N / 2 - 1$$

Exemple 1 : [42]

Transformée directe :

Pour réaliser la première itération, il faut calculer la moyenne (fonction d'échelle) et la différence (fonction d'ondelette) de chaque paire de donnée :

Données	7	1	6	6	3	-5	4	2
Moyenne	$(7+1) / 2 = 4$		$(6+6) / 2 = 6$		$(3+(-5)) / 2 = -1$		$(4+2)/2 = 3$	
Différence	$7-4 = 4-1 = 3$		$6-6 = 6-6 = 0$		$3-(-1)=-1-(-5)=4$		$4-3=3-2=1$	
1^{ère} Itération	Signal comprimé = moyenne				Détail du signal = différence			
	4	6	-1	3	3	0	4	1

Tableau 3-1 : 1ere itération de la multirésolution

La deuxième itération ne concerne plus que les 4 premières valeurs :

Données	4	6	-1	3
Moyenne	$(4+6) / 2 = 5$		$(-1+3) / 2 = 1$	
Différence	$4-5 = 5-6 = -1$		$-1-1 = 1-3 = -2$	
1^{ère} Itération	Signal comprimé		Détails du signal	
	5	1	-1	-2

Tableau 3-2 : 2eme itération de la multirésolution

Cette deuxième itération donne les valeurs suivantes :

$$5, 1, -1, -2, 3, 0, 4, 1$$

La troisième itération se poursuit sur les 2 premières valeurs (5,1) :

Moyenne = $(5 + 1) / 2 = 3$;

Différence = $(5 - 3) = (3 - 1) = 2$.

Cette troisième itération donne les valeurs suivantes :

$$3, 2, -1, -2, 3, 0, 4, 1$$

Deux mesures étant minimum possible, nous ne pouvons pas aller plus en avant dans ce processus, ce qui marque la fin de la multirésolution. On remarque que la première valeur obtenue par la transformée est égale à la moyenne des données initiales :

$$(7 + 1 + 6 + 6 + 3 - 5 + 4 + 2) / 8 = 3$$

Transformée inverse :

le calcul de la transformée inverse commence sur les deux premières valeurs (3 et 2) :

$$l = 3 + 2 = 5$$

$$r = 3 - 2 = 1$$

Le premier traitement donne les valeurs suivantes :

$$5, 1, -1, -2, 3, 0, 4, 1$$

Le traitement se poursuit sur les 4 premières valeurs :

$$5, 1, -1, -2$$

$$l1 = 5 + (-1) = 4$$

$$r1 = 5 - (-1) = 6$$

$$l2 = 1 + (-2) = -1$$

$$r2 = 1 - (-2) = 3$$

Le second traitement donne les valeurs suivantes :

$$4, 6, -1, 3, 3, 0, 4, 1$$

Le traitement se poursuit sur les 8 premières valeurs :

$$4, 6, -1, 3, 3, 0, 4, 1$$

$$l1 = 4 + 3 = 7$$

$$r1 = 4 - 3 = 1$$

$$l2 = 6 + 0 = 6$$

$$r2 = 6 - 0 = 6$$

$$l3 = -1 + 4 = 3$$

$$r3 = -1 - 4 = -5$$

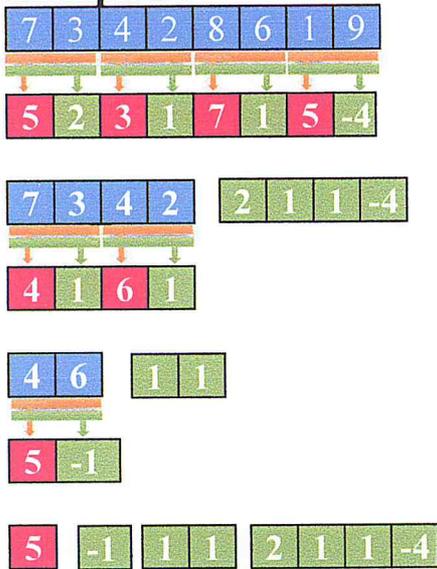
$$l4 = 3 + 1 = 4$$

$$r4 = 3 - 1 = 2$$

Le troisième et dernier traitement donne les valeurs suivantes :

$$7, 1, 6, 6, 3, -5, 4, 2 \text{ qui sont les valeurs initiales.}$$

Exemple 2 :



8-2- L'ondelette de Daubechies :

L'algorithme de transformée d'ondelette de Daubechies est composé d'une fonction d'ondelette et d'échelle. Les coefficients de la fonction d'échelle sont notés "h" et les coefficients de la fonction d'ondelette sont notés "g". L'application de la fonction d'échelle à un signal "y" donne un coefficient de type "e" et l'application de la fonction d'ondelette à ce même signal "y" donne un coefficient de type "o". [43]

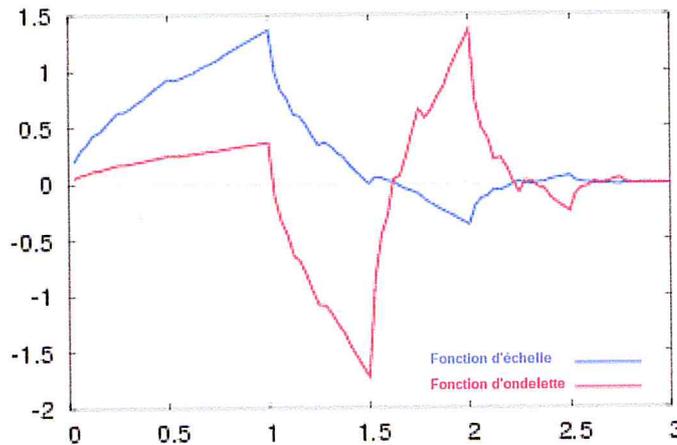


Figure 3-16 : L'ondelette de Daubechies 4.

La transformée est appliquée comme une multiplication matricielle du type (pour Daubechies 4) :

$$\begin{array}{|c|c|c|c|c|c|c|c|} \hline
 h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 \\ \hline
 g_0 & g_1 & g_2 & g_3 & 0 & 0 & 0 & 0 \\ \hline
 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 \\ \hline
 \end{array}
 \begin{array}{|c|} \hline
 y_1 \\ \hline
 y_2 \\ \hline
 y_3 \\ \hline
 \end{array}
 \begin{array}{|c|} \hline
 e_1 \\ \hline
 o_1 \\ \hline
 e_2 \\ \hline
 \end{array}
 \begin{array}{|c|} \hline
 e_1 \\ \hline
 e_2 \\ \hline
 e_3 \\ \hline
 \end{array}$$

$$\begin{array}{l}
 | 0 0 g_0 g_1 g_2 g_3 0 0 || y_4 | = | o_2 | \text{ On permute } | e_4 | \\
 | 0 0 0 0 h_0 h_1 h_2 h_3 || y_5 | \quad | e_3 | \text{ pour avoir } | o_1 | \\
 | 0 0 0 0 g_0 g_1 g_2 g_3 || y_6 | \quad | o_3 | \quad | o_2 | \\
 | h_2 h_3 0 0 0 0 h_0 h_1 || y_7 | \quad | e_4 | \quad | o_3 | \\
 | g_2 g_3 0 0 0 0 g_0 g_1 || y_8 | \quad | o_4 | \quad | o_4 |
 \end{array}$$

Tel que :

$$h_0 = \frac{1 + \sqrt{3}}{4\sqrt{2}}, h_1 = \frac{3 + \sqrt{3}}{4\sqrt{2}}, h_2 = \frac{3 - \sqrt{3}}{4\sqrt{2}} \text{ et } h_3 = \frac{1 - \sqrt{3}}{4\sqrt{2}}$$

En réappliquant le traitement sur les 4 premières valeurs après permutation :

$$\begin{array}{l}
 | h_0 h_1 h_2 h_3 || e_1 | \quad | e_1' | \quad | e_1' | \\
 | g_0 g_1 g_2 g_3 || e_2 | \quad | o_1' | \quad \text{On permute } | e_2' | \\
 | h_2 h_3 h_0 h_1 || e_3 | = | e_2' | \quad \text{pour avoir } | o_1' | \\
 | g_2 g_3 g_0 g_1 || e_4 | \quad | o_2' | \quad | o_2' |
 \end{array}$$

On obtient ainsi la multirésolution, dans ce cas le traitement s'arrête à ce stade, car il faut 4 valeurs pour le calcul.

Les coefficients de la transformée de Daubechies sont calculés (entre autre) pour que la matrice transposée soit la matrice inverse :

$$\begin{array}{l}
 | h_0 h_1 h_2 h_3 || h_0 g_0 h_2 g_2 | \quad | 1 0 0 0 | \\
 | g_0 g_1 g_2 g_3 || h_1 g_1 h_3 g_3 | = | 0 1 0 0 | \\
 | h_2 h_3 h_0 h_1 || h_2 g_2 h_0 g_0 | \quad | 0 0 1 0 | \\
 | g_2 g_3 g_0 g_1 || h_3 g_3 h_1 g_1 | \quad | 0 0 0 1 |
 \end{array}$$

Donc, à chaque niveau, la matrice de la transformée inverse est la matrice transposée de la transformée :

$$\begin{array}{l}
 | h_0 g_0 h_2 g_2 || e_1' | \quad | e_1 | \\
 | h_1 g_1 h_3 g_3 || o_1' | = | e_2 | \\
 | h_2 g_2 h_0 g_0 || e_2' | \quad | e_3 | \\
 | h_3 g_3 h_1 g_1 || o_1' | \quad | e_4 |
 \end{array}$$

$$\begin{array}{l}
 | h_0 g_0 0 0 0 h_2 g_2 || e_1 | \quad | y_1 | \\
 | h_1 g_1 0 0 0 h_3 g_3 || o_1 | \quad | y_2 | \\
 | h_2 g_2 h_0 g_0 0 0 0 || e_2 | \quad | y_3 | \\
 | h_3 g_3 h_1 g_1 0 0 0 || o_2 | = | y_4 | \\
 | 0 0 h_2 g_2 h_0 g_0 0 0 || e_3 | \quad | y_5 | \\
 | 0 0 h_3 g_3 h_1 g_1 0 0 || o_3 | \quad | y_6 |
 \end{array}$$

$$\begin{array}{|c|c|c|c|c|c|c|c|} \hline | 0 & 0 & 0 & 0 & h_2 & g_2 & h_0 & g_0 \\ \hline | e_4 & & & & & & & \\ \hline | y_7 & & & & & & & \\ \hline | 0 & 0 & 0 & 0 & h_3 & g_3 & h_1 & g_1 \\ \hline | o_4 & & & & & & & \\ \hline | y_8 & & & & & & & \\ \hline \end{array}$$

qui permet de retrouver le signal de départ.
Quelques types d'ondelettes de Daubechies :

D2 (Haar)	D4	D6	D8	D10	D12	D14	D16
1	0.6830127	0.47046721	0.32580343	0.22641898	0.15774243	0.11009943	0.07695562
1	1.1830127	1.14111692	1.01094572	0.85394354	0.69950381	0.56079128	0.44246725
	0.3169873	0.650365	0.8922014	1.02432694	1.06226376	1.03114849	0.95548615
	-0.1830127	-0.19093442	-0.03957503	0.19576696	0.44583132	0.66437248	0.82781653
		-0.12083221	-0.26450717	-0.34265671	-0.31998660	-0.20351382	-0.02238574
		0.0498175	0.0436163	-0.04560113	-0.18351806	-0.31683501	-0.40165863
			0.0465036	0.10970265	0.13788809	0.1008467	6.68194092 ^{e-4}
			-0.01498699	-0.00882680	0.03892321	0.11400345	0.18207636
				-0.01779187	-0.04466375	-0.05378245	-0.02456390
				4.71742793 ^{e-3}	7.83251152 ^{e-4}	-0.02343994	-0.06235021
					6.75606236 ^{e-3}	0.01774979	0.01977216
					-1.52353381 ^{e-3}	6.07514995 ^{e-4}	0.01236884
						-2.54790472 ^{e-3}	-6.88771926 ^{e-3}
						5.00226853 ^{e-4}	-5.54004549 ^{e-4}
							9.55229711 ^{e-4}
							-1.66137261 ^{e-4}

Tableau 3-3 : Types d'ondelettes de Daubechies. [44]

8-3- L'ondelette Biorthogonale :

Les ondelettes Biorthogonales sont des familles d'ondelettes symétriques à support compact. La symétrie des coefficients de filtrage est souvent souhaitable, car elle conduit à une phase linéaire de la fonction de transfert. Dans le cas bi-orthogonal, plutôt que d'avoir une mise à l'échelle et de la fonction d'ondelettes, il existe deux fonctions d'échelle qui peuvent générer différents analyses multirésolution, et par conséquent deux différentes fonctions d'ondelettes. Les fonctions doubles d'échelle et les fonctions d'ondelettes ont les propriétés suivantes :

- Elles sont égales à zéro à l'extérieur d'un segment.
- Les algorithmes de calcul sont maintenus, et donc très simple.
- Les filtres associés sont symétriques.
- Les fonctions utilisées dans les calculs sont plus faciles à construire numériquement que ceux utilisés dans les ondelettes de Daubechies.

[45]

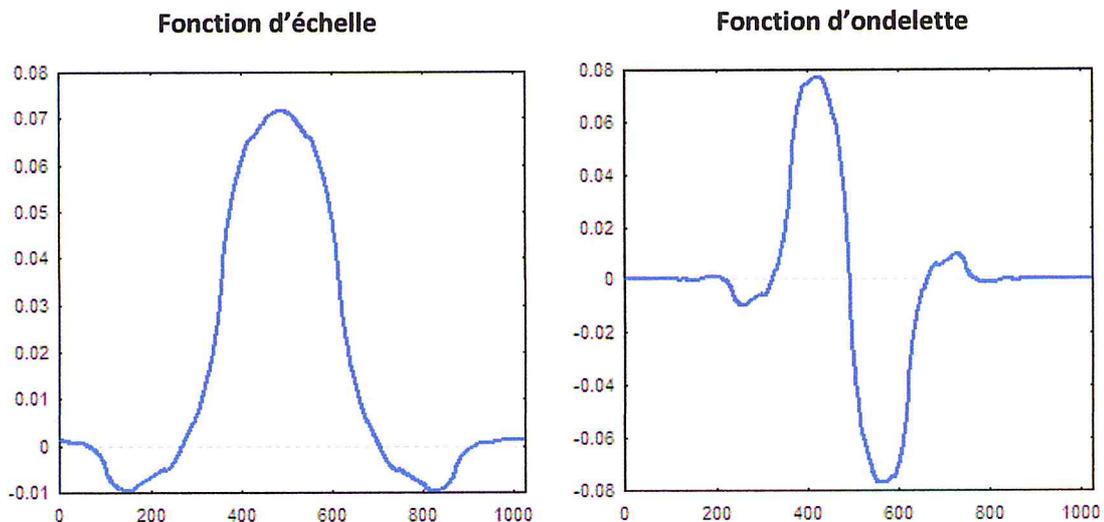


Figure 3-17 : L'ondelette Biorthogonale 1.3. [46]

9- Conclusion :

La transformée en ondelettes présente de nombreux avantages dans le domaine du traitement du signal et de l'image. L'analyse par ondelette ne se limite plus à l'image telle qu'elle nous apparaît, mais permet l'étude des objets présents dans l'image à différentes échelles. Elle permet de réduire la redondance pour améliorer la compression d'une image, elle peut également extraire les informations importantes (texture, contours, etc.) et aussi de réduire le bruit contenu dans l'image. C'est un outil puissant de transformation du signal qui permet de préparer le signal afin de faciliter le traitement envisagé.

Chapitre IV - Démarches de restauration existantes

1- Introduction :

Bien que la restauration des images naturelles soit largement traitée dans la littérature scientifique, la restauration des images de documents est à considérer comme un domaine de recherche émergent ; de nouveaux problèmes vont se rajouter aux problèmes classiques présents sur les images naturelles. Par problèmes classiques, nous faisons référence aux dégradations apparues lors de la capture, de l'acquisition, de la transmission, du changement de format ou du stockage de l'image.

Dans le cas des documents anciens, d'autres types de dégradations apparaissent. Ces dégradations sont intrinsèquement liées au support physique du document et aux conditions de sa conservation.

En effet, ce patrimoine écrit dont la valeur historique est souvent inestimable n'échappe pas, en dépit des précautions prises, à l'usure du temps (vieillesse, modification des couleurs, porosité du papier, acidité de l'encre, humidité et variation de la température de stockage, présence d'insectes, sinistres de toutes sortes.) qui détériore énormément le support.

Ces dégradations qui s'ajoutent à celles liées à la procédure de numérisation (variation d'éclairage, résolution, compression.) vont contribuer à la dégradation de la copie numérique.

Toutes ces dégradations peuvent se traduire par des modifications de couleurs, des trous dans le support (associés à des caractères disparus), des taches recouvrant des parties plus ou moins importantes du document, la superposition du recto et du verso en transparence, des caractères devenant flous et courbure de la page.

Compte tenu de ces différents nouveaux problèmes, la restauration (que l'on qualifiera d'informatique ou virtuelle) des images des documents anciens ne se limite pas à une simple transcription des outils classiques de traitement et d'analyse, et encore moins à une application immédiate des méthodes de restauration déjà définies pour les images naturelles dégradées.

Mais quelle que soit la nature du document (image naturelle ou image de document), l'objet principal de la restauration reste le même : il consiste à atténuer (voire à supprimer dans les cas les plus favorables) les dégradations présentes sur un document image. En outre, la restauration est un prérequis nécessaire pour un grand nombre d'usages d'images de document. Non seulement, elle permet d'améliorer leur qualité visuelle, mais aussi elle rend possible ou optimise les performances des outils

informatiques tels que les logiciels de reconnaissance optique des caractères (OCR). [47]

2- Type de restauration :

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-désires, 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. [4]

3- Les méthodes de restauration :

Parmi les travaux de recherche sur la restauration des images de documents, nous trouvons beaucoup d'articles qui abordent la séparation entre le recto et le verso.

3-1- Méthodes de restauration supervisée « Non-Aveugle » :

3-1-1-Séparation recto/verso par recalage :

Le recalage a été plus particulièrement appliqué pour la suppression des marques en transparence avec la présence de deux faces de document (recto/verso associé). Le principe consiste à mettre en correspondance les deux faces d'une même page pour extraire le texte original associé à chaque face. L'approche de Sharma applique un filtre linéaire adaptatif sur les deux faces du document simultanément. Le cas de passage en transparence de l'écrit n'a, semble-t-il, pas été traité et ils donnent dans la figure 4-1 le seul résultat publié par l'auteur ; il faut remarquer que les deux parties à séparer sont de nature différentes (texte et image). [48]

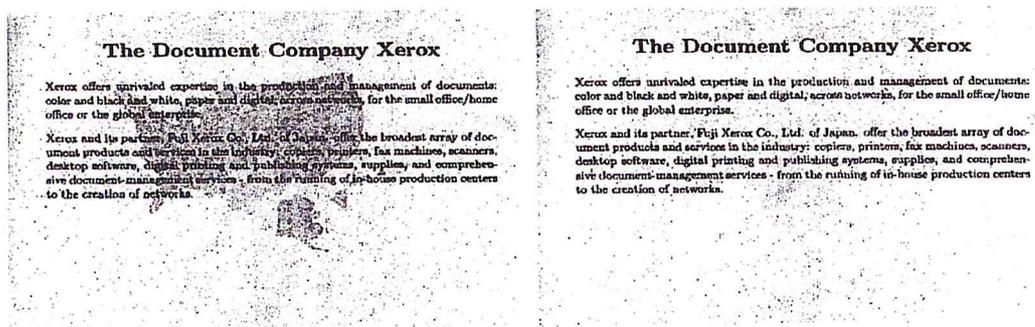
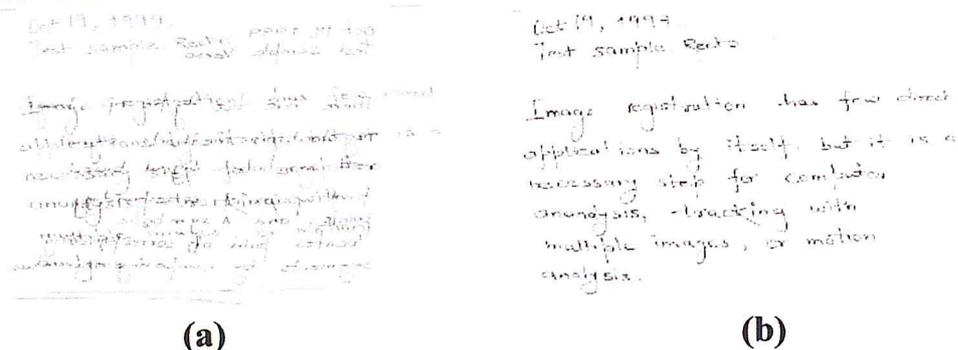


Figure 4-1 : Suppression des taches en transparence par recalage.

Une autre 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 plus élevée que celle du verso, sa valeur est conservée. La figure 4-2 montre un exemple d'application de cette méthode.



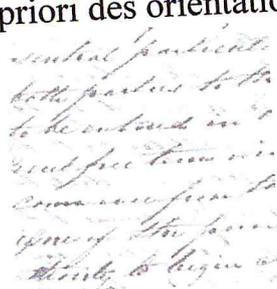
(a) Le recto d'un document de test avec une fuite significative du verso, (b) Le recto du document restauré après application de la méthode de Dubois et A. Pathak

Figure 4-2 : Suppression des taches en transparence par recalage

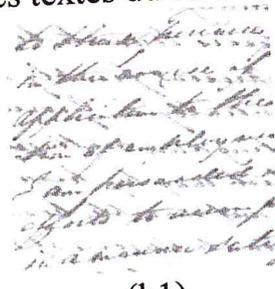
Si l'image du recto est géométriquement déformée par rapport à celle du verso, la méthode ne peut pas mettre une mise en correspondance parfaite des deux faces. [49]

3-1-2-Séparation recto/verso par analyse de l'orientation :

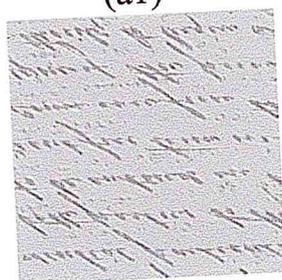
Certains documents sont caractérisés par des styles d'écriture très orientés. L'analyse de l'orientation s'avère donc efficace pour distinguer le texte du recto de celui du verso. Q. Wang et al. Proposent une analyse de l'orientation à base de la transformée en ondelettes. Le principe consiste à définir une transformée en ondelettes 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 (Figure 4-3). On classe donc cette approche dans les méthodes de restauration supervisées puisqu'elle nécessite la connaissance a priori des orientations des textes du recto et du verso. [50]



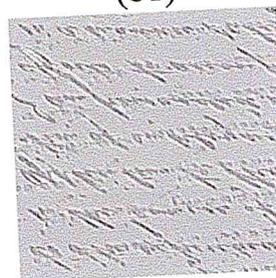
(a1)



(b1)



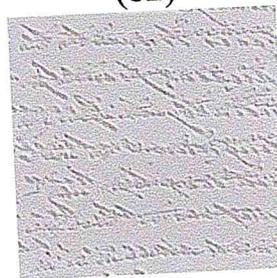
(a2)



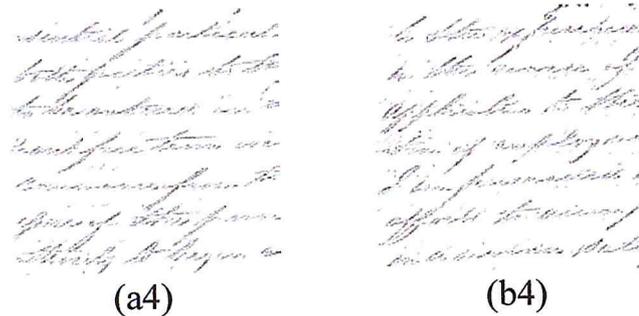
(b2)



(a3)



(b3)

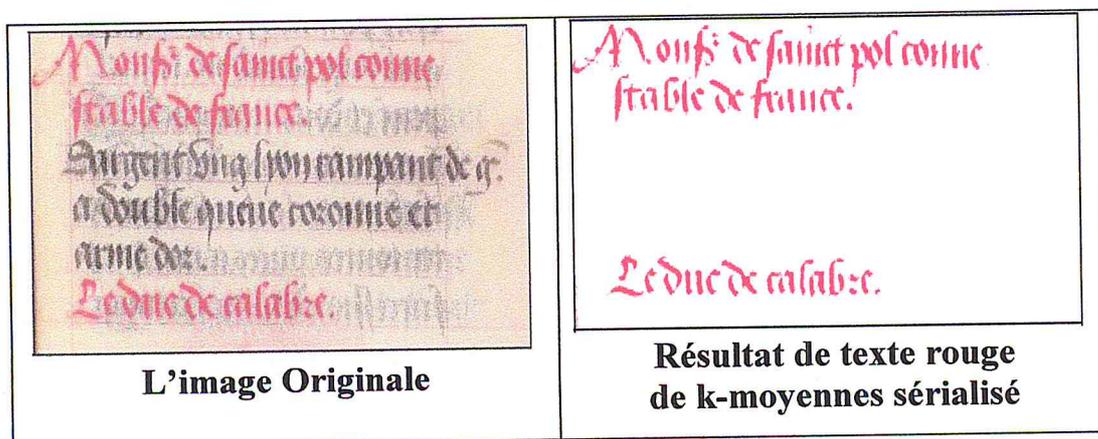


(a1), (b1) : image de test,
 (a2), (b2) : image associée à l'orientation 135°,
 (a3), (b3) : image associée à l'orientation 45°,
 (a4), (b4) : image restaurée de (a1), (b1) respectivement.
Figure 4-3 : Suppression des taches en transparence par analyse de l'orientation

3-1-3-Séparation recto/verso par analyse de la couleur :

Y. Leydier, F. Le Bourgeois et H. Emptoz présentent un algorithme de segmentation adaptée pour l'analyse de documents images en fonction de la sérialisation de l'algorithme de nuées dynamiques (k-moyennes) appliqué séquentiellement le long de chaque ligne de balayage. Ils proposent également de représenter chaque classe logique avec plusieurs clusters dans l'espace des caractéristiques.

L'initialisation des centres des pôles n'est pas une étape critique du fait de l'adaptative de l'algorithme, et le nombre de grappes par classe et la taille de la fenêtre ont une influence limitée sur les résultats de segmentation. Les résultats sont très bons sur toutes les anciennes images de manuscrits qu'ils ont traitées avec des paramètres optimaux (nombre et position des grappes, la taille de la fenêtre, les caractéristiques des choix). Ils prévoient de simplifier l'utilisation de cet algorithme et d'aider à l'utilisateur de définir automatiquement certains paramètres tels que le nombre de grappes pour chaque catégorie et des échantillons de couleur pour chaque classe. [51]



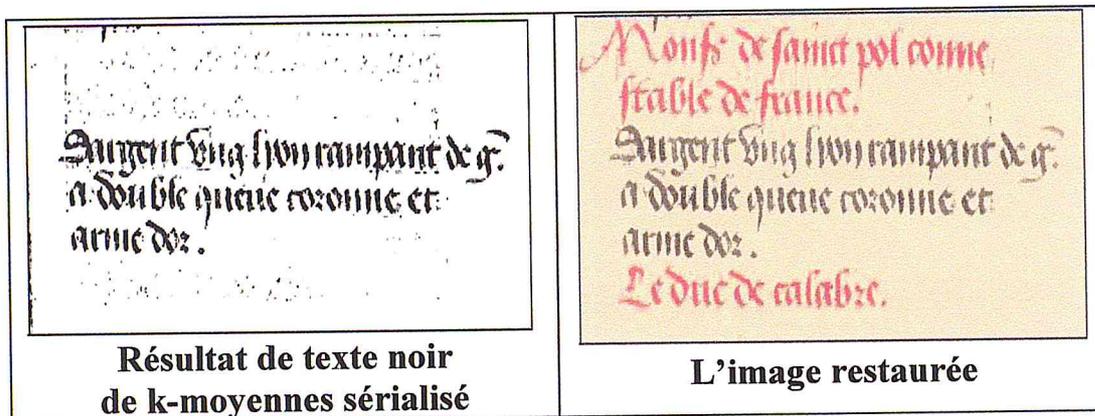


Figure 4-5 : Séparation recto/verso par analyse de la couleur

3-1-4-Méthode de séparation recto/verso proposée par Wang [52] :

Filtre médian et la méthode d'ondelette :

Le filtrage médian augmente l'intensité du pixel à l'aide du voisinage trois par trois par défaut, avec le pixel au centre remplacée par la valeur médiane de toutes les valeurs de pixel à l'intérieur de la fenêtre 3-en-3. Il a été vérifié qu'une telle performance agit comme un filtre efficace passe-bas. Mais quand il s'agit de livres numérisés, le filtre médian ne peut généralement pas spécifier le mélange de la région de l'encre noire saigné et le caractère encrée originale.

Pour classer activement l'encre noire saigné et le caractère, la méthode des ondelettes peut être appliquée.

Un avantage majeur offert par ondelettes est la capacité d'analyser une zone localisée d'un signal plus grand. L'analyse en ondelettes est souvent appliquée pour compresser ou débruiter un signal sans dégradation notable.

Algorithme basé sur le filtre médian et ondelettes :

Pour un caractère, les basses fréquences affichent les composants tant qu'une information approximative de la forme du caractère, tandis que les composantes de fréquence supérieure agissent habituellement comme détails. Comme le bruit de fond est également inclus dans l'échelle de fréquence plus élevée, l'analyse en ondelettes avec 2 niveaux perdrait les détails de caractère et les traits de caractère peuvent être liés ensemble. L'image originale serait filtrée simultanément et perdrait les informations de bord.

Afin de préserver les détails de hautes fréquences, la décomposition en ondelettes à plusieurs niveaux est choisi avec l'amélioration du coefficient haute fréquence multi-niveaux et de réduire les coefficients de fréquence de niveau 1 à 2 et supérieur.

L'ondelette de Daubechies est utilisée comme la fonction principale d'ondelette. Afin d'accroître l'efficacité de la transformée en ondelette,

l'ajustement de l'histogramme ainsi que la méthode de filtrage médian sont sélectionnés comme préprocesseurs pour gérer l'image numérisée.

Résultats et discussions :

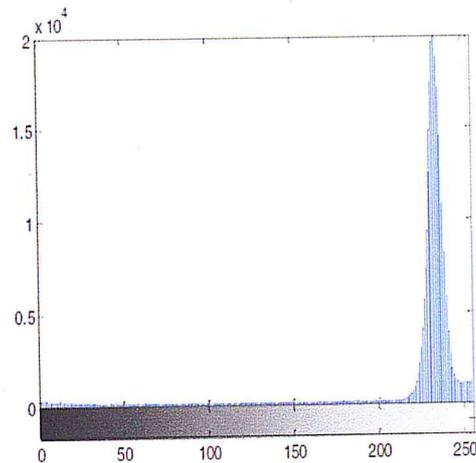
Pour l'image numérisée, le signal du flou est mélangé avec le texte. Puisque le texte anglais imprimé qui est composée par alphabets, est dans une certaine mesure simple par rapport au texte chinois basé sur des lignes, cet algorithme de récupération est principalement orienté à l'image de texte chinois.

A. Ajustement d'intensité du pixel pour réduire l'arrière-plan :

Le bruit du fond d'image est un facteur important affectant la précision de la reconnaissance. Dans la figure 4-6 (a), de résolution 512×512 les caractères anglais sont mélangés avec un gris dans le fond.

To explain the behavior of penetrant-po the Flory-Huggins theory, since it provide on the molecular structures of the comp series of parameters and very often th contribution methods are not so clear. We

- for molar volumes, the Sugden a average, equivalent results;
- the vaporization method provide: methods. Those methods also accumulated errors may exist;
- Three-dimensional (3D) approach



(a)

(b)

Figure 4-6 : (a) l'image originale avec le texte anglais (b) son histogramme.

Il y a assez de distinctions entre les caractères anglais imprimé à l'encre noire et le gris arrière-plan, ce qui est vérifié par l'histogramme de l'intensité des pixels dans la figure 4-6 (b). Le pic de niveau de gris 235 avec une largeur de pic de l'ordre de 40 est la répartition des pixels de fond. Les caractères pixels noirs sont distribués avec faible hauteur de l'histogramme autour de niveau de gris 0 donc l'arrière-plan et le caractère sont indépendants dans la région de pixel niveau de gris.

Selon cette analyse, l'ajustement de l'histogramme d'intensité de pixels pour réduire le bruit de fond est présenté comme suit :

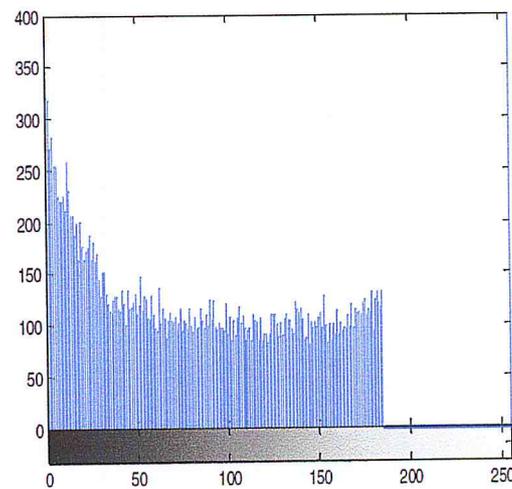
$$I'(x, y) = \begin{cases} I(x, y) & \text{if } I(x, y) < \textit{threshold} \\ 255 & \text{if } I(x, y) \geq \textit{threshold} \end{cases}$$

Les pixels sélectionnés du fond sont attribués à 255, et la zone noire ne change pas. Après ajustement, ça montre que le contraste entre le caractère anglais et son arrière-plan sur la figure 4-7 sont remarquablement améliorée.

To explain the behavior of penetrant-po the Flory-Huggins theory, since it provide on the molecular structures of the com series of parameters and very often th contribution methods are not so clear. We

- for molar volumes, the Sugden a average, equivalent results;
- the vaporization method provide: methods. Those methods also accumulated errors may exist;
- Three-dimensional (3D) approach

(a)



(b)

Figure 4-7 : (a) l'image ajustée par histogramme et (b) son histogramme.

L'ajustement d'intensité de pixel peut être utilisé pour effectuer la fonction de base de la réduction du bruit avant la récupération de caractère numérisé.

B. Médian et le filtrage par ondelette pour réduire le bruit mixte :

Dans cette section, une pièce de résolution 512×512 des caractères chinois flous numérisés d'un livre sont sélectionnés comme le montre la figure 4-8 (a) ; le nombre total de caractères chinois est de 100 par rapport à la figure 4-6 (a), l'arrière-plan de cette image est détériorée par le tableau au verso du papier et les erreurs de numérisation. Le caractère noir est mélangé avec le fond noir-gris affiché sur la figure 4-8 (b).

B. 在满足机械强度要求的前提下, 程序和光晕宽度。在满足光密度要求程序和光晕宽度。

C. 在保证感光度的条件下, 减小银与明胶的折射率存在极大的差异时, 其扩散漫射与胶层厚度成正比小, 清晰度越高, 分辨率越高。

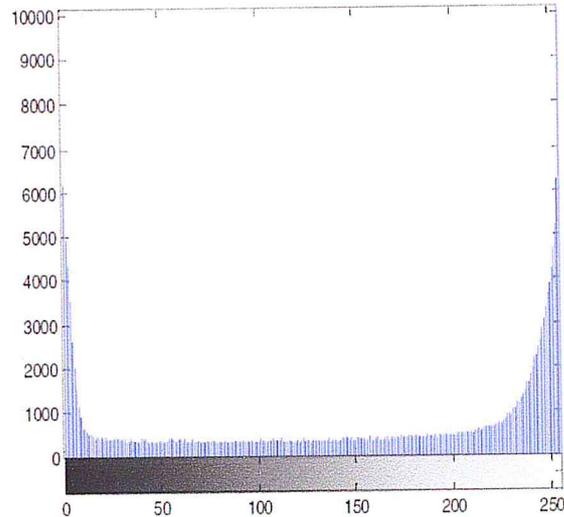
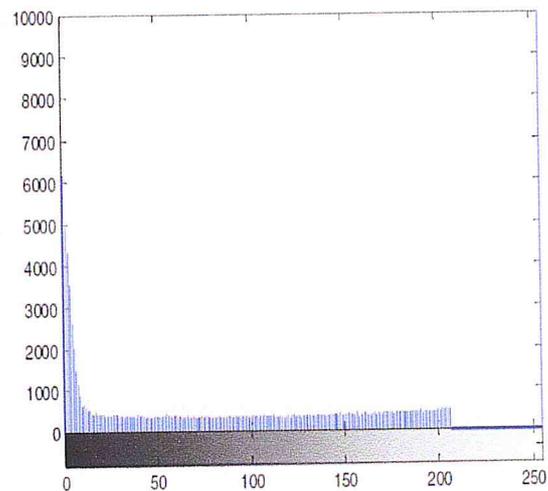


Figure 4-8 : (a) l'image floue de texte d'origine chinoise et (b) son histogramme.

Par réglage de l'intensité de pixel, l'arrière-plan ne s'est améliorée que faiblement. Le fond se confond également avec les caractères chinois, ce qui est différent de la figure 4-7 en raison du niveau de gris élevé du pixel d'arrière-plan, et le caractère n'est pas indépendant avec le fond.

B. 在满足机械强度要求的前提下, 程序和光晕宽度。在满足光密度要求程序和光晕宽度。

C. 在保证感光度的条件下, 减小银与明胶的折射率存在极大的差异时, 其扩散漫射与胶层厚度成正比小, 清晰度越高, 分辨率越高。



(a) (b)

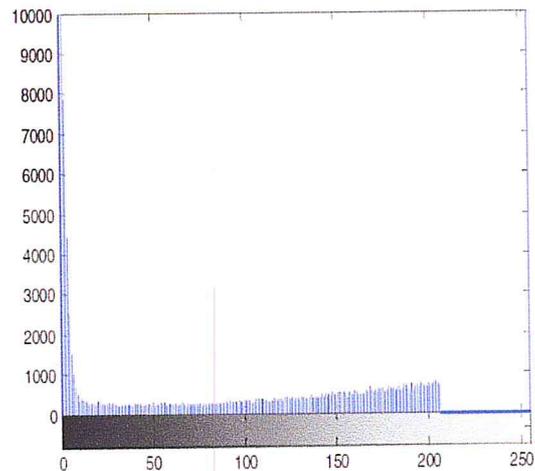
Figure 4-9 (a) l'ajustement de l'histogramme de l'image floue et (b) c'est histogramme.

Pour affaiblir le bruit de fond, le programme de filtrage médian est mis en avant pour la figure 4-9 (a). Après développement, le saignement de fond entre les caractères, en particulier dans les marges, a été réduit comme indiqué à la figure 4-10 (a). Mais il y a aussi l'image d'encre saigné mélangé avec les caractères. Dans la figure 4-10 (b), le pixel du sommet

de l'histogramme ne change pas remarquablement qu'à la figure 4-9 (b). Le filtre médian ne peut pas discriminer les informations de bords du caractère chinois de l'arrière-plan non uniforme, comme on le voit dans l'histogramme d'intensité.

B. 在满足机械强度要求的前提下，
程和光晕宽度。在满足光密度要求
程和光晕宽度。

C. 在保证感光度的条件下，减小
银与明胶的折射率存在极大的差异
时，其扩散漫射与胶层厚度成正比
小，清晰度越高，分辨率越高。



(a)

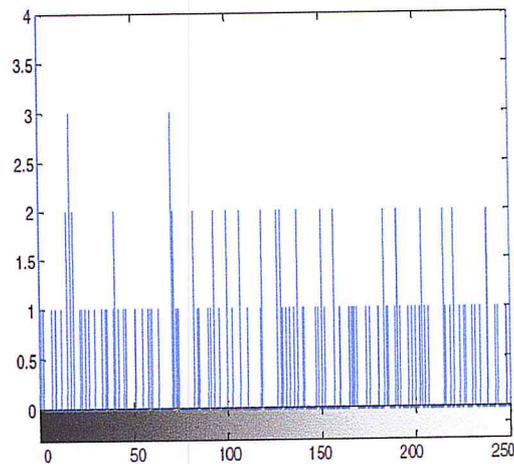
(b)

Figure 4-10 (a) l'image débruitée par le filtre médian et (b) son histogramme.

En outre, la décomposition en ondelettes à 3 niveaux est réalisée à l'image figure 4-10 (a) du filtrage médian. Comme le montre la figure 4-11 (a), les bruits noirs mixtes ont été réduits distinctement, tandis que la résolution de bord du caractère chinois a été préservée.

B. 在满足机械强度要求的前提下，
程和光晕宽度。在满足光密度要求
程和光晕宽度。

C. 在保证感光度的条件下，减小
银与明胶的折射率存在极大的差异
时，其扩散漫射与胶层厚度成正比
小，清晰度越高，分辨率越高。



(a)

(b)

Figure 4-11 (a) L'image débruitée par filtre à ondelette et (b) son histogramme.

Cela implique que les informations approximatives de la forme du caractère chinois dans la région de basse fréquence et les informations de détail de la bordure de caractère à haute fréquence ont été conservées. Le bruit de fond est également un composant à haute fréquence, mais a été réduit au cours du niveau 2 à 1 de la décomposition en ondelettes, en affectant des coefficients de fréquence. En outre, l'histogramme de l'intensité a été remarquablement changé, comme on le voit sur la figure 4-10 (b).

OCR Résultat des différents algorithmes de débruitage

Basé sur le logiciel OCR Han'wang, le taux de précision de la reconnaissance de l'image originale, histogramme ajusté, médiane filtré et ondelettes transformé a été démontré dans le tableau 1. Le taux de précision de l'image originale est de seulement 1% en raison de la perturbation du fond mixte. Ainsi, on vérifie que le fond est un facteur essentiel influant la précision de l'OCR.

Après ajustement de l'histogramme, la précision de reconnaissance a été réduite à 0%. Cette procédure doit être effectuée pour modifier l'arrière-plan uniforme comme on le voit sur la figure 4-8 (b) et figure 4-9 (b), bien qu'il semble que le bruit de fond mélangé avec des caractères n'est pas amélioré.

Par filtrage médian, le taux de précision a été amélioré à 18% après débruitage de l'arrière-plan dans une mesure limitée. Enfin, la décomposition en ondelettes à 3 niveaux a accru la précision de la reconnaissance de 45%. Les résultats indiquent que l'algorithme combiné a remarquablement amélioré le taux signal/bruit, et surtout affaibli l'encre de saignement mixé avec les caractères chinois.

Méthode	Caractères reconnus	Taux de Précision
Originale	1	1%
Après ajustement de l'histogramme	0	0%
Après filtrage médian	18	18%
Après 3 niveaux de décomposition en ondelettes	45	45%

Tableau 4-1 : Nombre de Caractères reconnus et taux de précision par des méthodes de débruitage différents

3-1-5-Restauration des formes par modèle :

Certaines méthodes sont basées sur la définition des modèles pour la restauration des formes reflétant plus ou moins fidèlement les dégradations. Un exemple d'un modèle est proposé par Baird [53], qui inclut un certain nombre de paramètres sur l'image.

Un autre modèle propose par Kanungo [54], simule certains défauts, comme le flou par des opérations de fermeture morphologique.

Dans [55], le modèle propose est basé sur les paramètres extraits de l'image dégradée produite.

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 (Figure 4-12). 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.

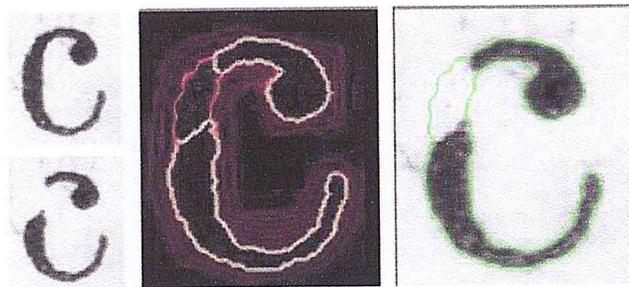


Figure 4-12 : Principe de la restauration d'une lettre cassée par les contours actifs [4]

3-1-6-Restauration des formes par prototypage :

Cette approche est appliquée généralement sur les caractères d'un document imprimé [56] Elle consiste à collectionner toutes les formes similaires d'une page, ensuite on construit le modèle de la forme idéale d'une lettre à partir de la superposition de toutes les formes dégradées de la même lettre. Une méthode de restauration des formes en utilisant le prototypage, est celle de Hobby et Baird [57].

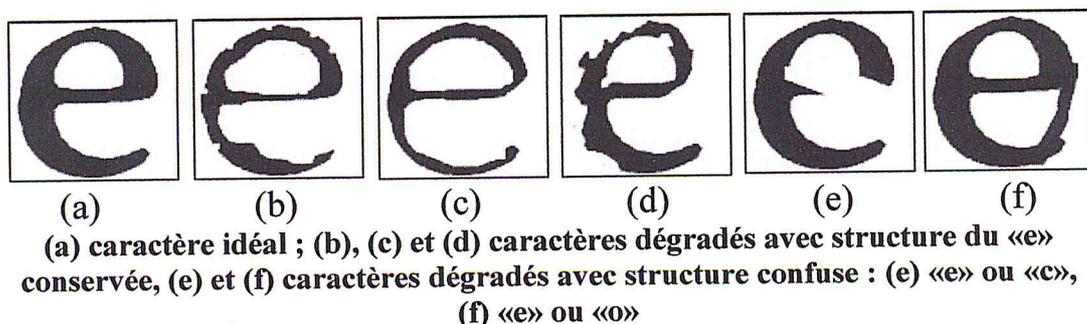


Figure 4-13 : Cas de dégradations de la lettre «e»

3-1-7-Restauration géométrique de la courbure [58][59][60][61] :

Un autre défaut qui peut apparaître lors de la numérisation des livres volumineux, est l'apparition de la courbure à proximité de la reliure du livre, ce qui réduit la lisibilité du document. Le traitement de ce défaut est alors nécessaire.

a- Restauration par un modèle 2D :

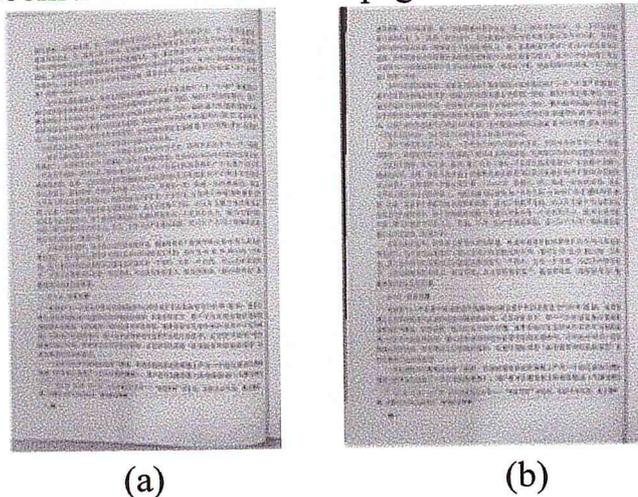
Les méthodes de restauration de la courbure basées sur un modèle 2D procèdent en deux étapes :

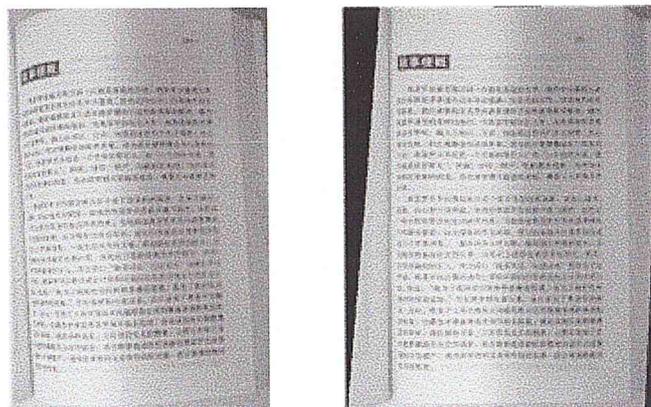
1. Calcul de la courbure : Le calcul de la courbure peut être fait à partir des lignes de textes ou par extrapolation des bords de la page.
2. Traitement géométrique pour la corriger.

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.

b- 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 photographique), soit particulière (deux appareils prises sous des éclairages différents). Pour une prise de vue normale, les déformations des feuilles sont généralement modélisées par une surface cylindrique aux abords de la reliure. Pour une prise de vue particulière, une mise en correspondance est calculée entre les images pour reconstruire le relief de la page.





(a')

(b')

(a) (a') : L'image originale - (b) (b') : L'image résultat
Figure 4-14 : Restauration géométrique de la courbure

3-2- Méthodes de restauration non-supervisées « Aveugle » :

3-2-1-Restauration du recto/verso par séparation de sources :

L'idée de base de la méthode proposée par F. DRIRA, F. LEBOURGEOIS et H. EMPTOZ consiste à classer chaque pixel de l'image du document dégradé dans l'une de ces trois classes : le fond, le texte original et le texte en transparence.

Le problème est donc un problème de séparation de couches superposées avec différents niveaux de couleurs. La classe n°3 doit être supprimée de l'image et remplacée par la couleur du fond (par exemple par la moyenne des pixels du support papier). Néanmoins, une classification globale non supervisée en trois classes sur l'image entière n'est pas suffisante pour correctement extraire le texte original.

Ils ont appliqué l'algorithme des « nuées dynamiques » sur les données décarrelées par analyse en composantes principales dans l'espace des couleurs Rouge-Vert-Bleu. Les résultats d'une classification automatique en trois classes sur une image de test sont donnés sur la figure 4-15.

Ces résultats montrent qu'une telle application ne permet pas de séparer le recto du verso. Par ailleurs, ils remarquent que la classe image correspondante au recto, contient toujours des informations reliées au verso même après avoir effectué cette classification.



Figure 4-15 : Résultat d'une classification globale en trois classes sur une image de document, de gauche à droite : l'image dégradée, image de classe 1, image de classe 2 et image de classe 3 respectivement.

Ils ont alors introduit une méthode de séparation recto/verso basée sur une technique de segmentation récursive. Pour simplifier l'analyse et optimiser sa complexité, nous nous sommes limités au cas d'un problème à deux classes : texte original du recto ou autre. L'approche proposée est une approche de segmentation récursive utilisant l'analyse en composantes principales et histogramme logarithmique. Le choix d'une telle méthode de dé-corrélation des données et d'un tel algorithme de segmentation, a été dirigé vers les outils couramment utilisés dans la littérature. Ceci a pour objectif de tester mieux l'efficacité de la méthode.

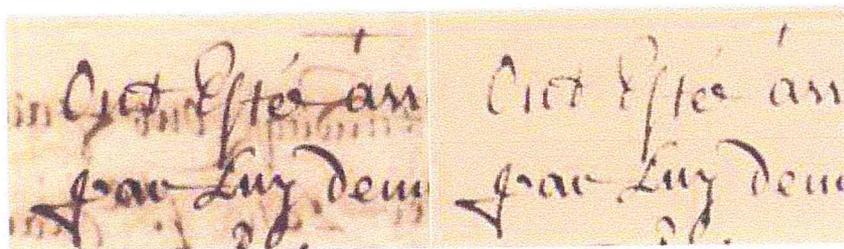


Figure 4-16 : Gauche : Image de document dégradé - Droite : image restaurée par la méthode de [47].

3-2-2-Restauration Colorimétrique [56] :

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.

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 image qui permet de quantifier la capacité de distinguer deux régions distinctes (Figure 4-18).

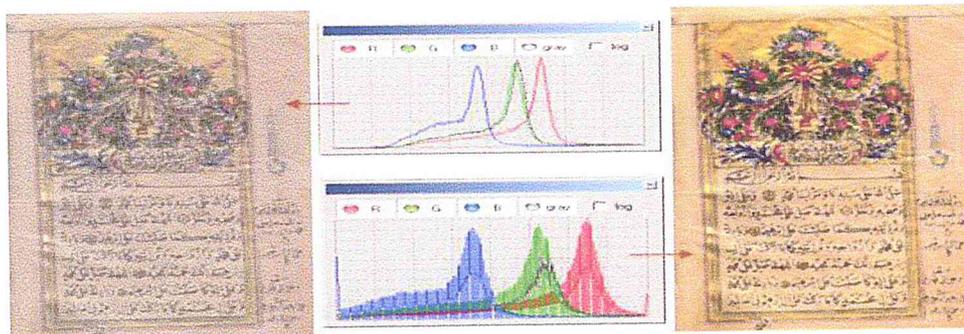


Figure 4-17 : Restauration de la dégradation colorimétrique par ajustement du contraste

4- Conclusion :

Dans ce chapitre nous avons mentionné différents types de restaurations dites supervisées « non-aveugle » qui nécessite un utilisateur qui connaît à priori le contenu du document mais qui ne soit pas toujours disponible d'où leur utilisation est difficile.

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 utilisés.

Il y a aussi d'autres types de restaurations dites non supervisées « aveugle » qui ne nécessitent pas l'apprentissage mais qui sont rarement utilisées.

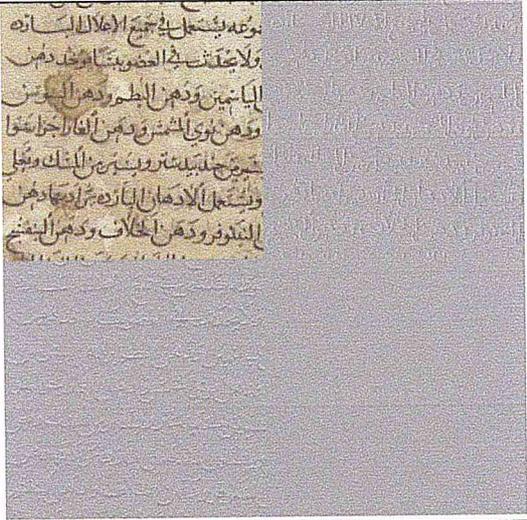
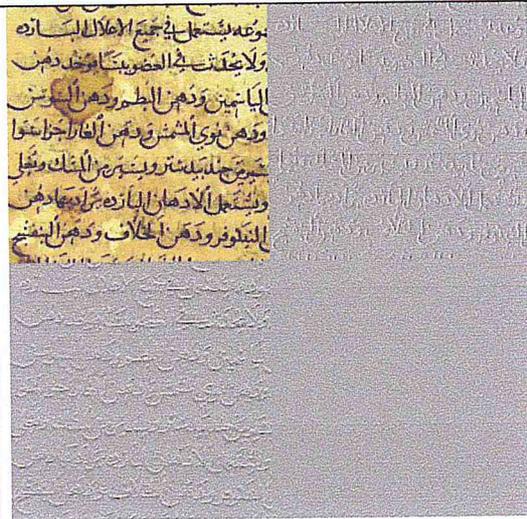
Chapitre V - Application (Méthodes)

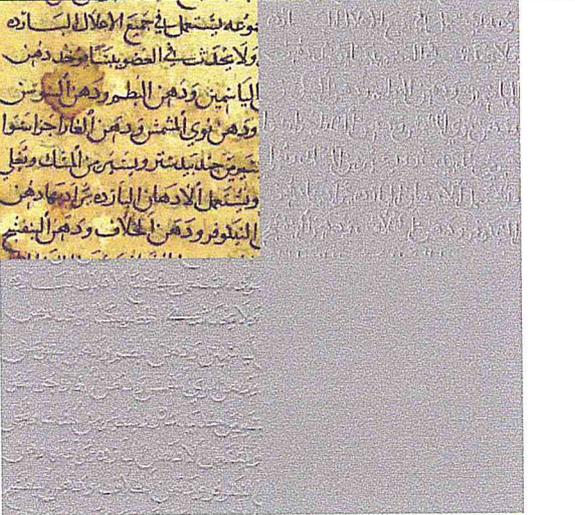
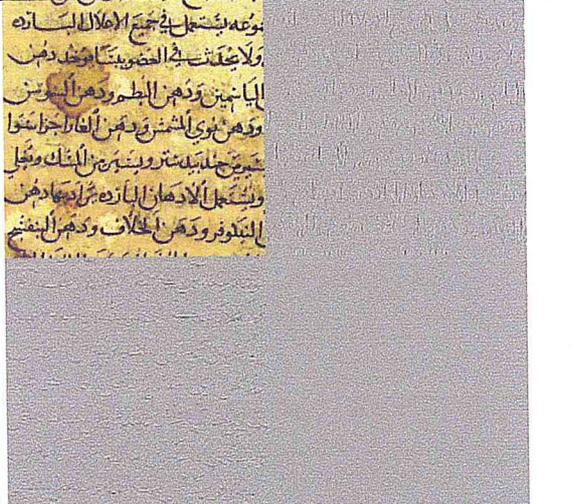
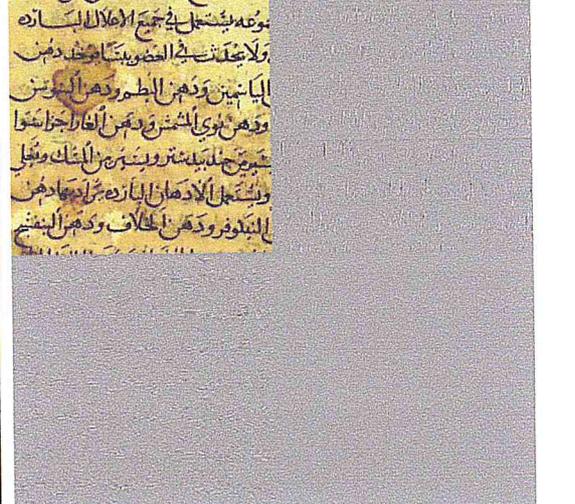
1- Introduction :

Dans ce chapitre nous allons présenter les méthodes que nous allons utiliser pour la restauration des anciens manuscrits avec la transformée en ondelette discrète, tel que la séparation recto/verso, la détection des contours et le débruitage.

2- Nos Méthodes :

Après avoir testé les résultats des différents types d'ondelettes dont nous disposons, nous avons constaté ce qui suit :

Type d'ondelettes	Image résultante	Remarques
Haar		Meilleur pour la détection des contours Bon pour la séparation fond/texte
Daubechies 2		Bon pour la détection des contours

<p>Daubechies 4</p>		<p>Bon pour la détection des contours</p>
<p>Daubechies 6</p>		<p>Peut être utilisé dans la détection des contours</p>
<p>Daubechies 8</p>		<p>Peut être utilisé dans la détection des contours</p>

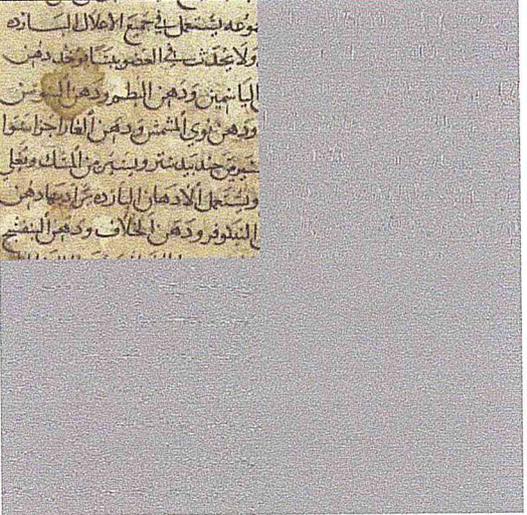
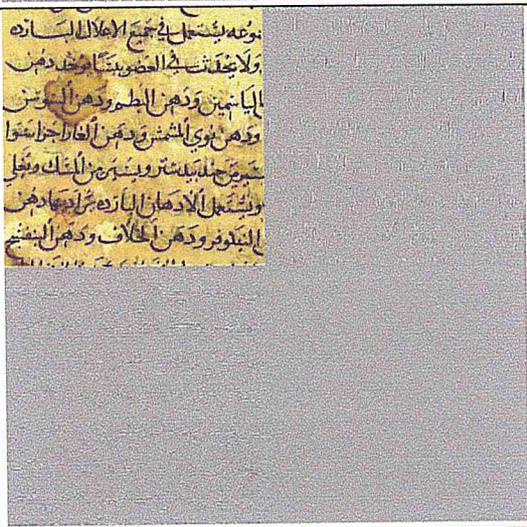
<p>Biorthogonale</p>		<p>Meilleur pour l'élimination du bruit</p> <p>Bon pour la séparation fond/texte (moins le bruit plus la précision du seuillage)</p>
<p>Fast Symmlet 8</p>		<p>Peut être utilisé dans la détection des contours</p>

Tableau 5-1 : Les résultats de la décomposition par différent types d'ondelettes

3-1 Séparation recto/verso :

Notre méthode consiste à décomposé l'image par ondelettes (on choisit l'ondelette de Haar et Biorthogonal9/7) avec un seul niveau d'itération, voici la figure 5-1 ci-dessous qui montre la décomposition en utilisant l'ondelette de Haar (gauche) et Biorthogonal9/7 (Droite) :

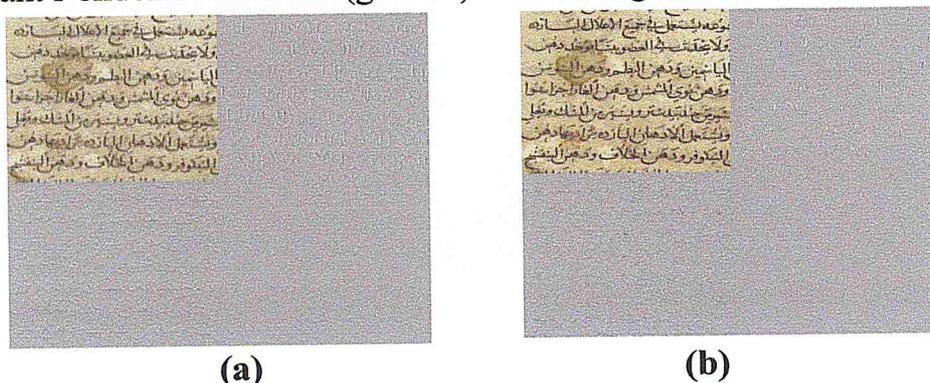


Figure 5-1 : décomposition avec niveau 1 en utilisant (a) Haar et (b) Biorthogonal9/7.

Après la décomposition nous avons choisi l'image d'approximation (LL) pour faire la séparation entre le fond et le texte (car le seuillage des coefficients ne donne pas le résultat souhaité) avec un algorithme qui calcule le point séparative entre le texte et le fond (seuil) qui se trouve entre les deux pics de l'histogramme de l'image texte où un pic représente le texte et le 2eme pic représente le fond.

Algorithme de calcul du seuil de séparation :

La 1^{ère} étape de cette méthode de séparation consiste à construire l'histogramme de l'image en niveaux de gris. Dans la plupart des manuscrits on trouve 2 pics, un pic représentant le texte et l'autre représentant le fond.

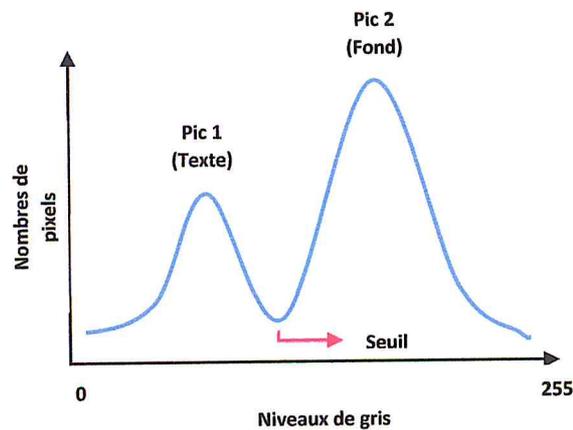


Figure 5-2 : Histogramme montré les pics du texte et du fond

Variables

i, j : entiers ;

Histogramme : vecteur d'entiers de longueur = 256 ;

Début

Pour $i=1$ à largeur (image) faire

 Pour $j=1$ à hauteur (image) faire

 Histogramme [(Rouge(Pixel(i, j)) + Bleu(Pixel(i, j)) + Vert(Pixel(i, j)) / 3)] ++ ;

 Fait ;

Fait ;

Fin ;

La 2^{ème} étape est de lisser l'histogramme pour faciliter la détection du seuil car l'histogramme peut avoir plusieurs déformations (faux sommets)

Remarque : le lissage se fait plusieurs fois pour éliminer tous les défauts.

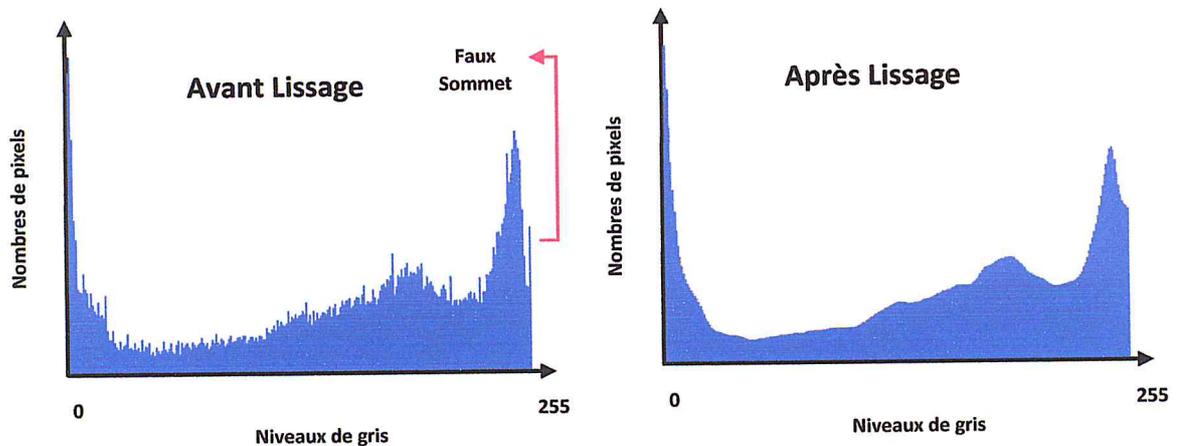


Figure 5-3 : Avant et après le lissage d'un histogramme

Variables

i : entier ;

Histogramme : vecteur d'entiers de longueur 256 ;

Début

Pour i=2 à longueur (Histogramme)-1

faire

Histogramme [i] <- (Histogramme [i-1] + Histogramme [i+1]) / 2 ;

Fait ;

Fin ;

La 3^{ème} étape consiste à trouver les 2 sommets (pics) car le seuil est la valeur minimum entre eux, le 1^{er} sommet est facile à trouver c'est simplement la valeur maximum de l'histogramme (généralement le pic du fond) ensuite en fait deux analyses à gauche et à droite du pic pour trouver le 2^{ème} pic. Pour les deux analyses tant que la valeur de l'histogramme diminue on avance jusqu'à atteindre la plus basse valeur possible, puis on continue tant que la valeur augmente jusqu'à atteindre la plus haute valeur possible, ensuite on compare les deux valeurs (à gauche et à droite) et la plus grande sera le pic 2 (généralement le pic du texte).

Variables

i, j, pic1, pic2, picGauche, picDroit, seuil : entiers ;

Histogramme : vecteur d'entiers de longueur = 256 ;

Début

pic1 <- 1 ;

Pour i=2 à longueur (Histogramme)

faire

Si Histogramme [pic1] < Histogramme [i] alors pic1 <- i ;

Fsi ;

Fait ;

```

i <- pic1
Tant que ( i > 1 et Histogramme [i-1] <= Histogramme [i] )i-- ;j=i ;
Tant que ( i > 1 et Histogramme [i-1] >= Histogramme [i] )i-- ;
Si i=j alors picGauche <- min(histogramme) sinon picGauche <- i ;
i <- pic1 ;
Tant que ( i < longueur(Histogramme) et Histogramme [i+1] <= Histogramme [i] )i++ ;j=i ;
Tant que ( i < longueur(Histogramme) et Histogramme [i+1] >= Histogramme [i] )i++ ;
Si i=j alors picDroit <- min(histogramme) sinon picDroit <- i ;
si Histogramme [picGauche] > Histogramme [picDroit] alors pic2 <- picGauche ;
sinon pic2 <- picDroit ;
fsi ;
seuil <- pic1;
Pour i=min(pic1, pic2) à max(pic1, pic2) faire
    Si Histogramme [seuil] > Histogramme [i] alors seuil <- i ;
    Fsi ;
Fait ;

```

Fin ;

Après l'utilisation de cet algorithme, on va réduire l'intensité des pixels inférieurs au seuil et augmenter celle des pixels supérieurs au seuil, voici l'image résultante :

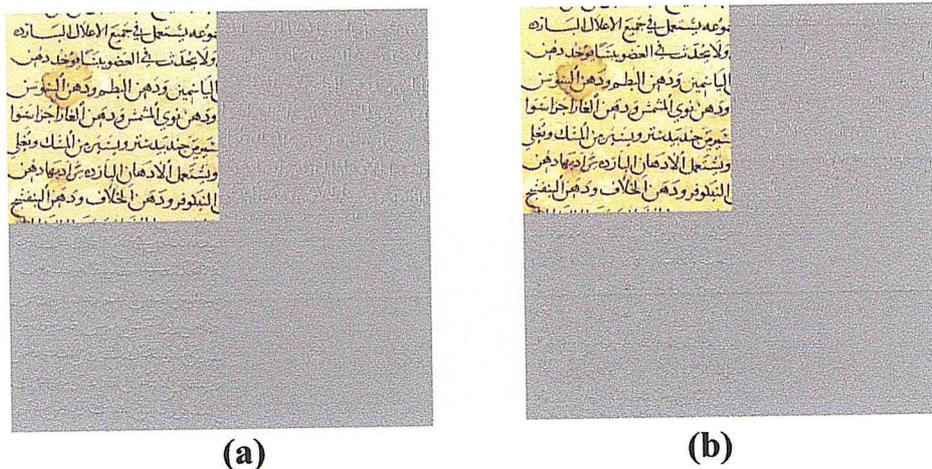


Figure 5-4 : Seuillage d'image d'approximation après la décomposition de (a) Haar et (b) Biorthogonal9/7.

Cette transformation permet de supprimer plus de bruits et conserver totalement les détails dans les coefficients horizontaux, verticaux et diagonaux, et à la fin on va faire la reconstruction de l'image avec la transformé en ondelette inverse, la figure 5-5 montre le résultat final :

نوعه يستعمل في جميع الاعلال الباردة
ولا يحدث في العضو بيتا فوجد دهن
الياسمين ودهن البطم ودهن السنوسن
ودهن نوي المشمش ودهن الغار اجزاسوا
شعير من جند بيدشتر وبيتير من المشك وبعلي
ويستعمل الادهان الباردة من ادبها دهن
النبتوفر ودهن الخلاف ودهن البنفسج

(a)

نوعه يستعمل في جميع الاعلال الباردة
ولا يحدث في العضو بيتا فوجد دهن
الياسمين ودهن البطم ودهن السنوسن
ودهن نوي المشمش ودهن الغار اجزاسوا
شعير من جند بيدشتر وبيتير من المشك وبعلي
ويستعمل الادهان الباردة من ادبها دهن
النبتوفر ودهن الخلاف ودهن البنفسج

(b)

نوعه يستعمل في جميع الاعلال الباردة
ولا يحدث في العضو بيتا فوجد دهن
الياسمين ودهن البطم ودهن السنوسن
ودهن نوي المشمش ودهن الغار اجزاسوا
شعير من جند بيدشتر وبيتير من المشك وبعلي
ويستعمل الادهان الباردة من ادبها دهن
النبتوفر ودهن الخلاف ودهن البنفسج

(c)

Figure 5-5 : Résultat final après seuillage en utilisant l'ondelette de (a) Haar, (b) Biorthogonal9/7 et (c) sans ondelettes.

Pour obtenir des résultats plus positifs et optimaux, on doit refaire le seuillage par binarisation sur l'image résultante ce qui va rendre le texte noir et le fond blanc, (voir la figure (5-6)) :

نوعه يستعمل في جميع الاعلال الباردة
ولا يحدث في العضو بيتا فوجد دهن
الياسمين ودهن البطم ودهن السنوسن
ودهن نوي المشمش ودهن الغار اجزاسوا
شعير من جند بيدشتر وبيتير من المشك وبعلي
ويستعمل الادهان الباردة من ادبها دهن
النبتوفر ودهن الخلاف ودهن البنفسج

(a)

نوعه يستعمل في جميع الاعلال الباردة
ولا يحدث في العضو بيتا فوجد دهن
الياسمين ودهن البطم ودهن السنوسن
ودهن نوي المشمش ودهن الغار اجزاسوا
شعير من جند بيدشتر وبيتير من المشك وبعلي
ويستعمل الادهان الباردة من ادبها دهن
النبتوفر ودهن الخلاف ودهن البنفسج

(b)

نوعه يستعمل في جميع الاعلال البارده
ولا يحدث في العضو بيتا يؤخذ دهن
الياسمين ودهن البطم ودهن السنبل
ودهن نوري المشمش ودهن الفاناجزاسوا
بشيرة جندبيد شتر ويسر من المسك وبعلي
ويستعمل الادهان البارده من ادبها دهن
البنكوف ودهن الخراف ودهن البنفسج

(c)

Figure 5-6 : Le seuillage après l'utilisation de notre méthode – (a) Haar, (b) Biorthogonale9/7 et (c) sans ondelettes.

Remarque :

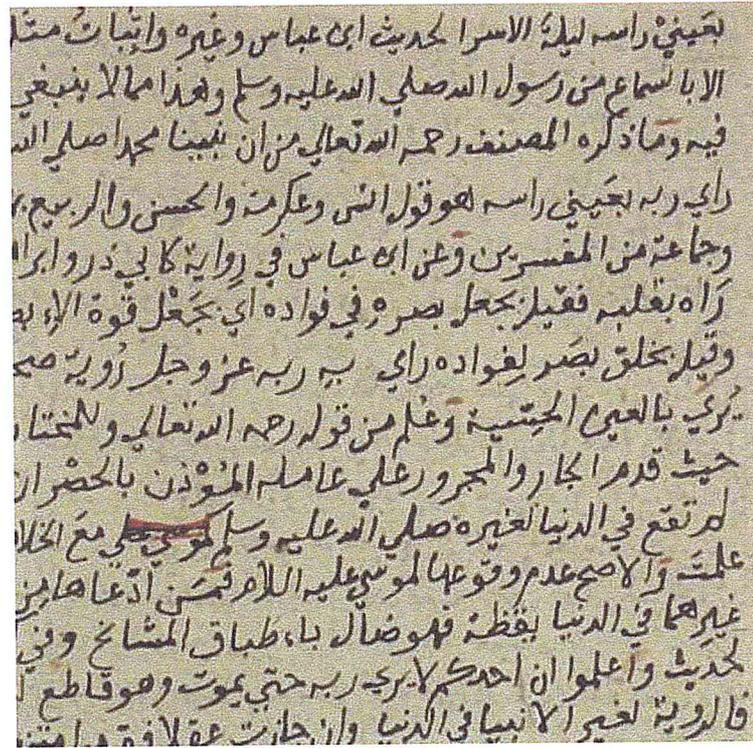
- La séparation par ondelettes fait un bon seuillage grâce à la disparition de quelque bruit qui gêne le calcul du seuil.
- La binarisation fait un bon résultat par rapport au seuillage par augmentation du contraste car elle élimine toute dégradation du fond.

3-2- Détection de contours :

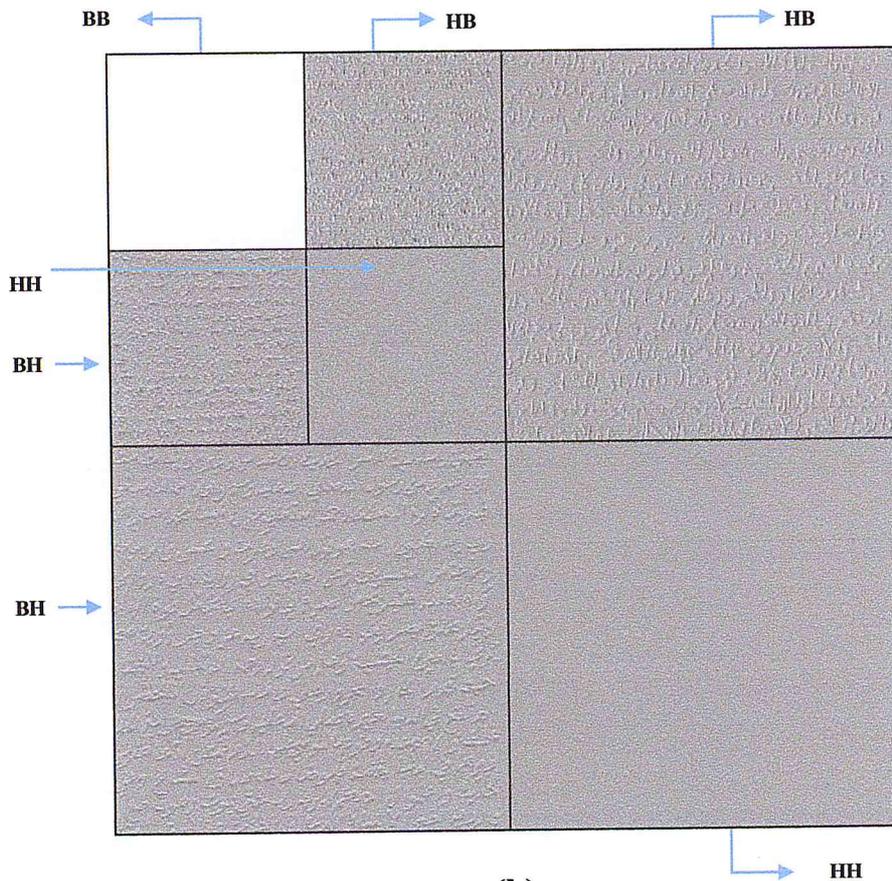
Dans cette partie, nous n'allons pas nous intéresser directement à des détecteurs de contours. Nous aborderons la technique de la transformation en ondelette qui permet de représenter les hautes fréquences de l'image à différentes résolutions. Puisque les contours d'une image peuvent être considérés comme les hautes fréquences, la transformée en ondelette peut alors permettre d'afficher les contours.

3-2-1. Méthode d'élimination de l'image d'approximation (BB) :

La figure 5-7 montre le résultat de l'application de la transformée en ondelette de Haar à une image avec deux niveaux de décomposition et identifie les groupes de fréquence (BB, HB, BH et HH) obtenus à chacune des deux premières résolutions. BB représente les basses fréquences du filtrage des lignes et des colonnes alors que HB exprime les hautes fréquences des lignes et les basses fréquences des colonnes. BH délimite les hautes fréquences des colonnes et les basses fréquences des lignes alors que HH représente les hautes fréquences du filtrage des lignes et des colonnes.



(a)



(b)

Figure 5-7 : (a) Image originale, (b) après transformation de (a) avec deux niveaux de décomposition en utilisant l'ondelette de Haar

Cette analyse du signal va nous permettre d'extraire des contours par les ondelettes. En effet, nous remarquons que la zone **HB** correspond aux contours verticaux, **BH** aux contours horizontaux et **HH** aux contours diagonaux. Par la suite, la détection des contours par ondelette reviendra essentiellement à choisir une ondelette appropriée ainsi que la bonne résolution.

Notre méthode consiste à éliminer la zone **BB** (Basses fréquences) de l'image transformée (c'est-à-dire remplacer chaque pixel de la zone **BB** par un pixel blanc), et après en va faire la transformation inverse pour obtenir les contours, voir la figure 5-8.

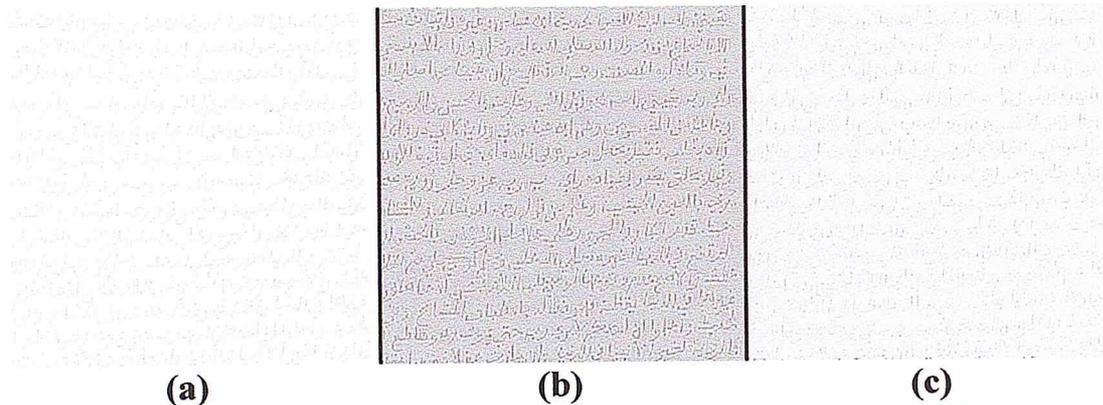


Figure 5-8 : Détection des contours avec l'ondelette de (a) Haar, (b) Bior9/7, (c) Daubechies 2.

Inconvénient : diminution de la qualité de l'image résultante.

Remarque :

- Cette méthode n'est pas la meilleure méthode de détection mais elle donne un résultat pas mal.
- Après avoir essayé plusieurs niveaux pour la détection des contours, nous remarquons que le 2^{ème} niveau donne le meilleur résultat.

3-2-2. Méthode de fusion des coefficients (détails) :

Le processus de cette méthode consiste à décomposer l'image par une ondelette ensuite fusionner les images de détails (coefficients) en mode lumière linéaire « Linear Light » [63] qui est le plus convenable parmi les autres modes testés dans Adobe Photoshop CC, voir la figure ci-dessous :

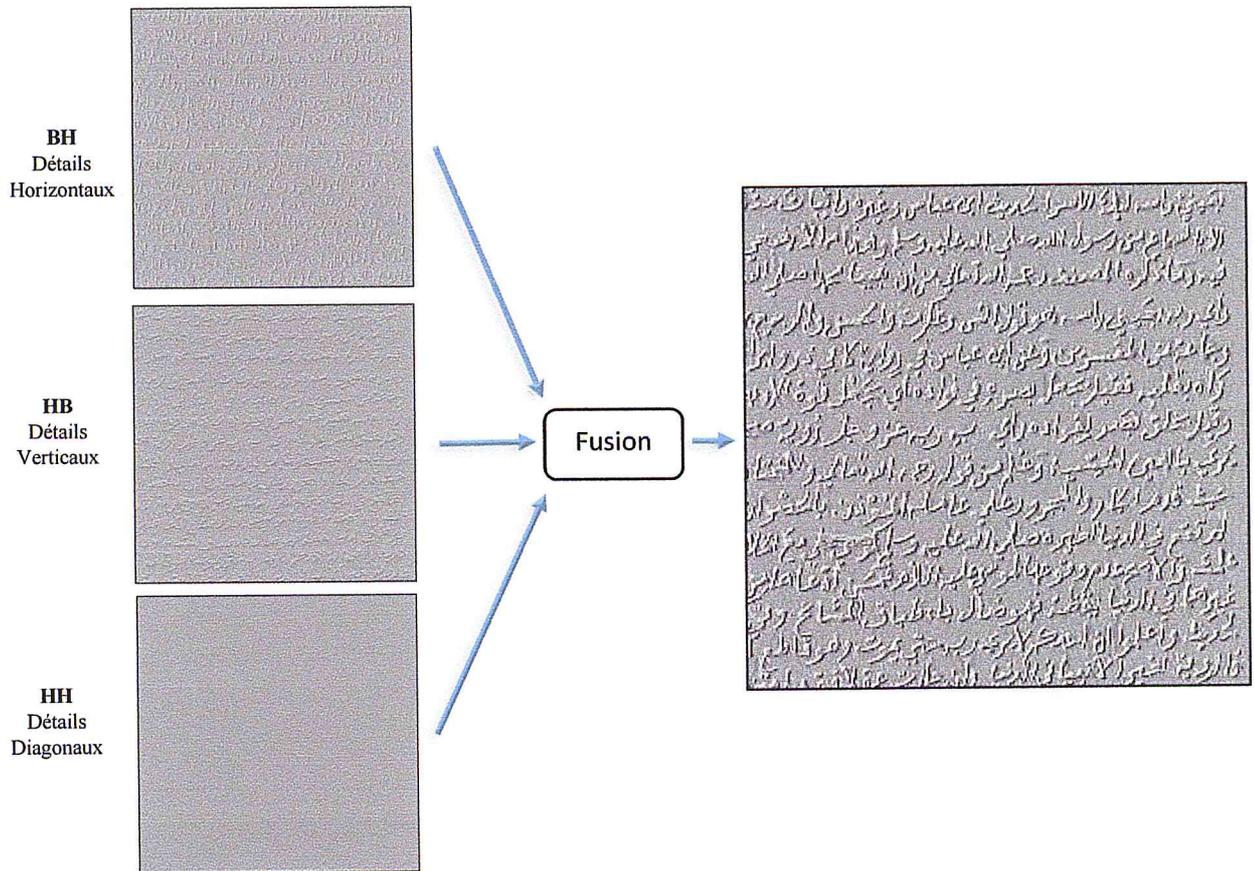


Figure 5-9 : Détection des contours par la méthode de fusion des coefficients (détails)

Remarque :

- Diminution de la résolution de l’image résultante.
- Cette méthode n'est pas la meilleure méthode de détection mais elle donne un bon résultat.

Comparaison entre les deux méthodes :

On remarque que la deuxième méthode (fusion) donne des contours plus clairs que la 1^{ère} méthode.

3-3 Débruitage :

3-3.1 Débruitage des coefficients (détails) :

La méthode proposée a pour but la réduction de degré du bruit dans l’image (Noise). D’un autre coté cette méthode se base essentiellement sur la transformation des images par ondelettes.

Principe général de la méthode proposée :

La méthode proposée est basée sur l'image transformée en ondelettes, en appliquant le processus de réduction du bruit et l’appliqué sur l’image transformée, puis inverser la transformation pour obtenir

l'image restaurée. La figure 5-10 représente le graphe de la technique proposée.

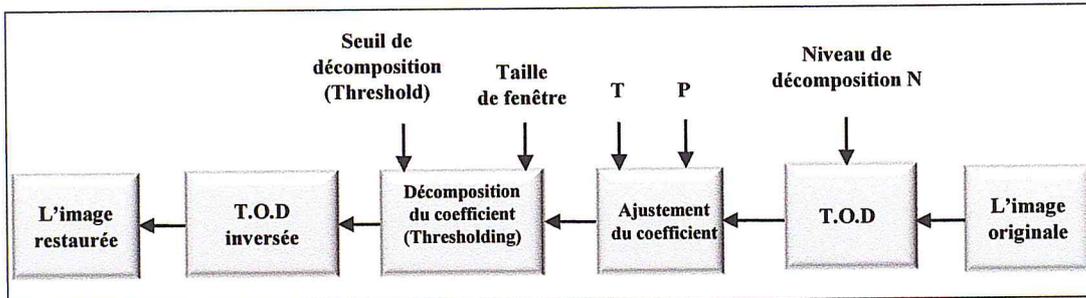


Figure 5-10 : Graphe de la technique de restauration des documents suggéré

Nous allons présenter et expliquer la méthode proposée et ses différentes étapes en ce qui suit.

Après avoir étudié les avantages et les inconvénients des différentes techniques utilisées dans le processus de débruitage de documents numériques et pour être en mesure de réduire les taux de bruit numérique émis par le processus de numérisation, en utilisant les ondelettes autant que possible, nous avons choisis la fonction mathématique qui nous permet de réduire le bruit situé dans les coefficients obtenus après l'application de la transformée en ondelettes.

Le processus de sélection de l'ondelette approprié représente une tâche difficile et importante en même temps pour obtenir un processus de restauration acceptable.

Dans ce contexte, nous nous sommes appuyés sur la fonction mathématique suivante du processus de débruitage approuvé (3) :

$$F_{T,P}(x) = \begin{cases} (|X|^p \times T^{1-P}) \times \text{Sign}(X) & \text{if } 0 \leq |X| \leq T \\ (1 - (1 - |X|)^p \times (1 - T)^{1-P}) \times \text{Sign}(X) & \text{if } T < |X| \leq 1 \end{cases} \quad (3)$$

Sachant que :

$$\text{Sign}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases} \quad (4)$$

Et : $T \in [0, 1], P \in [1, \infty]$

[64]

- Choisir le degré de séparation «T» :

Le processus de sélection du degré de séparation «T» permet de déterminer la proportion de bruit numérique qui peut être supprimé. Car la valeur optimale pour le degré de séparation «T» est la seule en mesure de réduire ou de supprimer le brouillage sans mettre en évidence les effets secondaires négatifs alors que la valeur exagérée pour «T» peut provoquer un affaiblissement des bords des formes dans l'image, une valeur non suffisante pour «T» peut conduire à une augmentation du bruit dans l'image numérique.

- Choisir l'élément «P» :

L'élément «P» Contrôle le degré de soutien du contraste entre le texte et le fond du document. Après avoir étudié l'effet de l'élément «P» sur les résultats des expériences que nous avons fait on a décidé de donner à l'élément «P» la valeur 3 dont la validité et l'efficacité de cette valeur a été vérifiée basée sur de vieux documents.

- Réduction du bruit numérique :

Dans le cas suppression du bruit numérique, le processus de sélection des ondelettes et du niveau de décomposition (Decomposition Level) de manière expérimentale et en se basant sur les images destinées à l'essai. En tenant compte des recherches antérieures, nos observations expérimentales nous ont dirigé vers l'utilisation des ondelettes de type (Biorthogonale 9/7) en se basant sur trois degrés de dégradation.

On montre Ci-dessous, les étapes de travail de la technique proposée :

- La division de l'image en utilisant des ondelettes (Biorthogonale 9/7)
- L'application du seuil sur les coefficients obtenus après l'utilisation de la transformée en ondelettes sur l'image bruitée.
- Reconstruire l'image en utilisant la transformé en ondelette inverse.

Nous offrons également L'algorithme suivant utilisé et basé sur les étapes mentionnées ci-dessus :

Soit I une image de taille M*N :

$$(A_i, H_i, V_i, D_i)_{i=1..n} = T.O(I)$$

Avec n : Niveau de décomposition

Ajustement : amélioration

pour i=1 jusqu'à n

faire

$$\tilde{H}_i = F_{T,P}(H_i)$$

$$\begin{aligned}\tilde{V}_i &= F_{T,P}(V_i) \\ \tilde{D}_i &= F_{T,P}(D_i)\end{aligned}$$

fin pour

$$I_{pretreated} = T \cdot O^{-1} \left(A_i, \tilde{H}_i^{k,a}, \tilde{V}_i^{k,a}, \tilde{D}_i^{k,a} \right)_{i=1..n}$$

$$F_{T,P}(x) = \begin{cases} (|X|^p \times T^{1-P}) \text{Sign}(X) & \text{if } 0 \leq |X| \leq T \\ \left(1 - (1 - |X|)^p \times (1 - T)^{1-p} \text{Sign}(X) \right) & \text{if } T < |X| \leq 1 \end{cases}$$

$$T \in [0, 1], P \in [1, \infty]$$

$$\text{Sign}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

On va appliquer cet algorithme sur les deux images bruitées suivantes :

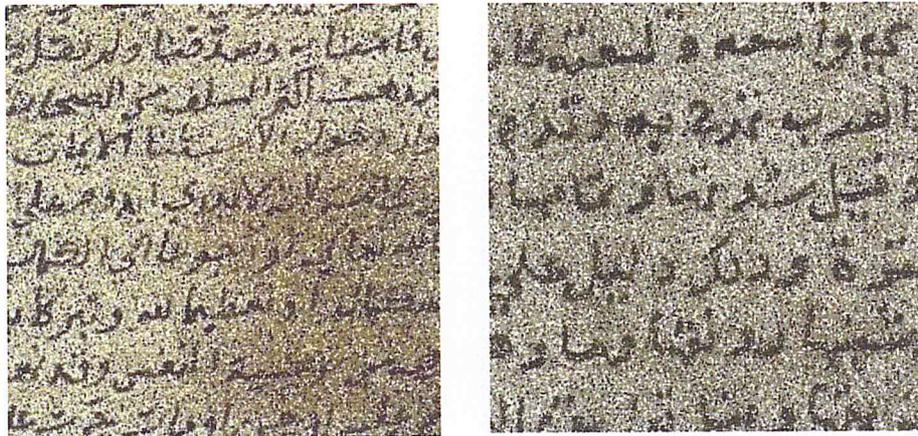
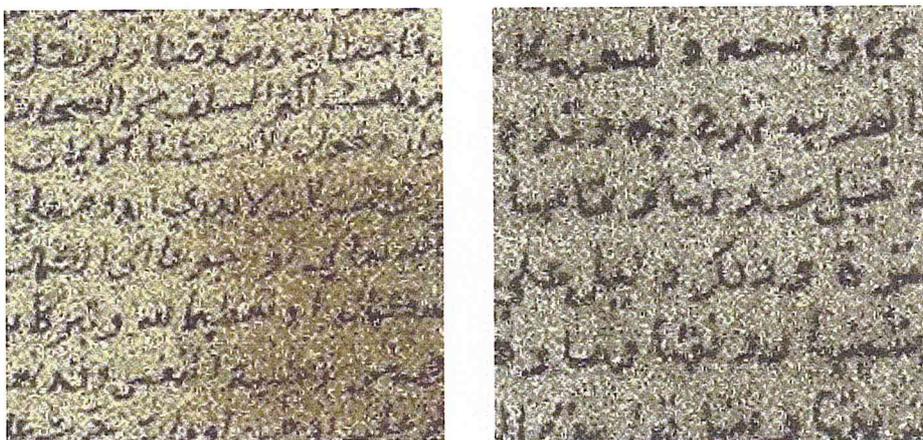
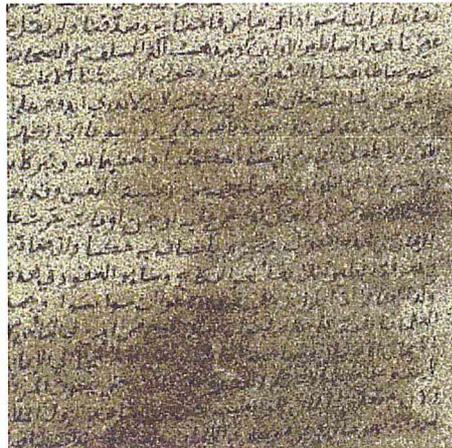


Figure 5-11 : Deux images manuscrites bruitées

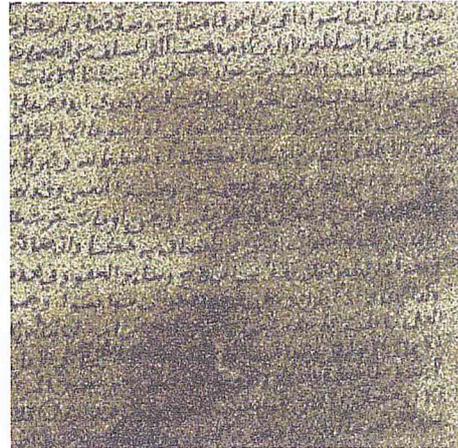
Voici le résultat pour les différentes ondelettes :



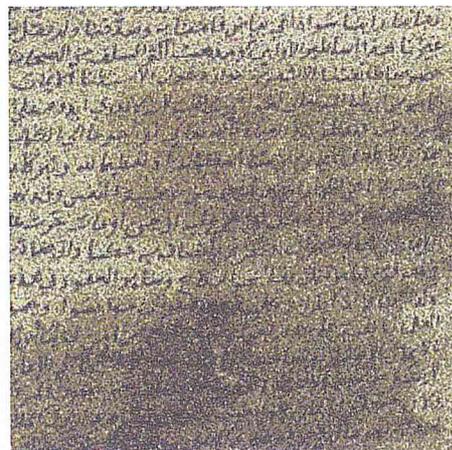
(a) Biorthogonale



(b) Haar



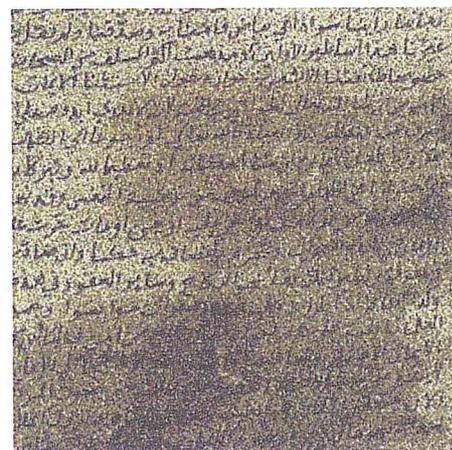
(c) Daubechies 2



(d) Daubechies 4



(e) Daubechies 6



(f) Daubechies 8



(g) Fast Symmlet 8

Figure 5-12 : Des images manuscrites débruitées par la méthode proposée en utilisant les différents types d'ondelettes

Remarque :

On a obtenu un bon résultat de débruitage :

- Nous avons constaté que parmi les types d'ondelettes utilisées, l'ondelette Biorthogonale donne le meilleur résultat ensuite l'ondelette de Haar car on aperçoit :
 - o Diminution remarquable du bruit.
 - o Préservation de la qualité de l'image.
 - o Elimination non-totale du bruit.
- Les autres types donnent un résultat non satisfiable.

3-3.2 Elimination des coefficients (détails) :

Le principe général de cette méthode consiste à décomposer l'image en ondelette à un niveau de décomposition, ensuite éliminer tous les informations situées dans les coefficients (détails) (la figure 5-13), en fin on fait la transformation en ondelette inversé, voici un exemple démonstratif de cette méthode :

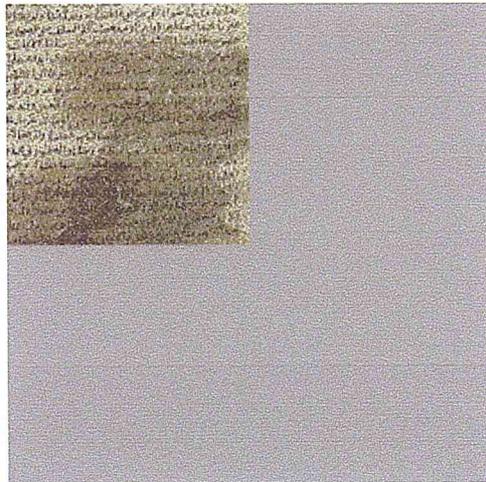
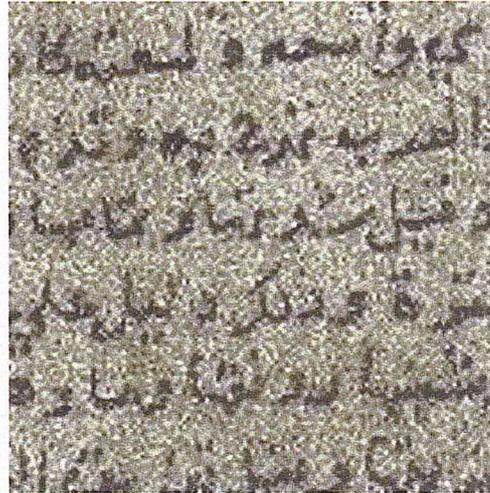
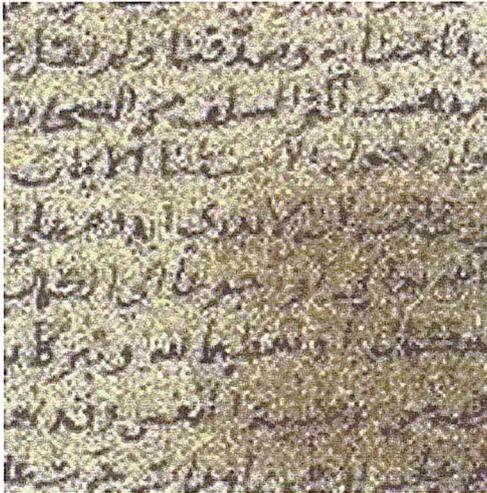
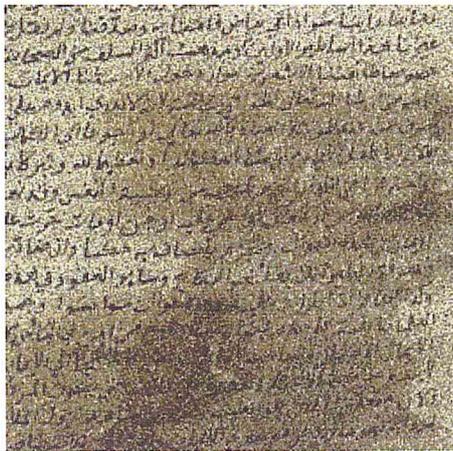


Figure 5-13 : Image après l'élimination des informations

Voici le résultat pour les différentes ondelettes :



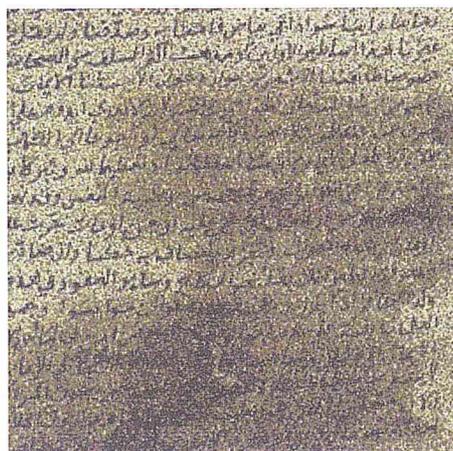
(a) Biorthogonale



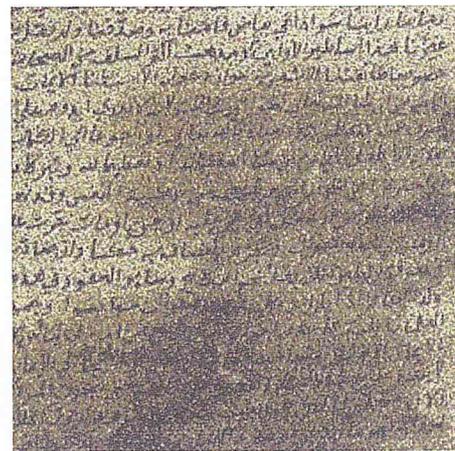
(b) Haar



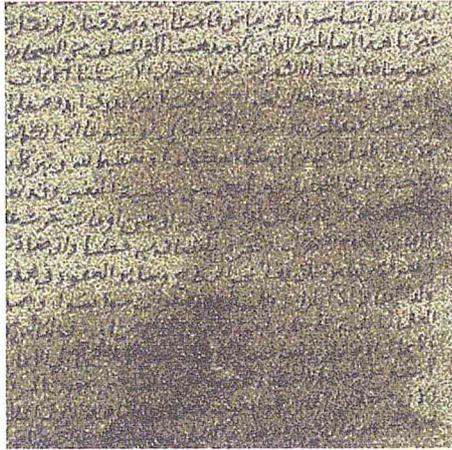
(c) Daubechies 2



(d) Daubechies 4



(e) Daubechies 6



(f) Daubechies 8



(g) Fast Symmlet 8

Figure 5-14 : Résultats après débruitage par élimination des coefficients (détails) avec les différents types d'ondelettes.

Remarque :

Cette méthode donne presque le même résultat que la méthode précédente mais plus mieux :

- Nous constatons que l'ondelette Biorthogonale donne le meilleur résultat aussi l'ondelette de Haar car on aperçoit :
 - o Diminution remarquable du bruit.
 - o Diminution de la qualité de l'image.
 - o Elimination presque totale du bruit.
- Les autres types donnent un résultat non satisfiable.

4- Conclusion :

Après avoir appliqué les méthodes de restauration (Seuillage, Détection des contours et Débruitage) en utilisant les ondelettes nous avons déduit que la restauration des anciens manuscrits par ondelettes ne donne pas des résultats optimaux mais des résultats acceptables.

Chapitre VI – Implémentation

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 :

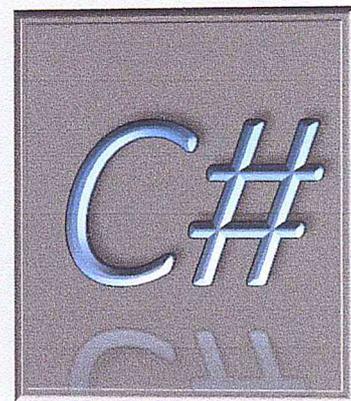
Pour implémenter notre système, nous avons utilisé les outils suivants :

2-1- Choix du langage :

Nous avons opté pour le langage C# pour l'implémentation car il offre plusieurs outils simples à utiliser qui aident beaucoup dans le domaine de traitement d'images.

Aperçu :

C# est un langage de programmation conçu pour la création d'une large gamme d'applications qui s'exécutent sur le .NET Framework. C# est simple, puissant, de type sécurisé et orienté objet. Avec ses nombreuses innovations, C# permet le développement rapide d'applications tout en conservant la simplicité et l'élégance des langages de style C. [65] [66]



- Il est compilé à un langage intermédiaire (CIL), indépendamment de la langue, il a été développé ou l'architecture cible et le système d'exploitation.
- La collecte des ordures automatique.
- Pointeurs ne sont plus nécessaires (mais en option).
- Capacités de réflexion.
- Ne pas besoin de s'inquiéter sur les fichiers d'en-tête ".h".
- Définition des classes et des fonctions peut se faire dans n'importe quel ordre.
- Déclaration de fonctions et de classes pas nécessaire.
- Inexistence de dépendances circulaires.
- Les classes peuvent être définies dans les classes.
- Il n'y a pas de fonctions ou variables globales, tout appartient à une classe.

- Toutes les variables sont initialisées à leurs valeurs par défaut avant d'être utilisé (ce qui est automatique par défaut, mais peut être fait manuellement à l'aide des constructeurs statiques).
- Vous ne pouvez pas utiliser des variables non booléennes (entiers, flottants ...) que les conditions. C'est beaucoup plus propre et moins d'erreurs.
- Les applications peuvent être exécutées dans un bac à sable restreint.

2-2- Choix de l'environnement de développement :

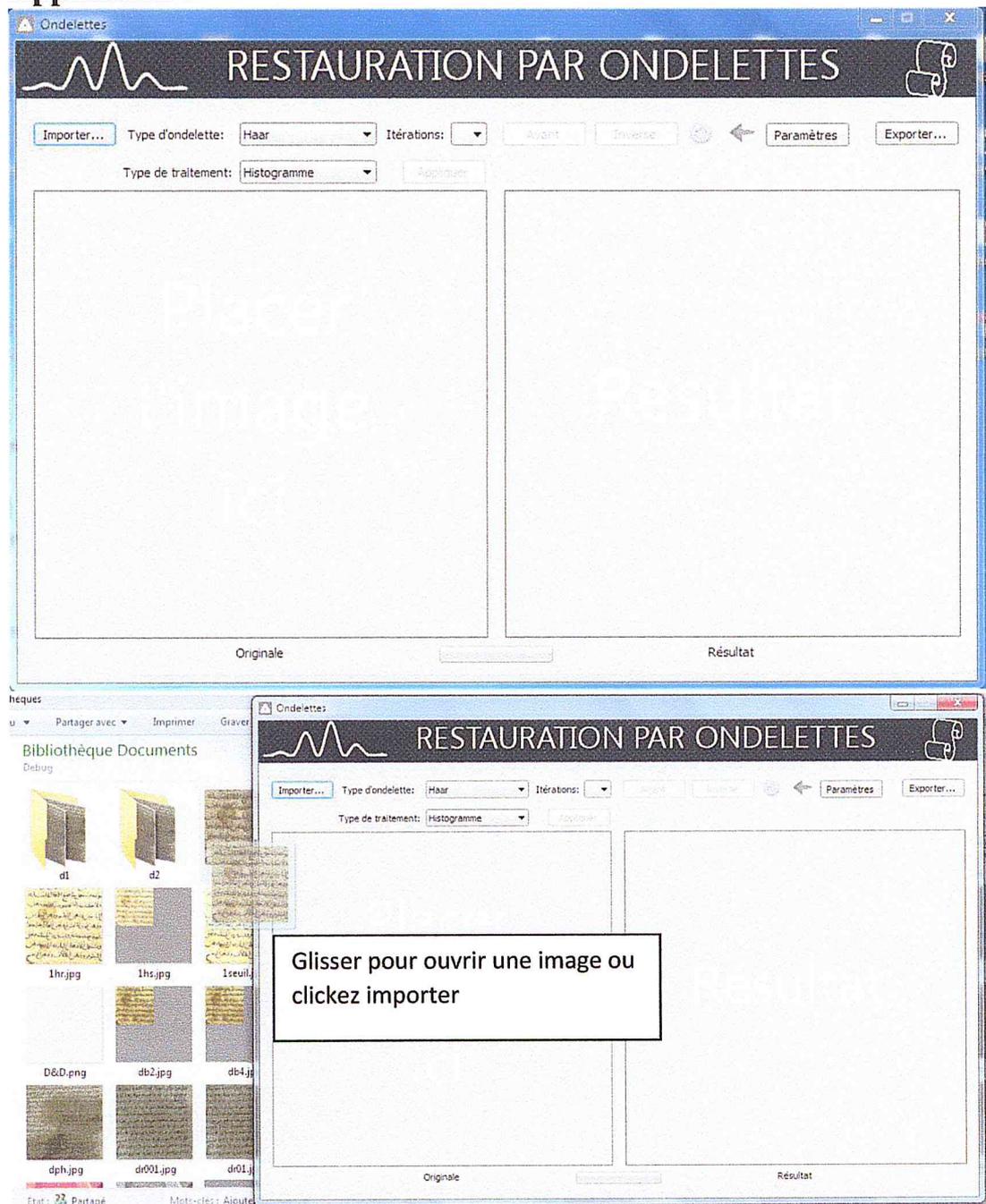
Nous avons choisis Visual Studio 2012 pour le développement de notre application, car il est simple et facile à utiliser, dont on a travaillé avec au pare avant.

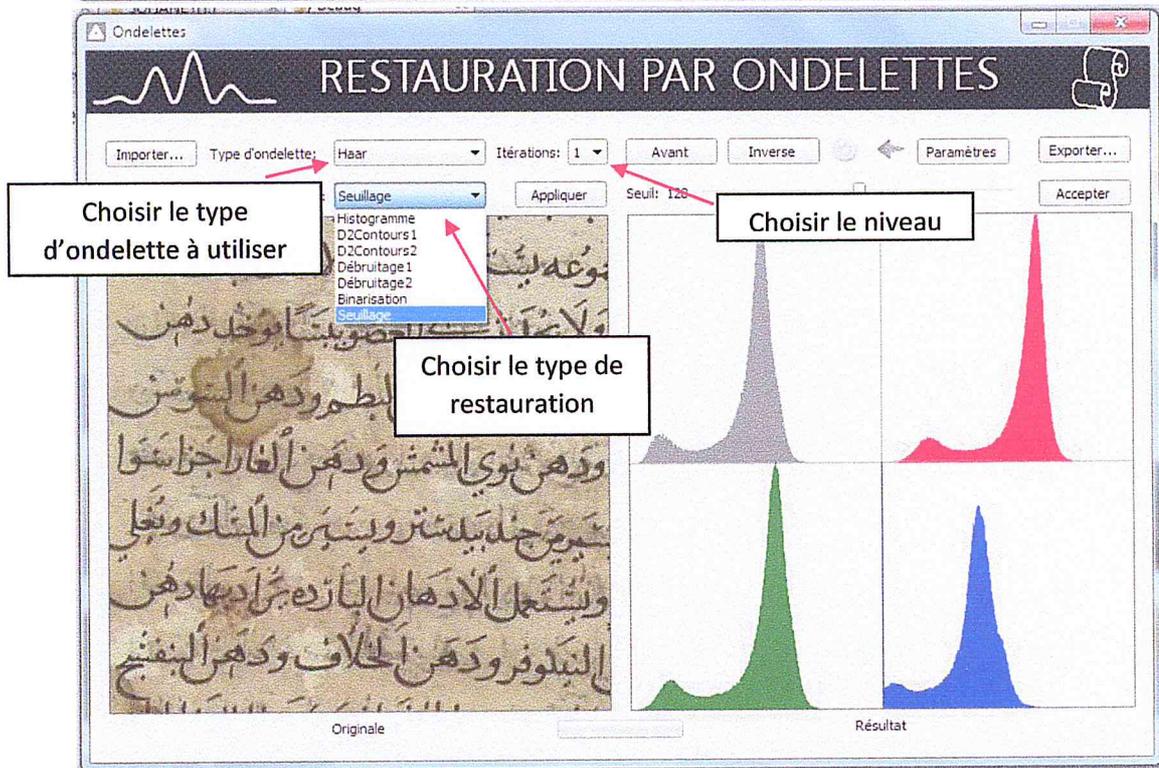
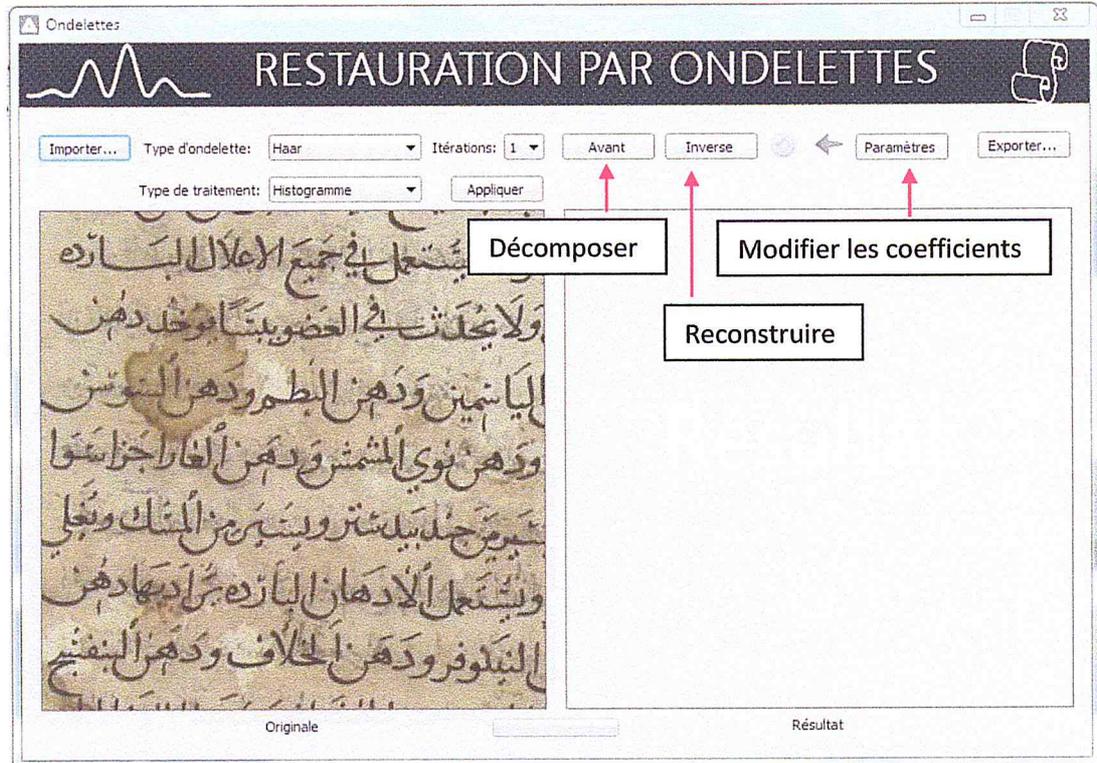
Aperçu :

Visual C# est une implémentation du langage C# par Microsoft. Visual Studio prend en charge Visual C# avec un éditeur de code complet, un compilateur, des modèles de projet, des concepteurs, des Assistants Code, un débogueur puissant et simple d'utilisation, ainsi que d'autres outils. La bibliothèque de classes .NET Framework donne accès à de nombreux services de système d'exploitation et d'autres classes utiles, bien conçues qui permettent d'accélérer considérablement le cycle de développement.



3- Application :







Conclusion Générale :

Le but global de notre recherche c'est de trouver l'apport des ondelettes au domaine de restauration des anciens manuscrits.

Après avoir examiné les résultats des tests nous avons pu déduire que les ondelettes peuvent être utilisés dans les débruitage et la détection des contours, mais cela ne veut pas dire que la transformée en ondelette est le meilleur moyen pour les réaliser, contrairement à la séparation fond/texte qui a donné de bons résultats.

A travers ce travail nous avons découvert le vaste domaine du traitement d'images qui constitue une grande partie de l'informatique de nos jours, et le domaine de la restauration des anciens documents qui sont une richesse du patrimoine, aussi nous avons appris plusieurs méthodes de restauration qui peuvent être utilisées dans plusieurs domaines autre que les anciens manuscrits.

Bibliographie :

- [1] : Web Science, 04/03/2014, « *Images numériques et stockage optique* », <http://www.web-sciences.com/documents/terminale/tedo20/teco20.php>.
- [2] : NOUGARET, Jean Claude, 09/03/2014, « *L'image Numérique* », www.nougaret.com/textes/image%20numerique,definitions.pdf.
- [3] : WACKER, Serge, 08/03/2014, « *les formats d'images numériques* », http://www.montpellier.iufm.fr/technoprimaire/c2i/revisions/formats_image.pdf
- [4] : MOKHBAT, Cherifa et Zineb, SIDI MOUSSA, « *Restauration des documents anciens : Texture Inpainting* », Université Saad Dahleb Blida, 2011/2012.
- [5] : Wikipedia, 08/03/2014, « *Rapports d'aspect et résolutions* », http://fr.wikipedia.org/wiki/D%C3%A9finition_d%C3%A9cran.
- [6] : Comment Ça Marche, 05/06/2014, « *Introduction au traitement d'images* », <http://www.commentcamarche.net/contents/1216-traitement-d-images>.
- [7] : Linternaute, 10/03/2014, « *Le contraste* », <http://www.linternaute.com/hightech/salon-multimedia/08/bien-choisir/tv-ecran-plat/2.shtml>.
- [8] : TECFA Education & Technologies, 10/03/2014, « *La résolution et les niveaux de gris* », <http://tecfa.unige.ch/themes/FAQ-FL/resolution/resolution.html>.
- [9] : MUNGUAKONKWA BIRINGANINE, Jean-marie, « *La liaison automatique des plusieurs images perçues sur un scanner* », Institut Supérieur Pédagogique de Bukavu, 2008.
- [10] : Primatice, 10/08/2014, « *Le codage des couleurs* », <http://primatice.net/logiciels/chromoweb/aide/codage.htm>
- [11] : Comment Ça Marche, 18/08/2014, « *Le codage CMYK (CMJN)* », <http://www.commentcamarche.net/contents/1193-le-codage-cmyk-cmjn>.
- [12] : JonTeaches, 10/08/2014, « *Color Theory for the Web* », <http://jonteaches.com/110/color-theory-for-the-web>.

- [13] : UVED, 10/08/2014, « *Le modèle TSL* », <http://e-cours.univ-paris1.fr/modules/uvved/envcal/html/compositions-colorees/representations-couleur/modeles-perceptuels/tsl.html>.
- [14] : Comment Ça Marche, 10/08/2014, « *Le codage CIE / Lab (L^*a^*b)* », <http://www.commentcamarche.net/contents/1192-le-codage-cie-lab-l-a-b>.
- [15] : DORANGE Pierre-Alain, 10/08/2014, « *Comprendre le codage couleur YUV* », <http://www.garage-video.com/spec/YUV.html>.
- [16] : Comment Ça Marche, 10/08/2014, « *Le format YIQ* », <http://www.commentcamarche.net/contents/1220-le-format-yiq>.
- [17] : Wikipedia, 07/03/2014, « *Manuscrits* », <http://fr.wikipedia.org/wiki/Manuscrit>.
- [18] : GENDREY, Patricia, « *Quelle stratégie numérique pour les éditeurs de livres ?* », MBA Marketing et Commerce sur Internet, 2009/2010.
- [19] : Bibliothèque Royale de Belgique, 06/08/2014, « *Le livre manuscrit au moyen âge, du VE au XVE siècle en occident* », http://www.kbr.be/educ/pedagog/profLeerkracht/ms_fr.pdf.
- [20] : MONELLE, 06/08/2014, « *Livres incunables : La Grande Histoire et les Légendes* », <http://monelle.eklablog.com/les-incunables-a53421745>.
- [21] : Seine-Saint-Denis Tourisme, 06/08/2014, « *Visite des Archives Nationales* », <http://www.tourisme93.com/document.php?pagendx=300>.
- [22] : MORDRET, Guy, « *Informatique éditoriale : nouveaux outils, nouvelles pratiques* », Parc d'Innovation de Mescoat 29800 Landerneau, France, 09 janvier 2009.
- [23] : FANNY, Lanz, « *Le livre au Moyen-âge* », Bibliothèque d'Etude et du Patrimoine / Mairie de Toulouse, France, 2005, p. 1-41.
- [24] : KALLDRËMXHIU, Enriketa, « *Les logiciels de numérisation des livres anciens* », Université Claude Bernard - Lyon 1, avril 2000
- [25] : FAYE, Emiliane, 07/08/2014, « *La numérisation des documents : opportunités et contraintes* », http://www.codesria.org/IMG/pdf/Emiliane_Faye.pdf.

[26] : DE LA SELLE, Xavier et Thierry, DELCOURT, « *La numérisation en bibliothèque et en archives* », Les services d'archives et les centres de documentation de Champagne-Ardenne, janvier 1999.

[27] : Bibliothèque de l'Université Cornell - Département de Recherches, 09/08/2014, « *Didacticiel d'imagerie numérique : Création d'images* », <https://www.library.cornell.edu/preservation/tutorial-french/technical/technicalB-03.html>.

[28] : « *La numérisation des documents – Méthodes et recommandations* », Direction générale des archives - Bibliothèque et Archives nationales du Québec, Mai 2012.

[29] : « *Stage d'initiation au manuscrit médiéval* », Institut de recherche et d'histoire des textes, Vol 4, 5 Oct 2009.

[30] : YOUS, Mohamed et Ahmed, BENELKAID, Mémoire magister de « *Reconnaissance des caractères numériques par la transformée en ondelette et le réseau de neurone multicouche (WT-MLP)* », Université Saad Dahleb Blida, Novembre 2000.

[31] : SOLTANI, Ouarda, « *Restauration D'Images Satellites Via la Transformée en Ondelette* », Université de Batna, 2011.

[32] : M. Brucher, « *La transformée de Fourier rapide et d'une de ses dérivées* », <http://matthieu-brucher.developpez.com/tutoriels/algo/fft/>, Mai, 2007.

[33] : H. Florent, « *La Transformée de Fourier en Traitement d'Images* », Avril 2007.

[34] : Adapt Science, 16/09/2014, « STFT », <http://www.adeptscience.co.uk/wp-content/uploads/2014/03/STFTwithTimeSeries.png> .

[35] : GRAPS, Amara., « *An introduction to Wavelets* », *IEEE Computational Science and Engineering*, Vol. 2, N° 2, 1995, p. 50-61.

[36]: O. D. Faugeras, « *Digital color image processing within the framework of a human visual model* », *IEEE Trans Signal Processing*, vol. ASSP-27, no. 4, 1979, 380-393p.

[37] : O. Rioul, et Vetterli M., « *Wavelets and signal processing* », *IEEE Trans. Signal Process*, vol. 8, no. 4, 1991, 14-38p.

- [38]: L. Kovacs et Sziranyi T., « *Focus area extraction by blind deconvolution for defining regions of interest* », IEEE Trans. Patt. Anal. Mach. Intell., vol. 29, no. 6, 2007, 1080-1085p.
- [39] : M. M., Chang, Tekalp, A. M., et A. T., Erdem, « *Blur identification using the bispectrum* », IEEE Trans. Image Process., vol. 39, no. 10, Oct. 1991, 2323–2325p.
- [40] : HOVHANNES, Bantikyan, 17/03/2014, « *Discrete Haar Wavelet Transformation* », <http://www.codeproject.com/Articles/683663/Discrete-Haar-Wavelet-Transformation>.
- [41] : MathWorks, 17/03/2014, « *Lifting Method for Constructing Wavelets* », <http://www.mathworks.com/help/wavelet/ug/lifting-method-for-constructing-wavelets.html>.
- [42] : Lentilles gravitationnelles, 17/03/2014, « *Transformée en ondelettes de Haar* », http://home.base.be/epm6604b/ondelette_haar.html.
- [43] : Lentilles gravitationnelles, 17/03/2014, « *Transformée en ondelettes de Daubechies 4* », http://home.base.be/epm6604b/ondelette_daubechies.html.
- [44] : Wikipedia, 18/03/2014, « *Daubechies wavelet* », http://en.wikipedia.org/wiki/Daubechies_wavelet.
- [45] : TEENA, Varma, Chitre, VIDYA et DIPTI, « *The Haar Wavelet and The Biorthogonal Wavelet Transforms of an Image* », National Conference on Emerging Trends in Engineering & Technology, 30/03/2012, p. 288-291.
- [46] : Lentilles gravitationnelles, 18/03/2014, « *Galerie d'ondelette Biorthogonale* », http://home.base.be/epm6604b/ondelettes/Galerie_ondelette_biorthogonales.html.
- [47] : DRIRA, F., F., LEBOURGEOIS et EMPTOZ, H., « *Séparation recto/verso des images de documents anciens par une approche aveugle* », Laboratoire d'InfoRmatique en Image et Système d'Information (LIRIS), INSA de LYON- Bât Jules Verne, 2006.
- [48] : SHARMA, Gaurav, « *Cancellation of show-through in duplex scanning* », Appears in Preceedings IEEE International Conference on Image processing, vol.2, 2000, p. 609-612.

- [49] : DUBOIS, Eric et Anita, PATHAK, «*Reduction of bleed-through in scanned manuscript documents*», dans *processing IS&T Image processing, Image Quaity, Image capture Systems*, 2001, p. 177-180.
- [50] : WANG, Qian, Tao, XIA, LIM, Chew et Lida, LI, «*Directional Wavelet Approach to Remove Document Image Interference* », ICDAR 2003, p. 736-740.
- [51] : LEYDIER, Y., F., LE BOURGEOIS et EMPTOZ, H., «*Serialized Unsupervised Classifier for Adaptative Color Image Segmentation : Application to Digitized Ancient Manuscripts* », Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04), IEEE Computer Society, LIRIS, I.N.S.A. de LYON, France.
- [52] : WANG, Xiaojun, «*Recovery of Blurring Scanned Manuscript Image Based on Wavelets Transform Algorithm* », 3rd International Congress on Image and Signal Processing (CISP2010), North China Electric Power University, 2010, p. 844-847.
- [53] : S.BAIRD, Henry, «*Document Image Defect Models and Their Uses* », Proc of the International Conférence on Document Analysis and Recognition, 1993, p. 62-67.
- [54] : KANUNGO, Tapas, Robert M., HARALICK, Henry, S. Baird, Stuezle, WERNER et MADIGAN David, «*A Statistical, Nonparametric Methodology for Document Degradation Model Validation*», IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, N° 11, NOVEMBER 2000, p. 1209-1223.
- [55]: ZHENG, Qigong et Tapas, KANUNGO, «*Morphological degradation models and thier use in document image restoration* », IEEE, 2001, p. 193-196.
- [56] : DRIRA, Fadoua, *thèse de « Contribution à la restauration des images de documents anciens »*, École Doctorale Informatique et Information pour la Société (EDIIS), Lyon, France, 14/12/2007.
- [57] : D. HOBBT, John et Tin, KAM HO, «*Enhancing Degraded Document Images via Bitmap Clustering and Averaging* », Proceeding of the international conference Analysis and Recognition, 4th ICDARn Germany1997, p. 394-400
- [58] : Qigong ZHENG and KANUNGO Tapas, «*Morphological degradation models and thier use in document image restoration* », IEEE, 2001, p. 193-196.

[59] : CAO, Huaigu, Xiaoqing, DING et CHANGSONG Liu, « *Rectifying the Bound Document Image Captured by the Camera A Model Based Approach* », IEEE, Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR), Department of Electronic Engineering, Tsinghua University, Beijing, China 2003.

[60] : BERGOUNIOUX, « *Quelques méthodes mathématiques pour le traitement d'image* », Cours Master 2, 04-01-2009.

[61] : FRITZ, Carole, Gilles, TOSELLO, AZÉMA, Marc, Olivier, MOREAU, PERAZIO, Guy et PÉRAL José, « *Restauration virtuelle de l'art pariétal paléolithique : le cas de la grotte de Marsoulas* », In Situ 13 revue des patrimoines, 2010, p. 1-10.

[62] TSIN, Yanghai, Visvanathan, RAMESH et KANADE, Takeo, « *Statistical Calibration of CCD Imaging Process* », IEEE, 2001, p. 480-487.

[63] : KEVEN, Jensen, 11/09/2014, « *lets learn math photoshop blend modes* », <http://www.venture-ware.com/kevin/coding/lets-learn-math-photoshop-blend-modes/>.

[64] : CHARRADA, Mohamed Aymen, Anis, KRICHA et SOKRI BEN AMARA, Nadjoua, « *ترميم و معالجة صور ووثائق الأرشيف القديمة* », Université de Sousa, Tunisie.

[65] : Microsoft Developer Network, 08/09/2014, « *Visual C#* », <http://msdn.microsoft.com/fr-fr/library/kx37x362.aspx>.

[66] : **Les avantages du C# : Ref** : URRIELLUnet, 08/09/2014, « *Advantages of C#* », <http://urriellu.net/en/articles-software/csharp-advantages.html>.